2021

The Judgment of Personality: An Overview of Current Empirical Research Findings

Tera D. Letzring

Nora Murphy

Jüri Allik

Andrew Beer

Johannes Zimmermann

See next page for additional authors

Authors
Tera D. Letzring, Nora Murphy, Jüri Allik, Andrew Beer, Johannes Zimmermann, and Daniel Leising

PERSONALITY Science

PsychOpen GOLD

Theory: State of the Art Review [Invited Paper]

Check for updates

The Judgment of Personality: An Overview of Current Empirical Research Findings

Tera D. Letzring[1§], Nora A. Murphy[2], Jüri Allik[3], Andrew Beer[4], Johannes Zimmermann[5], Daniel Leising[6§]

[1] *Department of Psychology, Idaho State University, Pocatello, ID, USA.* **[2]** *Department of Psychology, Loyola Marymount University, Los Angeles, CA, USA.* **[3]** *Department of Social Sciences, Institute of Psychology, University of Tartu, Tartu, Estonia.* **[4]** *Department of Psychology, University of South Carolina Upstate, Spartanburg, SC, USA.* **[5]** *Department of Psychology, University of Kassel, Kassel, Germany.* **[6]** *Faculty of Psychology, Technische Universität Dresden, Dresden, Germany.*

[§]*These authors contributed equally to this work.*

Personality Science, 2021, Vol. 2, Article e6043, https://doi.org/10.5964/ps.6043

Received: 2021-02-01 • **Accepted:** 2021-08-30 • **Published (VoR):** 2021-10-20

Handling Editor: John F. Rauthmann, Bielefeld University, Bielefeld, Germany

Reviewing: *Round 1* - Simon M. Breil; David A. Kenny. Open reviews are available [see Index of Supplementary Materials]

Corresponding Author: Tera D. Letzring, Idaho State University, 921 S. 8th Ave., Department of Psychology, Box 8112, Pocatello, ID 83209, USA. E-mail: teraletzring@isu.edu

Abstract

This article presents an overview of the current state of knowledge in personality judgment research. We discuss accuracy and bias in personality judgments, including types of inter-rater agreement and elements of criteria used to determine levels of agreement and accuracy. We then address 1) the words and phrases that people use to describe one another and themselves, 2) research investigating judgments of targets by perceivers per trait, and 3) research investigating judgments of targets by perceivers on profiles or sets of traits. We also provide 4) an outlook regarding important research questions that remain unanswered in this field.

Keywords

personality judgment, judgment accuracy, person perception, bias, inter-rater agreement, self-other agreement, consensus

Relevance Statement

This article serves as an overview of the current state of knowledge in personality judgment research. Core findings that are agreed on by the authors - who are from diverse labs and perspectives and use diverse methods - are clearly identified. Several unanswered questions are also described.

Key Insights

- Summarizes the most robust empirical evidence in personality judgment research.
- Personality judgments can be accurate but also may contain biases.
- Descriptors and language shape subsequent personality judgments.
- Variable-centered and profile approaches are used to study judgment accuracy.
- Identifies key unresolved issues for future research to address.

Why Is It Important to Study Personality Judgment, and What Can We Learn?

The history of personality judgment research is nearly as long as the history of formal psychological research in general, and for good reason: personality judgments form quickly and naturally and have important consequences for both the person making the judgments and the person being judged. What we think about the personality of others plays an important role in how we behave toward others, including whether we attempt to communicate or establish a relationship, offer them a job, or vote for them in an election. As these are some of the most important domains in human existence (love and companionship, work and productivity), acquiring knowledge about these processes constitutes a potentially vital contribution to psychological science in both theory and practice.

As personality and social psychologists, we strive to understand how judgments of personality come about and the factors that influence these judgments. Personality judgment research seeks to answer questions such as: Which personality judgments are typically correlated and can be combined into overarching descriptive dimensions (i.e., what is the *structure* of these judgments)? Under what conditions do judgments of targets by others become more similar to one another (*consensus*)? How well do the targets' judgments of themselves converge with judgments by other people (*self-other agreement*)? Do personality judgments agree with criteria reflecting what targets are really like (*accuracy*)?

In the present paper, we provide an overview of empirical research on personality judgment, with a focus on current findings of particularly high importance and consistency in relation to the preceding questions[1]. Due to space constraints, we have limited our citations to review articles when possible, or to the first article that supported a

specific finding. We review terminology used within the field, what researchers generally agree upon, and where there are disagreements or questions remaining to be answered.

Accuracy and Bias in Personality Judgments

A complete review of theoretical deliberations within personality judgment research is beyond the scope of this paper. Instead, we will briefly reiterate some core theoretical tenets that we assume are shared by most researchers in the field, and otherwise concentrate on the most robust empirical evidence. Popular theories about interpersonal judgment processes include the following: Brunswik's (1956) lens model (Osterholz et al., 2021), the Social Relations Model (SRM) and PERSON Model (Kenny, 2004, 2020), the Realistic Accuracy Model (RAM; Letzring & Funder, 2021), the Social Accuracy Model (SAM; Biesanz, 2021), and the Truth and Bias model (TAB; West & Kenny, 2011). There is considerable convergence among these models, although an overarching integration is still lacking. For a review of interpersonal perception theories and models, see Biesanz (2018).

When perceivers (or judges) evaluate the personality of targets using words or phrases, their descriptions may reflect some "substantive" or "true" characteristics of the targets. For example, a target who uses more words per minute than the average target may be judged as "talkative." To study personality judgment accuracy, perceiver judgments are compared with some standard, or *accuracy criterion*, that is assumed to reflect an approximation of "true" characteristics of targets, using some variant of correlational analyses.

