Article

Perceiver Effects in First Impressions Reflect Generalized Stereotypes: Evidence of Consistency Across Time, Groups, and Contexts

Richard Rau[1] , **Nicole M. Lawless DesJardins[2], Lisa Maria Niemeyer[1],**
Mitja D. Back[1], Sanjay Srivastava[2], and Steffen Nestler[1]

Abstract
In interpersonal perception, perceivers' tendencies for judging the average target (perceiver effects) are commonly assumed to reflect generalized stereotypes about "the other." This is empirically supported by findings of consistent rank-orders of perceiver effects across measurement occasions, but previous studies could not rule out important alternative explanations for consistency. Here, we present a strict test of the generalized stereotype account in two studies (*n*s 146 and 295) in which participants provided mutual impressions in distinct, uniquely composed groups. Results reveal that perceiver effects are quite consistent on the level of global evaluation and acquiescence but less consistent on the level of specific trait- or item-content. The finding that perceivers who saw others in globally positive or negative ways in one situation also saw other targets in similar ways in entirely different situations across time, groups, and contexts is strong evidence for the generalized stereotype account and has vital implications.

Keywords
perceiver effects, interpersonal perception, consistency, rank-order stability, generalized stereotype, structural equation modeling, implicit personality theory, judgment bias, positivity, acquiescence

People view their social environment through idiosyncratic lenses. For instance, Chris might be generally critical when forming impressions of others, whereas Leo is generally lenient. Such differences, often termed *perceiver effects*, potentially bear an interpretation of great theoretical and practical relevance. Theoretically, if the same differences manifest consistently across time and social contexts, this would corroborate the view that they reflect generalized stereotypes or learned assumptions about "the average other." Practically, since person impression data are collected in numerous contexts of modern life such as on dating platforms, during personnel selection, or in group therapies, it is instructive to know to what degree people's judgment tendencies generalize beyond these specific contexts. Here, we present the first investigation of perceiver effect consistency in a systematic cross-situational design.

Perceiver Effects

Perceiver effects are defined as the average rating provided by someone (i.e., perceiver) about several others (i.e., targets) concerning a certain trait (Kenny, 1994). Typically, researchers collect these ratings in small groups where everyone rates one another and acts as both perceiver and target. According to the Social Relations Model (SRM; Kenny, 1994), each judgment then decomposes into a constant (the mean judgment), the perceiver effect (how the perceiver's average judgment deviates from the mean), the target effect (how the average judgment about the target deviates from the mean), and a relationship effect (a residual term). If Chris and Leo belonged to the same basketball team and were to judge all of their team members' outgoingness below and above average, respectively, Chris's perceiver effect would be negative and Leo's would be positive.

Because researchers and lay people are often most interested in target-level phenomena (e.g., is there a team member who is consensually seen as most outgoing?) or in relationship-level phenomena (e.g., are some team

[1]University of Münster, Germany
[2]University of Oregon, Eugene, USA

Corresponding Author:
Richard Rau, University of Münster, Fliednerstr. 21, 48149 Münster, Germany.
Email: richard.rau@uni-muenster.de

members seen as outgoing by some but not by other team members?), it may be easily overlooked that perceiver effects typically produce a considerable portion of the variation found in personality judgments (e.g., some team members see *everyone* as more outgoing than other members do). Specifically, perceiver effects tend to explain between 20% and 30% of the variation in person perception data no matter the degree of acquaintanceship between perceivers and targets (see Kenny, 2019 for an overview).

An intriguing interpretation of perceiver effects is that they reflect people's working models or generalized stereotypes about "the other." In classic theorizing, the generalized stereotype is conceptualized as a pervasive force shaping perceivers' interpersonal experiences throughout life (Bowlby, 1988; Erikson, 1959, 1968). The idea is that, while some have had reassuring experiences early in life and have internalized positive expectations about future interaction partners, others bring skepticism and distrust to every new encounter they make. Yet, for such interpretation to be potentially justified, there exist important prerequisites in terms of how perceiver effects are measured.

Measurement

In which setting should perceiver effects be assessed? One criterion is that perceivers should have the same knowledge about targets. For instance, in a freshly composed team of basketball players, everyone bases their judgment on the same, limited amount of information, and perceiver effects will reflect personal judgment tendencies. If, in contrast, perceivers have unique prior experiences with the others, perceiver effects would also reflect the specific knowledge a perceiver has about targets. Imagine Chris had joined Leo's basketball team just recently: He might judge everyone's outgoingness rather neutrally, whereas Leo, having experienced the team being very outgoing in the past, provides higher ratings. As a consequence of this, perceiver effects would reflect a blend of perceivers' judgment tendencies and their actual experiences with targets. One way to avoid this would be to have perceivers rate targets they never met in person such as celebrities or strangers on photographs. However, designs treating perceptions as though they happened through "one-way mirrors" have been rightfully criticized for their lack of ecological validity (Kenny & Albright, 1987, p. 393) such that designs with newly composed groups of strangers are clearly the better option.

Apart from the advantage of avoiding unique knowledge about targets among perceivers, a second reason why perceiver effects are best studied in a low-acquaintanceship setting is that generalized stereotypes are especially relevant in this kind of context. That is, perceiver effects found in first impressions necessarily reflect people's *expectations*, thus warranting the assumption that they might stem from their working models about "the average other." In well-

acquainted groups, in contrast, perceiver effects likely partly also tap variation in how much someone enjoys being part of their group (Kenny, 2019), an experience that may be unrelated to generalized stereotypes.

For which attributes should perceiver effects be measured? Research on the factorial structure of perceiver effects has revealed that they reflect not only differences in overall leniency (i.e., *positivity*) but also in how people view others regarding particular trait content such as outgoingness (i.e., *trait-specificity*) and in how they use rating scales (i.e., *acquiescence*; Heynicke et al., 2021; Rau, Carlson, et al., 2021; Srivastava et al., 2010). For instance, Leo's high perceiver effect for outgoingness may be produced by a tendency to see the best in others but also by a tendency to see others as highly extraverted in particular or by his mere preference for the upper parts of rating scales. Importantly, the contribution of trait-specific influences has been found to be strongest for traits that are easily observable and evaluatively neutral (e.g., extraversion) but it is often negligible for less observable and more evaluative traits such as agreeableness (Rau, Carlson, et al., 2021). To allow for a modeling approach that formally disentangles positivity, trait-specificity, and acquiescence, perceiver effects ought to be assessed for multiple desirable and undesirable attributes. Moving beyond personality perceptions, perceiver effects might additionally be studied for exclusively evaluative judgments, for example, for *liking* but also *meta-liking* (i.e., how much the perceiver expects to be liked by the target).

Consistency

Assuming that perceiver effects have been appropriately measured and modeled, how would evidence for a generalized stereotype interpretation look? Because the essence of a generalized stereotype is its pervasiveness, strong evidence would need to concern the *consistency* of the perceiver effect. That is, a person who showed a certain judgment tendency in situation A (i.e., at a certain point in time in a certain group which meets in a certain context) would be expected to show a similar tendency in situation B (i.e., at a different point in time in a different group which meets in a different context). To the degree that perceiver effects were instable across any of these aspects (i.e., temporal, cross-group, cross-context), they would not reflect a generalized stereotype but rather a momentary or local belief about what others are like in a specific situation.