Several accuracy criteria are used in judgment research, including targets' descriptions of themselves, descriptions from acquaintances of the targets, and codings of targets' behaviors that are assumed to relate to the traits of interest. When ratings from different perceivers, including the self, are compared to each other, the level of correspondence can be referred to as *inter-rater agreement*, and researchers use a variety of terms to refer to different combinations of types of perceivers and criteria (Letzring & Funder, 2018). *Consensus* applies to correspondence between judgments by two or more perceivers other than the target. These perceivers could all be strangers or all be acquaintances (also called informants), or strangers' judgments could be compared to acquaintances' judgments. *Self-other agreement* applies to correspondence between judgments by others and targets' self-descriptions. Note that ratings by informants can be treated either as the judgments whose validity is to be determined, or as a criterion variable. There are several good reasons for using acquaintance-ratings as accuracy criteria. For example, close acquaintances of targets may reasonably be assumed to base

1) See Back and Nestler (2016) for a review that focuses on zero and minimal acquaintance and a Brunswikian lens model approach, and Letzring and Funder (2018) for a review that includes an overview of the history in this area and conceptual models.

judgments on a broad range of experiences that they had with the targets. Judgments by acquaintances also have the advantage of aggregating ratings across multiple acquaintances, which will increase the reliability of measurement. On the other hand, the targets themselves presumably know the truth best regarding certain variables to which they have "privileged access" (e.g., their own feelings; Vazire, 2010, p. 283). In many studies, these types of inter-rater agreement are interpreted in terms of accuracy, on the assumption that the different perceivers agree because they all observed or inferred something "real" that sets a given target's personality apart from other targets.

Sometimes, *composites* of several variables (e.g., self and other-ratings) are used as accuracy criteria, based on the assumption that the outcome will reflect the variance that is shared among all sources of information, and thus most likely reflect something real (the Realistic Accuracy Model approach; Letzring & Funder, 2021). Another strategy is to use accuracy criteria that are (mostly) unfiltered by human perception, such as behavior tracking using mobile sensing and digital footprints, recordings of ambient sound (e.g., Tackman et al., 2020), or direct observation of behaviors in a laboratory setting. However, one downside is that generating these criteria requires much time, effort, and resources. Another consideration is that the conceptual relationship between judgment and criterion variables will have to be justified. For example, if judgments of extraversion correlate with speaking time, should that be interpreted in terms of (partial) accuracy? Also, the decision as to which of the many available ("hard") variables are to be deemed important will still have to be made by human judges (i.e., researchers).

Personality judgments can also be influenced by variables other than the "substantive" characteristics of the target, and such influences are often called perceiver *biases*. For example, some perceivers tend to endorse all descriptor items, irrespective of content (known as *acquiescence*, see Section 2.4) and a perceiver's fondness for a target plays a role in what terms the perceiver will use to describe the target. Inter-rater agreement or accuracy also may be driven by *shared biases* such as stereotypes (e.g., perceivers sharing certain racial stereotypes may agree in their judgments of targets of different races, irrespective of targets' actual behaviors). While a "kernel of truth" may exist in some stereotypes and the variance associated with stereotypes can thus contribute to accuracy (Jussim et al., 2009; Kenny, 2004), an appropriate handling of the accuracy issue requires careful consideration regarding alternative, substance-unrelated reasons for measures to correlate with one another (e.g., Podsakoff et al., 2003). In summary, it is important to consider that personality judgments contain systematic variance between targets, between perceivers, between individual perceiver-target dyads (after controlling for differences between perceivers and between targets), and error (Kenny, 2020). So, personality judgments are not only informative regarding differences between targets - they also tell us something about perceivers and the specifics of each perceiver-target dyad.

What Do Personality Judgment Researchers Know and Agree On?

After considerable discussion, we have chosen topics that we agree are especially important in understanding personality judgments and judgment accuracy. Despite our attempt to aim for a collection of robust and representative findings, our selections are still somewhat subjective. In the following sections, we first focus on descriptor terms and phrases related to personality that people may use to describe targets (Section 1). These descriptors are used in most personality research and thus relevant to personality judgments. Then, we address issues of inter-rater agreement and accuracy in personality judgments, describing two commonly-used strategies: the *variable-centered* or *trait-wise* approach (Section 2) and the *profile* or *person-centered* approach (Section 3).

1. Core Findings Regarding Descriptors of Personality

1.1. The natural person-descriptive language contains hundreds of terms that people use to characterize personality traits (Anderson, 1968), such as "outspoken" or "impulsive." The *psycholexical hypothesis* (Allport & Odbert, 1936) posits that these terms exist because they denote personality characteristics that people found important to communicate about throughout time. Understanding the properties of such terms, and ways in which perceivers use them, is relevant to personality judgments because they constitute the material with which data are collected (e.g., questionnaires, rating items, etc.).

1.2. For most person-descriptive terms, the optimal level is not found at one of the poles of the response continuum, but at a point that is less extreme. Thus, many person-descriptions incorporate the idea that one may have "too much" of a good trait, or "too little" of a bad trait (Jung & Kenny, 2005).

1.3. Research shows that personality-descriptive terms differ in a variety of ways that may impact judgments, including how positive or negative a light they shed on a target (*social desirability*; Edwards, 1953), the stability of substantive characteristics to which they refer (*traitness*), how easily the substantive differences between targets that they refer to may be observed from outside (*observability* or *visibility*), and how common those characteristics are in the population (*base rate*). Terms also differ in how broad a range of substantive characteristics they refer to (*abstractness*), and how important it is for others to know that someone has been described this particular way (*importance*). All of these term properties can be reliably rated by a small number of (~10-15) raters.

1.4. The distributions of most of these term-properties are unimodal and centrally-peaked, with one clear exception: social desirability ratings have a bimodal distribution and clearly imply a positive or negative evaluation of a target. There seem to be more negative descriptors than positive ones (Anderson, 1968).

1.5. The rated social desirability of a term is also a strong predictor of how its use will depend on attitude differences, such as liking, between perceiver-target dyads. Perceivers

rate liked targets higher on positive terms, and disliked targets higher on negative terms, and they do so in at least partial independence of the targets' actual characteristics (Bäckström et al., 2009; Zimmermann et al., 2018).

1.6. The covariance structure of personality judgments may be captured at different levels of abstraction. At the highest level, a strong, evaluative ("general") factor distinguishes more positive from more negative descriptions (Biderman et al., 2018). This factor is most appropriately interpreted as reflecting the perceivers' attitudes toward the targets (which may be shared among perceivers to varying degrees). In self-ratings, these attitudes largely equal the targets' self-esteem (Anusic et al., 2009). The general factor aligns closely with Neuroticism, Depressivity, and other constructs saturated with self-evaluation. Beyond that, research has yielded five (John & Srivastava, 1999) or six (Ashton et al., 2014) factors that are more content-specific. These factors have been named Conscientiousness, Agreeableness, Neuroticism or Emotional Stability, Openness to Experience, Extraversion, and Honesty/Humility. However, the universality of these personality judgment factors across cultures is still the subject of considerable debate, with many proponents strongly advocating it, and others strongly opposing it (cf. Allik, Realo, Mõttus, Borkenau, et al., 2010; Thalmayer et al., 2020). Extraversion and Agreeableness largely capture the same item covariance that is also captured by the "big" two interpersonal dimensions: Dominance (or Status, Power, Agency, or Competence) and Affiliation (or Love, Warmth, or Communion; Wiggins, 1991). At an even lower level of abstraction, individual *aspects*, *facets*, and *nuances* of the broader personality factors may be distinguished from one another (John & Srivastava, 1999).

2. Core Findings Using the Variable-Centered Approach

The variable-centered approach provides information about how a specific trait may be perceived and is featured in several interpersonal perception models (e.g., Brunswik, 1956; Kenny, 2004). In this approach, several targets are rated by perceivers, either on a single term or phrase, or using several items from the same domain that are then aggregated. Importantly, in this approach agreement or accuracy is determined for individual variables (i.e., items or scale scores). This is useful when researchers are interested in studying judgments on a particular trait. It is also the only feasible approach in many cases where accuracy criteria other than (aggregated) judgments (e.g., visible behavior or objective test results) are used, because measuring such variables for more than just a few traits likely exceeds the resources available to most labs.

2.1. Scale scores are more reliable measures than scores of individual items, as long as the scale's items are intercorrelated. This higher reliability also makes it more likely for scales, rather than individual items, to correlate with other variables, and is probably the main reason why the use of personality scales is very common in contemporary personality psychology. However, because aggregation will accentuate the items' common variance at the cost of their unique variances, scale scores may sometimes be highly

reliable measures of things other than the targets' personalities (e.g., perceivers' attitudes toward the targets; Anusic et al., 2009; Leising et al., 2021).

2.2. Scales seem to be more strongly correlated with one another in other-ratings than in self-ratings, especially at lower levels of acquaintance (e.g., Beer & Watson, 2008a). This is likely due to attitudes contributing more to the overall variance in other-ratings (Leising et al., 2021), and/or to self-rating factors capturing more actual between-target variation in more internal experiences such as thoughts and feelings.

2.3. Some relatively small disparities in mean levels exist between judgments of the self versus others. For example, people rate themselves as more neurotic and less conscientious compared to external observer ratings, and these differences are cross-culturally replicable (Allik, Realo, Mõttus, Borkenau, et al., 2010).

2.4. Perceivers differ from one another in how positively they judge the average target, in what trait levels they attribute to the average target on specific traits, and in acquiescence (Rau et al., 2021). Estimates of the total influence of different types of perceiver variance on personality judgments often range around 30% (Kenny, 2020; Rau et al., 2021). However, when studies do not permit interaction between perceivers and targets, estimates can be considerably lower (e.g., Heynicke et al., 2021). One reason for this could be that, in studies with previous interactions, the variation between perceivers' judgment styles may be increased by variation in the average target's behavior toward particular perceivers (e.g., one perceiver actually being treated more nicely by everyone than another perceiver).

2.5. Perceivers differ systematically from one another in how variable their responses are across items, regardless of content (Baird et al., 2017). This is important because it may lead to erroneous conclusions about variability in targets (e.g., in a person's behavior) when, in reality, this variability is content-independent and rooted entirely in the perceiver. Notably, this individual difference is relatively stable over time (Wetzel et al., 2016).

2.6. People sometimes judge other people similarly to how they judge themselves (*assumed similarity).* The average degree of assumed similarity is moderately sized and varies across traits; meta-analytic results show the strongest and most consistent correlations for honesty-humility (r = .48), followed by openness (r = .23-.35) and agreeableness (r = .11-.25; Thielmann et al., 2020, Table 1). Assumed similarity also tends to be stronger for traits that are less visible and among people who have closer relationships (Kenny, 2020; Kenny & West, 2010).