To summarize, to test the generalized stereotype hypothesis, perceiver effects should (a) be measured using targets about whom the perceiver has no unique knowledge, (b) be measured in a setting that involves personal encounters, and (c) be discerned from acquiescence bias. Under these conditions, consistency across time, groups, and contexts would reflect strong evidence of a generalized stereotype account of perceiver effects.

Table 1. Existing Studies on Perceiver Effect Consistency

Reference	Type of consistency evidence	Average level of consistency[a]	Considered construct	(a) Did all perceivers have the same knowledge about targets?	(b) Did perceivers personally meet targets?	(c) Was acquiescence accounted for?
Marcus and Miller (2003; as cited in Kenny, 2019)	Cross-group (targets of same vs. opposite gender)	$r = .55$	Attractiveness	Yes	Yes	No
Kenny et al. (2015)	Cross-group (baseball fans of same vs. opposite teams)	$r = .34$	Liking	Yes	Yes	No
Rau et al. (in press)	Over time (several minutes to several days) + cross-context (various group activities)	$r = .83$	Positivity	Yes	Yes	Yes
Rau, Nestler, et al. (2021)	Over time (2 weeks) + cross-group (strangers vs. classmates)	$r = .43$	Positivity	Yes	No	Yes
Srivastava et al. (2010, Study 2)	Over time (1 week) + cross-context (various group activities)	$r = .69$	Big Five	Yes	Yes	No
Wood et al. (2010, Study 3)	Over time (1 year)	$r = .69$	Positivity	No	Yes	Yes

[a]Most studies reported several effect sizes (e.g., for different subgroups, at different timepoints, for different traits) and some studies reported correlations that were attenuated due to measurement error in perceiver effects. For ease of presentation, we manually combined and disattenuated effect sizes where needed.

Previous Evidence of Consistency

A few studies have provided evidence favoring a generalized stereotype account of perceiver effects. Specifically, they all found consistency in some aspects (i.e., temporal, cross-group, cross-context) but they did not test these aspects at once. Some of the studies focused on perceiver effects for particular traits (Srivastava et al., 2010), others focused on positivity across traits (Rau, Nestler, et al., 2021; Wood et al., 2010), and others on judgments of attraction (Kenny et al., 2015; Marcus & Miller, 2003). In addition, some studies compared perceiver effects toward an ingroup versus outgroup (Kenny et al., 2015; Marcus & Miller, 2003), while the others did not involve such distinction. Table 1 provides an overview of all previous studies we are aware of.

Most of the studies in Table 1 did not meet at least one of the above stated prerequisites for interpreting perceiver effects as generalized stereotypes, rendering the overall evidence for this account unclear. Specifically, three of the studies did not rule out that consistency was driven by stable differences in how perceivers use rating scales (Kenny et al., 2015; Marcus & Miller, 2003; Srivastava et al., 2010). However, such confounding is not unlikely, given that research outside the person perception realm has shown that people's mere preferences for higher or lower scale values (i.e., acquiescence bias) are indeed highly stable (Crow & Hammond, 1957; Wetzel et al., 2016). Furthermore, Wood et al. (2010) could not rule out that perceiver effects reflected people's unique knowledge about targets, and

Rau, Nestler, et al. (2021) could not—at least not in two out of three studies—rule out that consistency was limited to abstract others who would never be met in person.

The strongest piece of current evidence comes from Rau et al. (in press) where consistency was found across time and contexts. The authors analyzed data from three studies in which small groups met over the course of several weeks and completed different tasks (e.g., icebreaking games, discussions, problem-solving) and provided mutual personality ratings before and after each task. They found perceiver effects to be quite consistent but, importantly, the groups were composed of the same individuals at every meeting. It is thus possible that consistency was partly driven by group-specific stereotypes. To illustrate, those who enjoyed their group might have consistently evaluated others favorably but they would not have necessarily done so had they been part of a different group at every measurement occasion.

Indeed, an inspection of the effect sizes shown in Table 1 suggests that a change in groups is the most influential source of inconsistency, given that cross-group studies (Kenny et al., 2015; Marcus & Miller, 2003; Rau, Nestler, et al., 2021) have found much smaller timepoint-to-timepoint correlations than did same-group studies (Rau et al., in press; Srivastava et al., 2010; Wood et al., 2010). Furthermore, additional evidence for group-specific stereotypes was found both in Rau et al. (in press) and in Srivastava et al. (2010), where consistency was weaker early in the study (e.g., between the first and second meeting) and stronger later on (e.g., between the second to last

and last meeting). This suggests that perceivers formed and consolidated group stereotypes as they became better acquainted, and it points toward the necessity of investigating consistency in a design where participants are members of different groups (Srivastava et al., 2010).

The Present Investigation

The above has made clear that the strongest support for a generalized stereotype account of perceiver effects would come from a design where perceivers provide judgments of targets which they have not previously known and who they meet at different points in time, in different group constellations, and in different contexts. Obviously, collecting this kind of data is challenging as it requires running a classic zero-acquaintance group study several times, rotating the assignment of individuals to groups. Here, we present evidence for perceiver effect consistency from two such studies for the first time.

In Study 1, participants attended two group meetings spaced 2 to 4 weeks apart, each time completing mutual personality perceptions after discussing a moral dilemma in a newly composed group of strangers. Thus, the study allows for a strict test of temporal and cross-group consistency. As both meetings involved the discussion of a moral dilemma, however, Study 1 does not test cross-context consistency. In Study 2, participants attended a larger meeting in which they completed up to four different tasks, every time being part of a newly composed group of strangers. As all tasks were completed within one meeting, Study 2 does not test temporal consistency.

Study 1

Study 1 was part of a larger project on hindsight bias in personality judgments. The sample size ($n = 146$) was not planned for the present research but allows the detection of moderately sized correlations ($r = .25$) with high statistical power ($1 - \beta = .87$) at a 5% alpha level. We focus on describing those measures and procedures that are directly relevant for the present analyses. A comprehensive description of the entire project is retrievable from https://osf.io/ptfs8/ and data and R-code are retrievable from https://osf.io/cxp3b/. The data have been used in an existing publication (Rau, Carlson, et al., 2021) which was unrelated to perceiver effect consistency.

Participants and Procedure

Participants were recruited via e-mail lists, flyers, and social media postings and were incentivized by a monetary compensation at an hourly rate of approximately 8€. They were students from different disciplines (age: $M = 22.72$, $SD = 2.73$). 77% of them identified as female and 23% as male. Nationality was not assessed but we expect that between 80% and 90% of the sample were German. After

signing up, participants were scheduled for two group meetings. They were told the names of the other group members and rescheduled in case they were familiar with any of them. Eight participants showed up at only one of the meetings. Most of the 49 meetings were held in groups of six (84%) and the rest in groups of five (12%) or four (4%). Upon arriving to the first meeting, participants were briefed about the study goals and procedures and provided written informed consent.