2.7. Both consensus and self-other agreement are consistently significantly different from zero, but are often not very large. When not correcting for unreliability, consensus between two individual raters (on average, across traits) is ~r = .30-.40, whereas self-other agreement tends to be lower (~r = .20-.30; Connelly & Ones, 2010; Hall et al., 2008; Kenny & West, 2010). The discrepancy is usually explained in terms of the types of information (feelings and thoughts vs. visible behavior) that self- versus other-raters

have access to, and that are completely shared in the case of consensus but only partly shared in the case of self-other agreement (Kenny, 2020; Vazire, 2010).

2.8. Research has also looked into the extent to which people are aware of how others see them (*meta-accuracy*) and whether they notice the differences between others' views of them and their own self-perceptions (*meta-insight*; Carlson et al., 2011). Both types of agreement between judgments have small to moderate effect sizes, depending on the trait.

2.9. Levels of self-other agreement and accuracy vary across traits (Krzyzaniak & Letzring, 2021), which supports the notion that some traits tend to be judged more accurately than others. Specifically, extraversion is most consistently judged with moderately high levels of self-other agreement and accuracy, and neuroticism is usually associated with the lowest levels of self-other agreement and accuracy (Beer & Watson, 2008b). These differences are especially robust when perceivers are not well-acquainted with targets and are negligible or even absent when perceivers are well-acquainted (Connelly & Ones, 2010). Additionally, there is evidence that variance is an important predictor of self-other agreement such that traits with more variance tend to have higher agreement (Allik, Realo, Mõttus, Esko, et al., 2010).

2.10. There is substantial "cross-situational consistency" of personality judgments. When statistically controlling for "non-shared meaning" (i.e., between-perceiver differences in how the same terms are used to describe the same behaviors; Kenny, 2020), judgments by different observers who observed the same targets for a few minutes correlate ~r = .40-.60 with one another (e.g., Borkenau et al., 2004), which is the approximate proportion of variance in such ratings that is attributable to stable differences between targets. Judgments based on minimal exposure to targets in the lab also correlate with self and acquaintance ratings (~r = .10-.30, depending on trait and level of aggregation), both of which are based on more information from other situations (Borkenau et al., 2004).

2.11. Theoretically, some of the existing agreement between perceivers in judging the same targets (see Section 2.7) and of the existing cross-situational consistency (see Section 2.10) may be explained in terms of shared stereotypes based on static target characteristics (e.g., age, gender) rather than the targets' actual behavior (Kenny, 2004). For example, if different perceivers attributed higher levels of some traits to targets merely based on the targets' gender (i.e., exhibiting a shared bias), their consensus would improve. Note, however, that interpreting this in terms of "bias" is only justified to the extent that static and behavioral information are uncorrelated. Sometimes, static information may actually be informative regarding differences in the targets' experiences and behavior, which has been called the "kernel-of-truth" (Jussim et al., 2009; Kenny, 2004). There is some evidence, however, that inter-rater agreement barely changes when accounting for gender and age (Borkenau et al., 2004), suggesting that static information may not actually introduce much bias in personality judgments.

2.12. Many studies have examined whether consensus and accuracy of judgments increase in correspondence to the amount of personality-relevant information that perceivers have about targets (Allik et al., 2016; Beer, 2021). This is referred to as the *acquaintanceship effect* because longer acquaintance is assumed to be related to having more information. Higher levels of acquaintance typically involve more opportunities to communicate with targets and to observe targets' behavior in a larger variety of situations, which should improve the representativeness of cues that are used to make judgments. Judgments of personal acquaintances, especially ones who have higher levels of intimacy with targets, tend to be more accurate than judgments of strangers or less intimate acquaintances (with moderate vs. small effect sizes), especially for less visible traits such as emotional stability (Connelly & Ones, 2010). However, studies that experimentally manipulate the amount of information that is available to perceivers usually yield small to negligible effects, and most gains occur within the first minutes and hours (Beer, 2021). It is interesting to note that self-other agreement is often at above chance levels even when judgments are made by strangers and people with relatively low levels of acquaintance, especially for traits with more visible cues (Connelly & Ones, 2010).

2.13. Personality judgments are valid in that they connect to behavior in various contexts. Self- and informant-ratings of behavior frequencies and personality traits predict measured frequencies of some behaviors (Tackman et al., 2020; Vazire & Mehl, 2008). For example, judgments of targets' intelligence predict performance on intelligence tests and GPA at moderate to large effect sizes (Borkenau et al., 2004; Murphy, 2007), and judgments of conscientiousness and emotional stability predict academic and job performance (Connelly & Ones, 2010). Furthermore, judgments of personality traits predict important life outcomes: Extraversion and Conscientiousness predict longevity and Neuroticism predicts romantic relationship quality (Ozer & Benet-Martínez, 2006).

In laboratory settings, personality judgments by both observers and targets are related to targets' expressed verbal and nonverbal behavior. For example, both observer and self-judgments of Extraversion are positively associated with speech rate (Breil et al., 2021). However, the relation between observer judgments and measured target behavior (i.e., cue utilization) tends to exceed the relation between self or informant-rated personality traits and measured target behavior (i.e., cue validity) in magnitude. When cue validities correspond with cue utilization, judgment accuracy should be enhanced (Brunswik, 1956; Osterholz et al., 2021).

2.14. Both self- and other-judgments of personality do have some incremental predictive validity when compared to each other (Connelly & Ones, 2010; Vazire & Mehl, 2008), indicating that each perspective contains some valid information that the other perspective does not. For example, although self-ratings have some predictive validity by themselves, other-judgments of Conscientiousness, Emotional Stability, and Extraversion are better predictors of academic achievement than self-ratings, as are other-judgments of Conscientiousness for predicting job performance (Connelly & Ones, 2010).

3. Core Findings Using Analyses of Profiles of Personality Judgments

The second major approach to examining personality judgments is the *profile* or person-centered approach, in which agreement/accuracy is based on correspondence between judgments of *multiple* items and traits for a given perceiver-target pair. The profile approach uses multilevel modeling and profile correlations (Furr, 2008), and is featured in person perception models such as the Social Accuracy Model (Biesanz, 2021). The profile approach is useful when investigating personality judgment more broadly (i.e., not limited to a particular trait). Advantages of this approach include higher statistical power and the possibility to simultaneously include predictors at the level of perceivers, targets, dyads, and items. One disadvantage of this approach is being limited to using self- or other-judgments as validation criteria, because obtaining other types of (e.g., observational) data is often not feasible for more than just a few personality dimensions.

Here, we present some core findings based on the profile approach. Given that most of the relevant studies use the same type of data that could also be analyzed with a variable-centered approach, and that several parts of multilevel profile analyses may be directly translated into variable-centered analyses (Allik et al., 2015; Biesanz, 2021), we will focus on findings that are most specific to profile analyses. Also, given that there are currently some discrepancies in how certain technical terms (e.g., "normative accuracy") are used, we decided to avoid such terms altogether.

3.1. The judgment profile of the average target is strongly, but not perfectly (~r = .80-.90), correlated with ratings of the items' social desirability (Edwards, 1953). This is probably a direct reflection of the average perceiver's positive attitude toward the targets, as most perceivers are quite fond of their targets and thus describe them positively across items, regardless of those targets' actual personality characteristics. As a consequence, randomly paired judgment profiles are likely to correlate positively with one another. Not accounting for this fact may lead to erroneous conclusions regarding the "substantive similarity" of two profiles that are actually just rooted in their joined positivity (Wood & Furr, 2016).

3.2. Apart from social desirability, the average judgment profile also contains another kind of systematic variance which is best interpreted in terms of the average target's actual, substantive characteristics (Rogers & Biesanz, 2015). People, on average, show higher levels of some substance variables (e.g., working) than other substance variables (e.g., relaxing). This is another, independent reason why two randomly selected profiles of personality items should correlate positively with one another.

3.3. The two aforementioned profile components (social desirability and the average target's personality) may be disentangled from one another, and from a third component, which reflects the ways in which each target *differs* from the average target (Wessels et al., 2020). Studies using the profile approach show that perceivers agree to a moderate

extent in which characteristics they attribute to particular targets - as opposed to the average target (~r = .20; Human et al., 2019; Rogers & Biesanz, 2019; Wessels et al., 2020).

3.4. The three aforementioned profile components respond differently to differences between dyads in "knowing" (how well perceivers know targets) and attitudes such as how much perceivers like targets. The more a perceiver knows a target, the more accurately he or she will describe both the characteristics that this target has *in common* with most other targets, and the characteristics that set this target *apart* from other targets. At the same time, greater knowing is associated with lower positivity of personality judgments overall. For liking, a largely reverse pattern of associations with the three profile components is found (Wessels et al., 2020). This is important because shared liking may by itself account for profile similarity (Leising et al., 2013), and differences in liking across dyads may account for correlations between profile similarity and positive outcomes reported by the same perceivers. For example, the more two partners like each other, the more similar their ratings of each other will be, due to shared positivity. Also, this greater similarity will correlate, across dyads, with other measures that reflect the same perceivers' attitudes (e.g., relationship satisfaction).

3.5. The *good target* is someone who tends to be judged with higher accuracy than other targets. Research suggests that targets do vary substantially in this regard (Rogers & Biesanz, 2019). Some of this variance may be rooted in good targets' behaving more consistently across time and situations (Human et al., 2019).

3.6. The *good judge* is someone who perceives others with higher accuracy than other perceivers. Research demonstrates that perceivers do differ from one another in this regard (Biesanz, 2021; Rogers & Biesanz, 2019), but there is typically less variability in accuracy across perceivers than variability in accuracy across targets; and this lower variability may explain why it is difficult to find characteristics of perceivers that are consistently related to accuracy (Allik et al., 2016; Colman, 2021). Also, being a good perceiver requires having a good target (i.e., perceivers need valid information on which to base judgments; Rogers & Biesanz, 2019).

4. What Important Questions Remain Unanswered in Personality Judgment Research?

There are still many important unanswered questions in the field, some of which remain debated among researchers of personality judgments, including amongst ourselves. In this section, we briefly touch upon some questions we think are important to address in future personality judgment research.

4.1. What exactly are the substantive characteristics underlying the personality judgments that people make using natural language terms? For example, what is it about a target that makes a perceiver say the person is "dominant" or "kind?" We know some of these indicators (e.g., judgments of intelligence are predicted by eye gaze at an interaction partner; Murphy, 2007; see also Breil et al., 2021), but the picture is far from

complete. Perhaps even more important are questions about the factor structure of these substantive differences. Empirical findings (see Section 2.2) all reflect the structure of personality *judgments*, but this is not informative regarding the covariation among the underlying substance variables themselves. In fact, the Big Two/Five/Six may be rooted partly or even completely in how the average perceiver evaluates different behaviors by comparing them with the same set of goals or ideals (Borkenau, 1990). The question of how "substantive" personality judgment factors really are is highly relevant because only if they *are* substantive (i.e., there is co-variation among the target's actual behaviors) will it be reasonable to look for their between-target (e.g., biological) correlates or even origins.