In both group meetings, participants first introduced themselves and then engaged in a 12-min discussion about a moral dilemma. Both dilemmas had a similar structure but differed in terms of content to make sure a lively discussion would be triggered in both meetings. Verbatim instructions are shown in Appendix A. After the discussion, participants anonymously rated each other on several dimensions using tablets. These dimensions concerned interpersonal traits, familiarity with the target, attraction (liking and meta-liking), neuroticism (nervousness and calmness), and perceptions of specific cues (e.g., about the target's voice or facial expression). Here, we focus on perceptions of interpersonal traits and attraction. However, we used the familiarity judgments to verify that participants were indeed unacquainted prior to the study. On a 5-point scale asking "how well did you know this person before today?," 97.5% of the dyads reported a 1 (*not at all*) and 1.6% reported a 2 (*seen once*). Dyads who reported higher familiarity were excluded from the analyses.

Measures

We collected interpersonal trait perceptions with eight items from the Interpersonal Adjective List (IAL; Jacobs & Scholl, 2005) and judgments of liking and meta-liking using a 5-point scale (1 = *strongly disagree*; 5 = *strongly agree*; see Table 1). The IAL is designed to capture the space that is spanned by the two orthogonal dimensions of *agency*, which encompasses traits relevant for *getting ahead* of others, and *communion*, which encompasses traits relevant for *getting along* with others (Bakan, 1966; Hogan et al., 1985). Specifically, the IAL features items for each of the eight octants of the interpersonal circumplex and thereby explicitly taps the different combinations of high versus medium versus low agency and high versus medium versus low communion. For instance, *assertive* taps high agency/medium communion and *compliant* taps low agency/high communion. Although communion is typically conceptualized as a *content* domain, we note that this domain is heavily saturated with evaluativeness (John & Robins, 1993) such that judgments of communion are often practically indistinguishable from mere evaluations along a continuum from negative to positive (Leising et al., 2021; Rau, Carlson, et al., 2021).

For each of the 10 items, we ran an SRM analysis using the *R* (version 3.5.1; R Development Core Team, 2008) package *TripleR* (version 1.5.3; Schönbrodt et al., 2012)[1]

Table 2. Study 1 Standardized SRM Variance Components and Perceiver Effect Reliabilities by Item

Item	Mapping	Meeting 1		Meeting 2	
		P/T/R	Rel$p$	P/T/R	Rel$p$
Assertive	agy + / com 0	(.03)/.48/.49	.22	.11/.45/.44	.54
Cynical	agy + / com −	.27/.10/.64	.66	.32/.13/.55	.72
Hostile	agy 0 / com −	.33/.10/.56	.73	.19/.16/.65	.56
Unsociable	agy − / com −	.20/.23/.56	.63	.17/.23/.61	.55
Shy	agy − / com 0	.06/.41/.54	.32	.11/.41/.48	.52
Compliant	agy − / com +	.29/.20/.51	.73	.26/.22/.52	.70
Sensitive	agy 0 / com +	.18/.18/.65	.55	.22/.21/.57	.63
Outgoing	agy + / com +	.17/.34/.49	.62	.21/.31/.48	.66
How much do you like this person? (liking)		.26/.09/.65	.64	.24/.19/.57	.65
How much does this person like you? (meta-liking)		.39/(.03)/.58	.76	.41/.06/.53	.78

Note. SRM = Social Relations Model; P/T/R = standardized SRM variances for the perceiver / target / relationship (incl. error) component; coefficients in parentheses are not statistically different from zero ($p < .05$); Rel$p$ = reliability of individual perceiver effect scores; agy = agency; com = communion.

and saved perceiver effect estimates for later use in a structural equation model (SEM). As shown in Table 2, the variation in dyadic personality judgments was driven by between-perceiver differences to a considerable degree (19%–20% on average) but also by between-target and between-dyad differences (25%–27% and 56–53%, respectively). For liking and meta-liking in particular, perceiver differences were even more pronounced than for personality perceptions (up to 41% perceiver variance). Importantly, the majority of items yielded perceiver variances above 10% which is recommended as a minimum to warrant further investigation of an SRM component (Kenny, 1994). The only exceptions concern low perceiver variances for assertiveness and shyness in the first group meeting. They are likely explainable by the high observability (and, hence, high relative target variance) of these attributes, which is a common finding in SRM research. This also manifested in a low reliability of the individual perceiver effect scores for assertiveness and shyness, whereas reliability was acceptable for all remaining items.

Results and Discussion

To investigate the degree of consistency in perceiver effects for interpersonal traits across the two meetings, we specified an SEM that discerns positivity, acquiescence, and trait-specificity at each measurement occasion and, for each of these components, estimates the correlation across the two measurement occasions. Thereby, the model uncovers how each of the components contributes to consistency in perceiver effects. To find a good balance between a model flexible enough to fit well to the observed data on the one hand, and a model parsimonious enough to be robustly estimated while avoiding overfitting on the other hand, we compared several alternative specifications that imposed differing numbers of constraints on the model's parameters. We used the full information maximum likelihood estimator as implemented in the *lavaan* package in R (version 0.6-

7; Rosseel, 2012) to fit the models. We focus on the model with the best balance in the main text and present alternative specifications in Appendix B. A zero-order correlation matrix of all observed variables can be retrieved online (https://osf.io/cxp3b/).

Figure 1 illustrates the best-balanced model. Positivity was specified as a latent variable on which all perceiver effect variables were allowed to load freely (except for a marker variable used for model identification). Acquiescence was modeled as a latent variable on which all indicators loaded with a fixed value of 1. Finally, a latent variable accounting for agency-specific trait content was modeled. Based on interpersonal theory, we specified this variable with fixed loadings of 1 for *assertive*, −1 for *shy*, and 0 for *sensitive* and *hostile*. To allow for the possibility that loadings for "medium-agency" indicators may not be located perfectly midway between the circumplex axes (i.e., loadings of exactly .5 for *cynical* and *outgoing* and loadings of exactly −.5 for *compliant* and *unsociable*), they were estimated freely. Consistency correlations across measurement occasions were estimated freely for positivity, acquiescence, and agency, as well as for item-level residuals.

Judging against conventional cutoffs, absolute model-fit was borderline, $\chi^2(df) = 192(103)$, comparative fit index (CFI) = .900, root mean square error of approximation ($RMSEA$) = .077, standardized root mean square residual ($SRMR$) = .108. Yet, given the model's good performance relative to less parsimonious models (see Appendix B) and in light of previous research yielding borderline fit for short personality inventories (Rau, Carlson, et al., 2021; Srivastava et al., 2010), we deemed this level of fit acceptable. As can be seen in Figure 1, socially desirable attributes such as *sensitive* and *outgoing* had positive loadings on the positivity factors, whereas undesirable attributes such as *unsociable* and *hostile* had negative loadings. This justifies labeling the factor "positivity" and lines up well with earlier findings on the structure of perceiver effects (Heynicke et al., in press; Rau, Carlson, et al., 2021; Srivastava et al.,
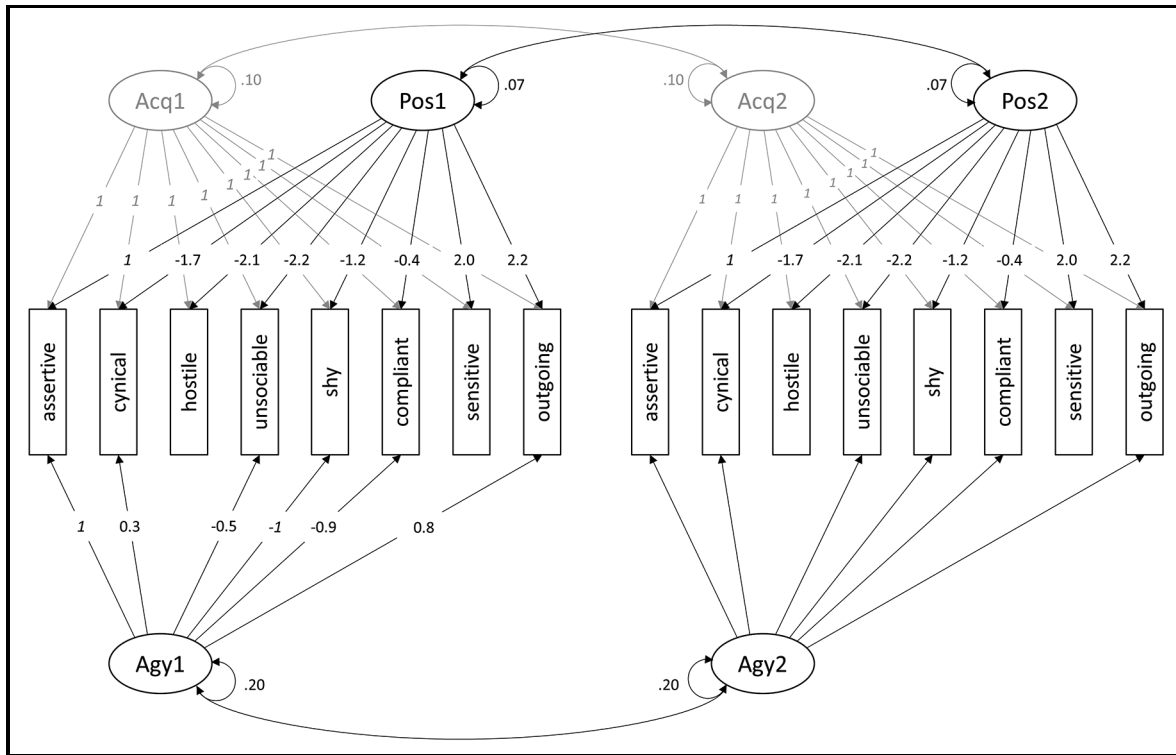
Figure 1. Study 1 Structural Equation Model With Perceiver Effects for Each Item as Observed Variables and Positivity (Pos), Acquiescence (Acq), and Agency (Agy) as Latent Variables *Note.* All parameters were constrained to be equal across timepoints (strict measurement invariance). Loadings in italics were fixed. Item-residual (co)variances are not displayed for clarity.

2010). As for the other latent variables, the variance of the acquiescence factor was significantly different from zero ($p < .001$) which confirms that ignoring differences in scale use risks inflating the variance of the positivity factor and blurring its meaning. Finally, the loading pattern for the agency factor was in line with expectations (i.e., loadings for *cynical* and *outgoing* fell between 0 and 1 and loadings for *compliant* and *unsociable* fell between 0 and −1). Note that we also tried specifying a communion factor but model comparisons suggested that this factor would not have had any explanatory utility above and beyond positivity and would overparameterize the model (see Appendix B).

How consistent were people's judgment tendencies across the two meetings? For positivity, the estimated correlation between measurement occasions was $r = .46$ (95% confidence interval [CI] = [.28, .63]) which suggests that almost half of the variance in the overall positivity versus negativity of trait judgments was indeed consistent across measurements. We also found substantial consistency for acquiescence, $r = .57$ (95% CI = [.40, .74]), which lines up with previous studies and highlights the danger of response-bias inflating stability estimates of rating data. Finally, we found support for consistency of agency-specific perceiver effects, $r = .29$ (95% CI = [.01, .56]), but this effect was more moderate compared with positivity

and acquiescence. Similar levels of consistency were found for most but not for all of the item-level residuals (for details, see Appendix C). This partial lack of consistency may go back to a lack of reliability in the perceiver effect variables we used as indicators (see Table 2).

Was there also consistency in perceiver effects for liking and meta-liking? To investigate this, we computed the Pearson correlation between the respective perceiver effect scores from the two meetings, yielding $r = .38$ (95% CI = [.23, .51]) for liking and $r = .44$ (95% CI = [.29, .56]) for meta-liking. Although significantly different from zero, these consistency correlations may appear smaller than the ones observed for global positivity in interpersonal trait perceptions. However, other than positivity, liking and meta-liking are manifest variables measured with error. When correcting for the unreliability of these scores (see Table 2), their consistency was quite sizable, again accounting for roughly half of the variance ($r = .59$; 95% CI = [.35, .80] for liking and $r = .57$; 95% CI = [.38, .73] for meta-liking).

In sum, the most important finding from Study 1 is that people's tendencies for globally positive versus negative perceptions were highly stable not only over the course of several weeks but also across different groups. This was also mirrored in the finding that perceivers who explicitly said that they liked and expected to be liked by their group

members in the first meeting expressed similar attitudes toward their second group. Overall, Study 1 yielded strong evidence for a generalized stereotype account of perceiver effects.

Study 2

In Study 2, we extended Study 1 in two respects. First, consistency in Study 1 might have partly been attributable to the fact that participants interacted with each other in very similar contexts across meetings. If participants had differential preferences for discussing moral dilemmas, their perceiver effects might have been colored by these preferences to some degree, thereby inflating consistency. In Study 2, contexts were varied between measurement occasions to rule out such inflation. Second, Study 1 yielded only weak support for trait- and item-specific perceiver effect consistency but this might have been due to the focus on interpersonal trait content. In Study 2, we applied an instrument that draws on the Five-Factor Model of personality and allows for a reassessment of the possibility of content-specific consistency within a more comprehensive trait framework. As in Study 1, Study 2 once again made fully rotated group assignments such that each group was composed of unique strangers at every measurement occasion.

Study 2 was conducted as part of a larger project on status hierarchies (Lawless DesJardins, 2016). The sample size ($n = 295$) was not planned for the present research but allows the detection of moderately sized correlations ($r = .25$) with very high statistical power ($1 - \beta = .99$) at a 5% alpha level. For brevity, we focus on the study's aspects directly relevant for the present work. A comprehensive description of all involved procedures and measures can be found in the primary publication from the larger project (Lawless DesJardins, 2016). Data and R-code are retrievable from https://osf.io/cxp3b/.