4.2. What characteristics of perceivers are consistently related to accuracy levels? Much of the existing good-judge research demonstrates meaningful relationships between accuracy and perceiver characteristics such as agreeableness, emotional stability, psychological adjustment, dispositional intelligence, and gender - with women being more accurate when a difference between gender is found (Colman, 2021; Letzring & Funder, 2021). These findings overlap with other interpersonal accuracy research (which is not specific to personality judgments per se) which indicates that accuracy is associated with many positive psychosocial characteristics (Murphy, 2016). On the other hand, there is research demonstrating a lack of relationships between perceiver characteristics and self-other agreement (Allik et al., 2016). Research indicates that there may be good judges of particular traits, rather than good judges of a profile of traits (Hall et al., 2018; Schlegel et al., 2017). Some of the inconsistencies in findings may be due to differences in acquaintanceship between perceiver-target dyads, the type of accuracy measure being used, and/or characteristics of the targets (e.g., Hall et al., 2018; Rogers & Biesanz, 2019).

4.3. Can the ability to be a good judge be trained, and if so, how? Several researchers have speculated that it is possible to instruct people in how to become better judges of others, yet empirical research is mixed. Most empirically-relevant research has focused on judging domains other than personality traits, such as emotion and deception (Blanch-Hartigan & Cummings, 2021). In these domains, the most effective type of training tends to include practice with feedback, with improvements in the small to moderate range. Future research could apply a similar paradigm to the study of personality judgment. If accuracy for judgments of personality may in fact be trained, that would be highly consequential to many fields such as personnel selection and forensic psychiatry.

4.4. Researchers have examined how perceptions of personality made by people are related to technology-based cues such as Facebook likes, language use in emails, and features of pictures that people post (Wall & Campbell, 2021). Although studies with very large samples already demonstrated that some technology-based cues are related to personality with low to moderate effect sizes (e.g., Youyou et al., 2015), further research is needed. This concerns the replicability of these findings, particularly with respect to the importance of individual cues in multivariate models of predicting personality, as

well as how judges use technology-based cues in their judgments and how accurate these judgments are. As technology continues to evolve, future research will have to be responsive to new technologies in relation to personality judgments.

4.5. To what extent do our lab-based findings regarding judgments and accuracy have practical relevance for people's everyday lives? Do the mechanisms of accuracy work the same way both inside and outside of the lab? There is some work in natural settings and much work has examined ratings of acquaintances and family members (Beer, 2021), yet more can be done to examine the ecological validity of personality judgment findings outside the lab. Furthermore, there seems to be a tacit assumption that better-than-chance results (i.e., statistical significance) indicate useful levels of accuracy, but statistical significance and common effect size metrics provide little insight into the practical relevance of such effects outside the lab (Murphy, 2016). For instance, a variable-centered accuracy finding of r = .22 translates to a 61% accuracy rate, which is 11% better than a 50% chance base rate (see Hall et al., 2008). Does an 11% increase in accuracy translate to meaningful differences in perceiver or target experiences?

4.6. There is considerable criticism that human behavioral studies, including psychology, mostly investigate WEIRD people – Western, Educated, Industrialized, Rich, and Democratic (Henrich et al., 2010), which limits generalizability. This pertains directly to the issue of the generalizability of certain structural models (see Section 2.2.), and to many other important issues such as how common it is to view and/or describe one's self positively. There is relatively little cross-cultural research investigating personality judgments (particularly in non-WEIRD populations and zero-acquaintance judgments beyond the Big Five). Investigating personality judgments in diverse populations is an important objective for future research in understanding human perception processes.

Conclusion

Research on personality judgment is a vibrant and active area of study. As demonstrated in this review, much has already been learned about the words and phrases used to describe personality, the structure of personality judgments, and some factors that are related to agreement and accuracy. But there is still much to be learned, and we encourage researchers to continue to address a variety of important issues to further increase our understanding of how accurate personality judgments are made.

Funding: The authors have no funding to report.

Acknowledgments: We are especially appreciative of detailed feedback provided by David A. Kenny and Simon Breil on an earlier draft of this paper.

Competing Interests: Juri Allik is a member of the editorial board, and Daniel Leising is a Methodological Consultant member of the journal.

Author Contributions: *Tera D. Letzring*—Idea, conceptualization | Writing | Feedback, revisions | Project coordination, administration. *Nora A. Murphy*—Writing | Feedback, revisions. *Jüri Allik*—Writing | Feedback, revisions. *Andrew Beer*—Writing | Feedback, revisions. *Johannes Zimmermann*—Writing | Feedback, revisions. *Daniel Leising*—Idea, conceptualization | Writing | Feedback, revisions.

Ethics Statement: No ethical issues and/or ethics approvals need to be disclosed.

Related Versions: The first submission of this manuscript was published via PsyArXiv at https://psyarxiv.com/8t63v/.

Author Note: This article is the result of an invitation issued to Tera D. Letzring to write a review article on the judgment of personality. Per the requested invitation, contributing authors are from different labs, institutions, and multiple countries, and most had not collaborated previously. Tera D. Letzring coordinated the recruitment of authors. Tera D. Letzring and Daniel Leising organized article drafts and should both be considered first authors, and all authors contributed to writing and revising the article.

Supplementary Materials

For this article an Open Peer-Review is available via PsychArchives (for access see Index of Supplementary Materials below).

Index of Supplementary Materials

Personality Science. (Ed.). (2021). *Supplementary materials to: The judgment of personality: An overview of current empirical research findings* [Open peer-review]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.5161

References

Allik, J., Borkenau, P., Hrebícková, M., Kuppens, P., & Realo, A. (2015). How are personality trait and profile agreement related? *Frontiers in Psychology, 6*, Article 785. https://doi.org/10.3389/fpsyg.2015.00785

Allik, J., de Vries, R. E., & Realo, A. (2016). Why are moderators of self-other agreement difficult to establish? *Journal of Research in Personality, 63*, 72-83. https://doi.org/10.1016/j.jrp.2016.05.013

Allik, J., Realo, A., Mõttus, R., Borkenau, P., Kuppens, P., & Hřebícková, M. (2010). How people see others is different from how people see themselves: A replicable pattern across cultures. *Journal of Personality and Social Psychology, 99*(5), 870-882. https://doi.org/10.1037/a0020963

Allik, J., Realo, A., Mõttus, R., Esko, T., Pullat, J., & Metspalu, A. (2010). Variance determines self-observer agreement on the Big Five personality traits. *Journal of Research in Personality, 44*(4), 421-426. https://doi.org/10.1016/j.jrp.2010.04.005

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47*(1), Article i–171. https://doi.org/10.1037/h0093360

Anderson, N. H. (1968). Likableness ratings of 555 personality trait-words. *Journal of Personality and Social Psychology, 9*(3), 272-279. https://doi.org/10.1037/h0025907

Anusic, I., Schimmack, U., Pinkus, R. T., & Lockwood, P. (2009). The nature of correlations among big five ratings: The Halo-Alpha-Beta-Model. *Journal of Personality and Social Psychology, 97*(6), 1142-1156. https://doi.org/10.1037/a0017159

Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality factors: A review of research and theory. *Personality and Social Psychology Review, 18*(2), 139-152. https://doi.org/10.1177/1088868314523838

Back, M. D., & Nestler, S. (2016). Accuracy of judging personality. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 98–124). Cambridge University Press. https://doi.org/10.1017/CBO9781316181959.005

Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality, 43*(3), 335-344. https://doi.org/10.1016/j.jrp.2008.12.013

Baird, B. M., Lucas, R. E., & Donnellan, M. B. (2017). The role of response styles in the assessment of intraindividual personality variability. *Journal of Research in Personality, 69*, 170-179. https://doi.org/10.1016/j.jrp.2016.06.015

Beer, A. (2021) Information as a moderator of accuracy in personality judgment. In T. D. Letzring, & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 132-148). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.9

Beer, A., & Watson, D. (2008a). Asymmetry in judgments of personality: Others are less differentiated than the self. *Journal of Personality, 76*(3), 535-560. https://doi.org/10.1111/j.1467-6494.2008.00495.x

Beer, A., & Watson, D. (2008b). Personality judgment at zero acquaintance: Agreement, assumed similarity, and implicit simplicity. *Journal of Personality Assessment, 90*(3), 250-260. https://doi.org/10.1080/00223890701884970

Biderman, M. D., McAbee, S. T., Chen, Z. J., & 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

Biesanz, J. C. (2018). Interpersonal perception models. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *The SAGE handbook of personality and individual differences: The science of personality and individual differences* (pp. 519–534). SAGE.

Biesanz, J. C. (2021). The Social Accuracy Model. In T. D. Letzring, & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 61-82). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.5

Blanch-Hartigan, D., & Cummings, K. H. (2021). Training and improving accuracy of personality trait judgments. In T. D. Letzring, & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 307-317). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.21

Borkenau, P. (1990). Traits as ideal-based and goal-derived social categories. *Journal of Personality and Social Psychology, 58*(3), 381-396. https://doi.org/10.1037/0022-3514.58.3.381

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

Breil, S. M., Osterholz, S., Nestler, S., & Back, M. D. (2021). Contributions of nonverbal cues to the accurate judgment of personality traits. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 195-218). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.13

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). University of California Press.

Carlson, E. N., Vazire, S., & Furr, R. M. (2011). Meta-insight: Do people really know how others see them? *Journal of Personality and Social Psychology, 101*(4), 831-846. https://doi.org/10.1037/a0024297

Colman, D. E. (2021). Characteristics of the judge that are related to accuracy. In T. D. Letzring, & J. S. Spain (Eds.). *The Oxford handbook of accurate personality judgment* (pp. 85-99). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.6

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

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *The Journal of Applied Psychology, 37*(2), 90-93. https://doi.org/10.1037/h0058073

Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality, 76*(5), 1267-1316. https://doi.org/10.1111/j.1467-6494.2008.00521.x

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Mast, M. S., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality, 42*(6), 1476-1489. https://doi.org/10.1016/j.jrp.2008.06.013

Hall, J. A., Back, M. D., Nestler, S., Frauendorfer, D., Schmid Mast, M., & Ruben, M. A. (2018). How do different ways of measuring individual differences in zero-acquaintance personality accuracy correlate with each other? *Journal of Personality, 86*(2), 220-232. https://doi.org/10.1111/jopy.12307

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2-3), 61-83. https://doi.org/10.1017/S0140525X0999152X

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

Human, L. J., Mignault, M. C., Biesanz, J. C., & Rogers, K. H. (2019). Why are well-adjusted people seen more accurately? The role of personality-behavior congruence in naturalistic social settings. *Journal of Personality and Social Psychology, 117*(2), 465-482. https://doi.org/10.1037/pspp0000193

John, O. P., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). Guilford Press.