Participants and Procedure

Participants were recruited from two introductory psychology courses and received course credit for participation. Most participants were young adults (age: $M = 19.13$, $SD = 2.44$), most were White (69%), and 69% of them identified as female, 30% as male, and 1% did not identify their gender. Participants attended one of the 14 sessions which took place in a large classroom furnished with round tables. Participants were directed to sit at any table to read and sign the consent form and were then assigned to their first group. The assignment was random, with the contingency that all groups would have nonoverlapping membership (i.e., no two participants would interact with one another more than once). For each upcoming task, group assignments (i.e., table numbers) were presented on a projection screen.

After having found their designated seats, participants received verbal instruction for the current task. After having worked on the task for 10 minutes, they were asked to stop their activity and provide impressions about their group members on a rating sheet. Ratings concerned the Big Five personality traits, familiarity, attraction, status, and judgments related to the specific task (e.g., expertise). Here, we focus on ratings of the Big Five and attraction. However, we again used the familiarity judgments to verify that participants were unacquainted. On a 5-point scale asking "Have you met any of these people before today?," 90.3% of the dyads reported a 0 (*No, I have never met this person before today*) and 4.0% reported a 1 (*I have seen this person around but we have never spoken before*). Dyads who reported higher familiarity were excluded.

After the ratings were completed, participants were assigned to their next group, and the procedure was repeated up to 4 times. Participants completed tasks with (a) a broadly defined social goal, (b) a competitive goal, (c) a cooperative goal, and (d) a knowledge-based problem-solving goal. The social task was simply to get to know each other as well as possible. The competitive task was a leaderless group discussion in which participants took the role of representatives in an alumni association and had to advocate for giving as much prize money as possible to one nominee they had been randomly assigned to (adapted from John & Robins, 1994; Staw & Barsade, 1993). The cooperative task was a modified Lost on the Moon task (Robins & Beer, 2001) in which the group needs to prioritize a list of items that would help them after having crash landed in an expedition to the moon. The problem-solving task involved 30 verbal and quantitative problems drawn from practice Scholastic Assessment Test questions (Khan Academy, 2015). Groups were asked to solve as many of them as possible. Tasks 2 and 3 have been used in previous work on perceiver effects (Srivastava et al., 2010).

The order of tasks was constant within sessions but randomized across sessions. In some sessions, there were not enough participants to create four sets of nonoverlapping groups. As a result, not all participants completed all tasks ($n_{Task1} = 236$; $n_{Task2} = 278$; $n_{Task3} = 219$; $n_{Task4} = 278$) and occasionally, there were "leftover" groups with fewer than four members or with members who had already met in an earlier task. After discarding these groups, there were 51 groups of five and 189 groups of four.

Measures

Mutual personality perceptions were assessed via the 10-item Big Five Inventory (Rammstedt & John, 2007). This instrument covers each personality trait with one positively and one negatively keyed item (Table 3). We also assessed liking and meta-liking. Participants were instructed to use a 0 to 10 scale anchored at 0 (*disagree strongly*), 5 (*neither agree nor disagree*), and 10 (*agree strongly*).

As in Study 1, judgments on all personality items were driven by between-perceiver differences (35%–43% on average) but also by between-target and between-dyad differences (15%–20% and 42–46%, respectively; see

Table 3. Study 2 SRM Variance Components and Perceiver Effect Reliabilities by Item

Item (Big Five dimension)	Task 1 (social)		Task 2 (competitive)		Task 3 (cooperative)		Task 4 (problem-solving)	
	P/T/R	Rel$_P$	P/T/R	Rel$_P$	P/T/R	Rel$_P$	P/T/R	Rel$_P$
Is outgoing, sociable (E +)	.15/.55/.30	.59	.22/.43/.36	.63	.19/.47/.34	.61	.19/.42/.39	.57
Is reserved (E−)	.25/.40/.36	.67	.32/.33/.35	.72	.22/.41/.37	.63	.21/.34/.44	.58
Is generally trusting (A +)	.47/.11/.42	.77	.45/.12/.43	.75	.41/.09/.50	.70	.41/.08/.50	.70
Can be cold and aloof (A−)	.49/.12/.39	.79	.53/.09/.38	.79	.44/.12/.44	.74	.45/.15/.41	.76
Does a thorough job (C +)	.48/.09/.42	.77	.42/.12/.47	.72	.28/.18/.55	.59	.40/.20/.40	.74
Tends to be lazy (C−)	.50/.09/.41	.77	.48/.12/.40	.77	.54/.12/.34	.83	.50/.12/.37	.80
Is relaxed, handles stress well (ES +)	.40/.13/.46	.71	.44/.10/.46	.74	.35/.14/.51	.67	.39/.10/.50	.68
Gets nervous easily (ES−)	.30/.25/.44	.67	.45/.15/.40	.77	.33/.24/.44	.69	.32/.21/.47	.66
Has an active imagination (O +)	.40/.12/.48	.70	.48/.08/.44	.76	.28/.15/.57	.59	.38/.07/.54	.67
Has few artistic interests (O−)	.33/.14/.53	.64	.39/.11/.50	.70	.37/.11/.52	.67	.37/.10/.53	.66
I like this person (liking)	.47/.13/.40	.77	.50/.12/.38	.79	.48/.16/.36	.79	.47/.20/.33	.81
This person likes me (meta-liking)	.66/.03/.30	.86	.65/.09/.26	.87	.64/.06/.31	.86	.69/.08/.23	.90