Jung, T., & Kenny, D. A. (2005). Judgments of moderator variables of agreement in person perception: A comparison of two cultures. *Asian Journal of Social Psychology, 8*(2), 155-171. https://doi.org/10.1111/j.1467-839x.2005.00164.x

Jussim, L., Cain, T. R., Crawford, J. T., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 199–227). Psychology Press.

Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review, 8*(3), 265-280. https://doi.org/10.1207/s15327957pspr0803_3

Kenny, D. A. (2020). *Interpersonal perception: The foundation of social relationships* (2nd ed.). The Guilford Press.

Kenny, D. A., & West, T. V. (2010). Similarity and agreement in self- and other-perception: A meta-analysis. *Personality and Social Psychology Review, 14*(2), 196-213. https://doi.org/10.1177/1088868309353414

Krzyzaniak, S. L., & Letzring, T. D. (2021). Characteristics of traits that are related to accuracy of personality judgments. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 119-131). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.8

Leising, D., Ostrovski, O., & Zimmermann, J. (2013). "Are we talking about the same person here?": Interrater agreement in judgments of personality varies dramatically with how much the perceivers like the targets. *Social Psychological & Personality Science, 4*(4), 468-474. https://doi.org/10.1177/1948550612462414

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

Letzring, T. D., & Funder, D. C. (2018). Interpersonal accuracy in trait judgments. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *The SAGE handbook of personality and individual difference: Vol 2.*

Applications of personality and individual differences (pp. 253-282). SAGE. https://doi.org/10.4135/9781526451248.n11

Letzring, T. D., & Funder, D. C. (2021). The Realistic Accuracy Model. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 9-22). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.2

Murphy, N. A. (2007). Appearing smart: The impression management of intelligence, person perception accuracy, and behavior in social interaction. *Personality and Social Psychology Bulletin, 33*(3), 325-339. https://doi.org/10.1177/0146167206294871

Murphy, N. A. (2016). What we know and the future of accuracy research. In J. A. Hall, M. Schmid Mast, & West, T. V. (Eds.), *The social psychology of perceiving others accurately* (pp. 404-424). Cambridge University Press. https://doi.org/10.1017/CBO9781316181959.019

Osterholz, S., Breil, S. M., Nestler, S., & Back, M. D. (2021). Lens and dual models. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 45-60). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.4

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57*, 401-421. https://doi.org/10.1146/annurev.psych.57.102904.190127

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology, 88*(5), 879-903. https://doi.org/10.1037/0021-9010.88.5.879

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

Rogers, K. H., & Biesanz, J. C. (2015). Knowing versus liking: Separating normative knowledge from social desirability in first impressions of personality. *Journal of Personality and Social Psychology, 109*(6), 1105-1116. https://doi.org/10.1037/a0039587

Rogers, K. H., & Biesanz, J. C. (2019). Reassessing the good judge of personality. *Journal of Personality and Social Psychology, 117*(1), 186-200. https://doi.org/10.1037/pspp0000197

Schlegel, K., Boone, R. T., & Hall, J. A. (2017). Individual differences in interpersonal accuracy: A multi-level meta-analysis to assess whether judging other people is one skill or many. *Journal of Nonverbal Behavior, 41*(2), 103-137. https://doi.org/10.1007/s10919-017-0249-0

Tackman, A. M., Baranski, E. N., Danvers, A. F., Sbarra, D. A., Raison, C. L., Moseley, S. A., Polsinelli, A. J., & Mehl, M. R. (2020). 'Personality in its natural habitat' revisited: A pooled, multi-sample examination of the relationships between the big five personality traits and daily behaviour and language use. *European Journal of Personality, 34*(5), 753-776. https://doi.org/10.1002/per.2283

Thalmayer, 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

Thielmann, I., Hilbig, B. E., & Zettler, I. (2020). Seeing me, seeing you: Testing competing accounts of assumed similarity in personality judgments. *Journal of Personality and Social Psychology, 118*(1), 172-198. https://doi.org/10.1037/pspp0000222

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

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology, 95*(5), 1202-1216. https://doi.org/10.1037/a0013314

Wall, H. J., & Campbell, C. (2021). Accuracy of personality trait judgments based on environmental and social media cues. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment* (pp. 219-234). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.14

Wessels, N. M., Zimmermann, J., Biesanz, J. C., & Leising, D. (2020). Differential associations of knowing and liking with accuracy and positivity bias in person perception. *Journal of Personality and Social Psychology, 118*(1), 149-171. https://doi.org/10.1037/pspp0000218

West, T. V., & Kenny, D. A. (2011). The Truth and Bias model of judgment. *Psychological Review, 118*(2), 357-378. https://doi.org/10.1037/a0022936

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

Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl: Vol. 2. Personality and psychopathology* (pp. 89–113). University of Minnesota Press.

Wood, D., & Furr, R. M. (2016). The correlates of similarity estimates are often misleadingly positive: The nature and scope of the problem, and some solutions. *Personality and Social Psychology Review, 20*(2), 79-99. https://doi.org/10.1177/1088868315581119

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences of the United States of America, 112*(4), 1036-1040. https://doi.org/10.1073/pnas.1418680112

Zimmermann, J., Schindler, S., Klaus, G., & Leising, D. (2018). The effect of dislike on accuracy and bias in person perception. *Social Psychological & Personality Science, 9*(1), 80-88. https://doi.org/10.1177/1948550617703167

eapp

Personality Science (PS) is an official journal of the European Association of Personality Psychology (EAPP).

leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.