Note. SRM = Social Relations Model; P/T/R = standardized SRM variances for the perceiver/target/relationship (incl. error) component; all coefficients are statistically different from zero ($p < .05$); Rel$_P$ = reliability of individual perceiver effect scores; E = extraversion; A = agreeableness; C = conscientiousness; ES = emotional stability; O = openness to experience.
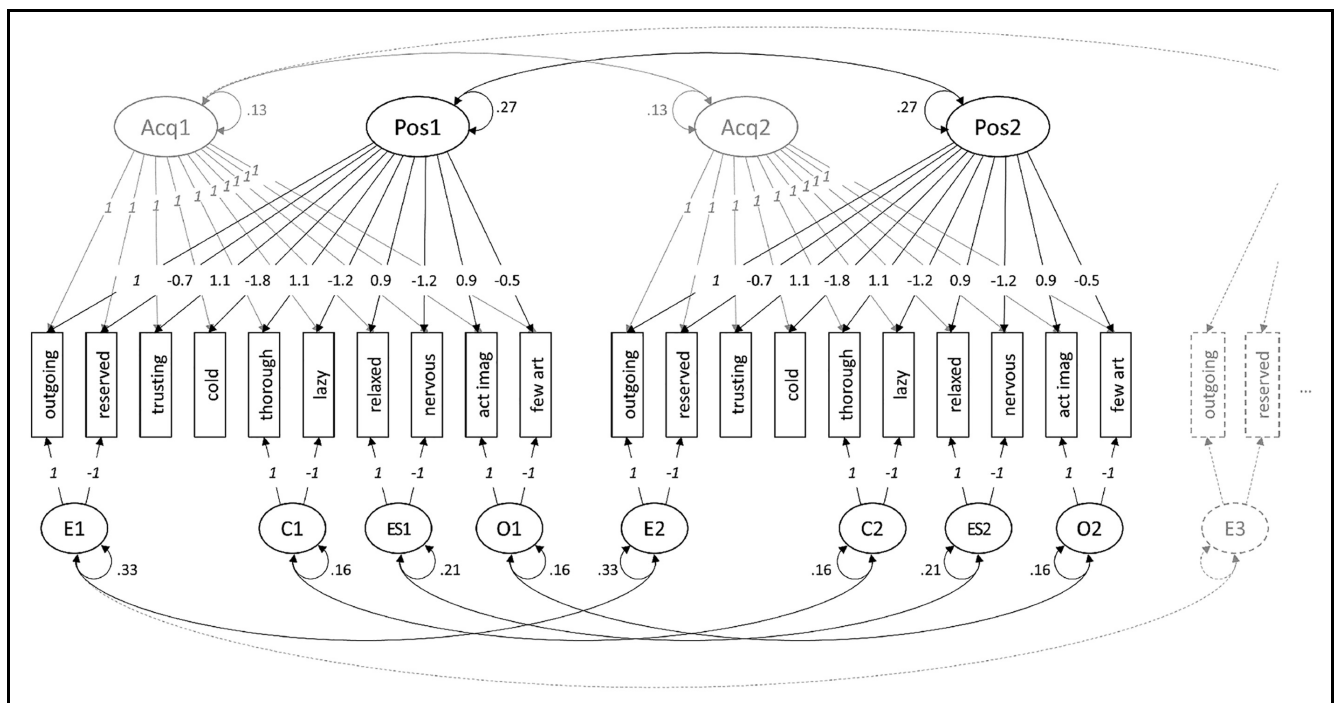

Figure 2. Study 2 Structural Equation Model With Perceiver Effects for Each Trait as Observed Variables and Positivity (Pos), Acquiescence (Acq), Extraversion (E), Conscientiousness (C), Emotional Stability (ES), and Openness to Experience (O) as Latent Variables *Note.* Loadings in italics were fixed. All parameters were constrained to be equal across timepoints (strict measurement invariance). For clarity, only two measurement occasions are displayed and residual (co)variances are omitted.

Table 3). For liking and meta-liking in particular, perceiver differences were even more pronounced (up to 69% perceiver variance). All traits yielded perceiver variances above 10% and acceptable reliabilities for individual perceiver effect scores. The larger portion of perceiver variance as compared with Study 1 may be explained by the use of a much wider scale (0–10 as compared with 1–5).

Results and Discussion

To estimate the consistency of the different perceiver effect components, we took a similar modeling approach as in Study 1 (Figure 2). Specifically, we specified positivity factors (free loadings), acquiescence factors (fixed loadings), and trait-specific factors for extraversion, conscientiousness, emotional stability, and openness with fixed loadings

of 1 and −1 for the respective indicators. Paralleling the approach of Study 1 where no trait-specific factor was modeled for communion due to its strongly evaluative nature, no specific factor for agreeableness was modeled in Study 2. For one of the four timepoints (problem-solving context), we also omitted the trait-specific factor for openness after observing that modeling this factor produced an improper solution (i.e., a variance estimate < 0). Consistency correlations for latent variables and item-residuals were constrained to be equal across all pairs of tasks. As for Study 1, alternative specifications are presented in Appendix B and a zero-order correlation matrix of all observed variables can be retrieved online (https://osf.io/cxp3b/).

Fit-indices were $\chi^2(df) = 1{,}094(729)$, $CFI = .829$, $RMSEA = .041$, $SRMR = .082$. As before, given the level of model-fit for less parsimonious models (Appendix B) and in previous studies (Rau, Carlson, et al., 2021; Srivastava et al., 2010), this level of fit appeared acceptable. As can be seen in Figure 2, the substantive meaning of the positivity factors was supported by a pattern of positive loadings for desirable personality attributes (e.g., *trusting, thorough*) and negative loadings for undesirable attributes (e.g., *cold, lazy*).

How consistent were people's judgment tendencies across groups and tasks? Similar to Study 1, the major portion of variance both in positivity and acquiescence was consistent across contexts, $r_{Positivity} = .53$ (95% CI = [.44, .62]); $r_{Acquiescence} = .58$ (95% CI = [.46, .69]). For extraversion and emotional stability, the level of consistency was more moderate, specifically $r = .30$ (95% CI = [.13, .48]) and $r = .35$ (95% CI = [.03, .68]), respectively. For openness, there was a large point estimate but this estimate was associated with a quite wide CI, $r = .57$ (95% CI = [.12, 1.00]), which warrants caution in interpreting this effect as particularly large. Finally, no evidence for conscientiousness-specific consistency was found, $r = .09$ (95% CI = [−.38, .55]), and on the level of item residuals, consistency was weak at best (see Appendix C) which in part goes back to measurement error in the indicator variables (cf. Table 3). For liking and meta-liking, the level of consistency was again substantial and comparable to the consistency of positivity (after correcting for measurement error), $r_{liking} = .42$ (95% CI = [.25, .57]); $r_{meta-liking} = .56$ (95% CI = [.40, .70]). In sum, although Study 2 used different personality items and involved variation in contexts, findings were quite similar as in Study 1 such that there was clear evidence for sizable consistency in positivity, acquiescence, and (meta-)liking and less robust evidence for moderate consistency in trait- and item-specific perceiver effects.

General Discussion

Across the first two studies that allowed for a strict test, we have found strong support for a generalized stereotype account of perceiver effects. Specifically, we found consistency correlations for overall positivity in personality judgments and for liking and meta-liking that were quite substantial in size, indicating that people who tend to see the best (or the worst) in others in one situation also tend to see the best (or worst) in others in entirely different situations, which can occur at a different point in time and involve different interaction partners and contexts. This is consistent with the interpretation that people have a relatively fixed set of learned assumptions about "the average other" which shapes how they perceive their social environment in a given situation (Bowlby, 1988; Erikson, 1959, 1968) and it emphasizes the role of perceiver effects as an important individual difference variable (Kenny, 1994, 2019).

We also observed an impressive amount of consistency in acquiescence bias in both studies which suggests that people have rather stable preferences for using the high or low ends of rating scales regardless of their content. This is in line with previous research (Crow & Hammond, 1957; Wetzel et al., 2016) and highlights the danger of response styles inflating stability estimates of rating data. Taken together, effect sizes for positivity, liking, meta-liking, and acquiescence all hovered around $r = .50$ suggesting that roughly half of the variance in these perceiver effects was shared across measurement occasions. Of note, this was true for both studies which suggests that unique design features (i.e., temporal gaps in Study 1 and cross-context variation in Study 2) diminished consistency to comparable extents.

Beyond global tendencies, we found situationally consistent trait-specific perceiver effects for some content domains (i.e., agency, extraversion, emotional stability, openness), whereas for other domains, perceiver effects were either very specific to situations (conscientiousness) or they did not emerge as relevant factors above and beyond global positivity to begin with (communion, agreeableness). These results suggest that consistent trait-specific perceiver effects might only exist for personality content that is sufficiently neutral in terms of evaluativeness. In line with this, previous work has found perceiver effects for agentic but not for communal attributes to be predictive of interpersonal behavior and social outcomes (Rau et al., 2019, 2020).

Finally, we found evidence of cross-situational consistency at the level of item-residuals for some but not for all items. This might indicate that perceivers have characteristic ways of interpreting specific terms (i.e., idiosyncratic "meaning systems," Kenny, 1994). However, only approximately one third of the variance both in trait-specific and in item-specific perceiver effects was consistent across situations which contrasts the much higher levels of consistency we found for global dimensions such as positivity or liking.

Outlook

As argued in the introduction, the strongest piece of evidence for a generalized stereotype interpretation of perceiver effects would have come from a study that varies time, groups, and contexts at once. Whereas the studies presented here varied only two out of the three aspects each, we are

confident that, by combining complementary designs, our results provide strong support for the idea that persistent generalized stereotypes affect first impressions.

This finding has vital implications for applied settings. For instance, personnel selectors who observe applicants' tendencies to see the best (or worst) in their competitors in an assessment center will be given a fairly good idea of how these applicants would view their future colleagues. Similarly, the judgment tendencies of psychotherapeutic patients that manifest in a group therapy may be a fruitful point of intervention for therapists, given that the same tendencies are likely to color patients' perceptions in everyday encounters outside the therapeutic setting. Future research will need to explore the potential of such applications.

ORCID iD

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

Notes

1. This analysis involves a correction for the fact that each perceiver rated a slightly different set of targets.

References

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

Bowlby, J. (1988). *A secure base*. Basic Books.

Crow, J., & Hammond, K. R. (1957). The generality of accuracy and response sets in interpersonal perception. *Journal of Abnormal Psychology*, *54*(3), 384–390. https://doi.org/10.1037/h0042990

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, *22*(3), 541–562. https://doi.org/10.1037/met0000083

Erikson, E. H. (1959). *Identity and the life cycle*. International Universities Press.

Erikson, E. H. (1968). *Identity, youth and crisis*. W. W. Norton.

Heynicke, M., Rau, R., Leising, D., Wessels, N., & Wiedenroth, A. (2021). Perceiver effects in person perception reflect acquiescence, positivity, and trait-specific content: Evidence from a large-scale replication study. *Social Psychological and Personality Science*. Avance online publication. https://doi.org/10.1177/19485506211039101

Hogan, R., Jones, W. H., & Cheek, J. M. (1985). Socioanalytic theory: An alternative to armadillo psychology. In B. R. Schlenker (Ed.), *The self and social life* (pp. 175–198). McGraw-Hill.

Jacobs, I., & Scholl, W. (2005). Interpersonale Adjektivliste (IAL). *Diagnostica*, *51*(3), 145–155. https://doi.org/10.1026/0012-1924.51.3.145

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.

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology*, *66*(1), 206–219. https://doi.org/10.1037/0022-3514.66.1.206

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

Kenny D. A. (2019). *Interpersonal perception: The foundation of social relationships* (2nd ed.). Guilford Press.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin*, *102*(3), 390–402. https://doi.org/10.1037/0033-2909.102.3.390

Kenny, D. A., Gomes, S. B., & Kowal, C. (2015). The intergroup social relations model: ISRM. *Group Dynamics: Theory, Research, and Practice*, *19*(3), 152–165. https://doi.org/10.1037/gdn0000028

Khan Academy. (2015). *Scholastic assessment test [Online practice test]*. https://www.khanacademy.org/test-prep/sat

Lawless DesJardins, N. M. (2016). *Social status across contexts* [Doctoral dissertation, University of Oregon].

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

Marcus, D. K., & Miller, R. S. (2003). Sex differences in judgments of physical attractiveness: A social relations analysis. *Personality and Social Psychology Bulletin*, *29*(3), 325–335. https://doi.org/10.1177/0146167202250193

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

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., Carlson, E. N., Dufner, M., Geukes, K., Kraft, L., Krause, S., Nikoleizig, L., Nestler, S., & Back, M. D. (in press). Positive peer perceptions over time: The role of personality and likability in predicting the development of perceiver effects. *Journal of Personality and Social Psychology*.

Rau, R., Nestler, S., Geukes, K., Back, M. D., & Dufner, M. (2019). Can other-derogation be beneficial? Seeing others as

low in agency can lead to an agentic reputation in newly formed face-to-face groups. *Journal of Personality and Social Psychology*, *117*(1), 201–227. https://doi.org/10.1037/pspp0000250

Rau, R., Nestler, W., Dufner, M., & Nestler, S. (2021). Seeing the best or worst in others: A measure of generalized other-perceptions. *Assessment*, *28*(8), 1897–1914. https://doi.org/10.1177/1073191120905015

Rau, R., Thielmann, I., Breil, S. M., Geukes, K., Krause, S., Nikoleizig, L., Back, M. D., & Nestler, S. (2020). Do perceiver effects in interpersonal perception predict cooperation in social dilemmas? *Collabra: Psychology*, *6*(1), 35. https://doi.org/10.1525/collabra.332

R Development Core Team. (2008). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing.

Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology*, *80*(2), 340–352. https://doi.org/10.1037/0022-3514.80.2.340

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Schönbrodt, F. D., Back, M. D., & Schmukle, S. C. (2012). Tripler: An R package for social relations analyses based on round-robin designs. *Behavior Research Methods*, *44*(2), 455–470. https://doi.org/10.3758/s13428-011-0150-4

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 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

Staw, B. M., & Barsade, S. G. (1993). Affect and managerial performance: A test of the sadder-but-wiser vs. happier-and-smarter hypotheses. *Administrative Science Quarterly*, *38*(2), 304–331. https://doi.org/10.2307/2393415

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

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

Appendix A

In the first group meeting of Study 1, participants were given the following text:

The married couple Marianne and Reinhard live together with their two children in a little village in Hesse. Both Marianne and Reinhard have been unemployed for a longer period of time, so that the family doesn't have a lot of money. A couple of weeks ago, the doctors diagnosed Reinhard with a rare illness that causes a lot of pain. Unfortunately, the illness has not yet been thoroughly researched and the only medicine that could ease Reinhard's pain is not available in Germany so that the family would have to buy it for a high price abroad. Unfortunately, the family cannot afford the medicine. Marianne is desperate and can't stand the situation any longer. Because of his pain, Reinhard has become a nursing case and the family climate also suffers because of the situation. To distract herself, Marianne has been seeing her friend Karl-Heinz for some time. He enjoys being close to Marianne because he has been secretly in love with her for a long time. To help Marianne, he decides to make her an offer. He would buy the medicine if she spent the night with him. Outraged, Marianne refuses. Reinhard's brother Gordon has recently made a lot of money by gambling on the stock exchange. He plans to fulfill his dream of owning his own restaurant. In spite of the difficult relationship between himself and his brother, Reinhard decides to ask him for help. Gordon, who is about to invest his money in the restaurant, fears that he would never see his money again, so he refuses after having thought about it for a long time. Marianne, who can't stand the situation any longer decides to accept Karl-Heinz's offer. Directly after the night spent with Karl-Heinz, Marianne immediately regrets her decision and tells Reinhard everything. Reinhard feels betrayed and leaves the family. Feeling hurt, Marianne calls her friend Tom, who is outraged by Reinhard's behavior and feels pity for Marianne. Tom drives to Reinhard to insult and beat him.

After reading the text, groups were asked to discuss the behavior of each person and to agree on a "moral" order for the characters from good to bad. The same procedure was repeated in the second group meeting with the following text:

Peter and Astrid, being in their late thirties, would like to have a baby. Until now, they were hesitant to conceive a baby because Peter's brother passed away after a long time of suffering due to amyotrophia. It is possible for Peter to hold the same gene his brother held and with a chance of 50%, he could transfer this gene to a descendant. Therefore, Peter would like to use preimplantation diagnostics. This means an artificial insemination of Astrid's ovocytes whereupon only healthy embryos are used. Using this technique, all other fertilized embryos would be destroyed. Astrid can't stand the thought that only a few embryos survive. Peter on the other hand would accept that some of the embryos will not be born. The problems about longing for a child result in discussions between Astrid and Peter. Their fights threaten their relationship. In her desperation, Astrid turns to her high school friend Thorsten because she wants to talk to an unbiased person. Meeting Thorsten helps Astrid a lot because Thorsten patiently listens to her and, during this difficult situation, he gives her the feeling of being understood. Thorsten decides to take advantage of Astrid's difficult situation by seducing her after having some glasses of wine.

After spending the night with Thorsten, Astrid finds herself being pregnant, expecting Thorsten's child. Astrid is shocked about those serious consequences of her indiscretion. She visits her friend Birgit and cries. Birgit has known Astrid and Peter for a long time, she already knows about their difficulty. Knowing Birgit as someone who tends to keep a clear head, Astrid asks Birgit to give her advice. She wants to know what

Birgit would do if she was in Astrid's shoes. Birgit gives the advice to pretend that Peter was the father of Thorsten's unborn child. This way, Astrid could save her relationship with Peter and have a healthy child. After carefully thinking about Birgit's advice and feeling guilty, Astrid decides to lie to Peter. She tells him that she forgot to take her pill and became pregnant. Peter is terrified about not knowing whether the child holds the gene. He wants Astrid to take an early mniocentesis, which could be dangerous for the unborn child. Astrid is desperate because she cannot admit her infidelity to Peter, but she doesn't want to jeopardize the life of the unborn child either. She agrees to take the amniocentesis because Peter pushes her to do so. Meanwhile Birgit feels guilty about giving Astrid the advice to lie to Peter. She learns about the upcoming amniocentesis. She knows that the amniocentesis will be dangerous for the unborn child. Immediately, she drives to the hospital where the amniocentesis is about to take place. She tries to stop the procedure by telling Peter the whole story. When Peter learns that Thorsten spent the night with his wife, he becomes very angry and drives to Thorsten's place. Finding Thorsten in front of his house, Peter vents all of his wrath, disappointment and desperation on Thorsten. Even when Thorsten lies on the ground, Peter keeps beating him.

Appendix B

Table B1. Model-Fit for Increasingly Parsimonious Specifications by Study

Study	Model	Number of free parameters	Fit	
			CFI; RMSEA; SRMR	BIC
1	1: All loadings free (except marker variables)	49	Does not converge	
	2: Trait-specific factor loadings partially fixed	43	.944; .060; .079	5140,5
	3: Communion factor removed	36	.913; .073; .088	5140,4
	4: Equality constraint on latent variances	**33**	**.900; .077; .108**	**5140,1**
	5: Equality constraint of positivity factor loadings	26	.676; .134; .165	5310,8
2	1: All loadings free (except marker variables)	139	Does not converge	
	2: Trait-specific factor loadings fixed	134	.840; .042; .080	29582,5
	3: Agreeableness factor removed	128	.838; .043; .080	29560,8
	4: Equality constraint on latent variances	111	.836; .042; .080	29485,0
	5: Equality constraint on consistency correlations	**41**	**.829; .041; .082**	**29174,8**
	6: Equality constraint of positivity factor loadings	32	.795; .045; .092	29208,5

Note. Models decrease in complexity from top to bottom. All models are nested such that any model's constraints were added on top of the constraints of the previous model. Model selection was based on the Bayesian Information Criterion (BIC; Schwarz, 1978). Selected models are printed in bold. Whereas Models 1 and 2 reflect a classic bifactor specification, Models 3 through 6 reflect an S − 1 specification recommended to stabilize the estimates for bifactor models (Eid et al., 2017). CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Appendix C

Table C1. Study 1 Consistency Correlations for Item-Level Residuals

Item	r [95% CI]
Assertive	.08 [−.22, .38]
Cynical	.30 [.09, .51]
Hostile	.06 [−.21, .32]
Unsociable	.43 [−.08, .94]
Shy	.18 [−.08, .44]
Compliant	.40 [.24, .57]
Sensitive	.16 [−.07, .38]
Outgoing	.31 [.05, .56]

Note. CI = confidence interval.

Table C2. Study 2 Consistency Correlations for Item-Level Residuals

Item	r [95% CI]
Outgoing	.11 [−.13, .34]
Reserved	.17 [.08, .26]
Trusting	.23 [.15, .31]
Cold	.25 [.15, .34]
Thorough	.22 [.09, .35]
Lazy	.20 [.11, .29]
Relaxed	.07 [−.03, .18]
Nervous	.15 [.05, .25]
Active imagination	.16 [.07, .25]
Few artistic interests	.18 [.09, .26]

Note. CI = confidence interval.

Author Biographies

Richard Rau works as a postdoc at the Department of Psychology at University of Münster, Germany. He studies interpersonal perception and teaches personality psychology, assessment, and statistics.

Nicole M. Lawless DesJardins earned her PhD at the University of Oregon studying status hierarchies in small groups. She now works as a data scientist.

Lisa Maria Niemeyer earned her PhD at the University of Münster studying maladaptive personality traits in forensic contexts. She now works as a forensic psychologist.

Prof. Mitja D. Back is head of the work group for Personality and Assessment at University of Münster, Germany. He studies the interplay of personality and social relationships.

Prof. Sanjay Srivastava is head of the Personality and Social Dynamics Lab at University of Oregon, Eugene, USA. He studies behavioral expressions of personality, impression formation, and personality development.

Prof. Steffen Nestler is head of the work group for Statistics and Psychological Methods at University of Münster, Germany. He develops methods for the study of interpersonal perception and intraindividual variability.