



AQ: 1  
AQ: 2  
AQ: 3  
AQ: 4

# Personality Evaluated: What Do People Most Like and Dislike About Themselves and Their Friends?

Jessie Sun<sup>1, 5</sup>, Becky Neufeld<sup>2, 5</sup>, Paige Snelgrove<sup>3, 5</sup>, and Simine Vazire<sup>4, 5</sup>

<sup>1</sup> Department of Psychology, University of Pennsylvania

<sup>2</sup> Department of Psychology, University of Utah

<sup>3</sup> Department of Psychology, California State University, Sacramento

<sup>4</sup> Melbourne School of Psychological Sciences, The University of Melbourne

<sup>5</sup> Department of Psychology, University of California, Davis

What do people think their best and worst personality traits are? Do their friends agree? Across three samples, 463 college students (“targets”) and their friends freely described two traits they most liked and two traits they most disliked about the target. Coders categorized these open-ended trait descriptors into high or low poles of six trait domains (extraversion, agreeableness, conscientiousness, emotional stability, openness, and honesty-humility) and judged whether targets and friends reported the same specific best and worst traits. Best traits almost exclusively reflected high levels of the major trait domains (especially high agreeableness and extraversion). In contrast, although worst traits typically reflected low levels of these traits (especially low emotional stability), they sometimes also revealed the downsides of having high levels of these traits (e.g., high extraversion: “loud”; high agreeableness: “people-pleaser”). Overall, targets and friends mentioned similar kinds of best traits; however, targets emphasized low emotional stability worst traits more than friends did, whereas friends emphasized low prosociality worst traits more than targets did. Targets and friends also showed a moderate amount of self–other agreement on what the targets’ best and worst traits were. These results (a) shed light on the traits that people consider to be most important in themselves and their friends, (b) suggest that the desirability of some traits may be in the eye of the beholder, (c) reveal the mixed blessings of different traits, and, ultimately, (d) provide a nuanced perspective on what it means for a trait to be “good” or “bad.”

**Keywords:** evaluative judgments, person perception, personality judgment, self–other agreement, trait desirability

**Supplemental materials:** <https://doi.org/10.1037/pspp0000388.supp>

Personality psychology has traditionally focused more on *describing* what people are like than on *evaluating* whether their personalities are good or bad. Indeed, lexical studies from which

the Big Five were derived systematically excluded highly evaluative person descriptors (e.g., “annoying,” “charming,” “strange,” “moral”) from the domain of personality (Allport & Odbert, 1936;

Jessie Sun <https://orcid.org/0000-0001-6764-0721>  
Becky Neufeld <https://orcid.org/0000-0002-7270-5259>  
Paige Snelgrove <https://orcid.org/0000-0001-5102-5900>  
Simine Vazire <https://orcid.org/0000-0002-3933-9752>

Simine Vazire acquired funding, designed, and supervised data collection for the Sample 1 data. Jessie Sun designed the data collection for Samples 2 and 3. Jessie Sun and Becky Neufeld collected the data for Samples 2 and 3. Jessie Sun and Simine Vazire conceptualized this project and coded the broad trait domains. Jessie Sun, Becky Neufeld, and Paige Snelgrove curated and analyzed the data. Jessie Sun collected the trait rating data and conducted all supplemental analyses. Jessie Sun and Becky Neufeld drafted the method and results. Jessie Sun drafted the introduction, discussion, and [online supplemental materials](#). Simine Vazire provided extensive feedback and critical revisions. All authors interpreted the results, edited the manuscript, and approved the final version for submission.

Data collection for Sample 1 was supported by a grant from the National Science Foundation to Simine Vazire (BCS-1125553). The study materials and the data and code required to reproduce the analyses reported in this article are available at <https://osf.io/jce7k>. We are grateful to collaborators and research assistants who contributed to the design and data collection for Sample 1; to Geoff Goodwin for contributing to the design of the data collection for Samples 2 and 3; to the research assistants who coded the specific trait data; to Sophie Faircloth for her assistance with collecting the trait rating data; to Kun Zhao and Luke Smillie for feedback on the coding scheme for Samples 2 and 3; to Joe Simmons for an insightful interpretation of the randomization test results; and to Joshua Wilt, Damien Crone, and Luke Smillie for comments on a draft of this article.

Correspondence concerning this article should be addressed to Jessie Sun, Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104, United States. Email: [jessun@sas.upenn.edu](mailto:jessun@sas.upenn.edu)

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Cattell, 1943; Norman, 1967). However, many personality psychologists believe that personality judgment is a fundamentally evaluative endeavor (Borkenau, 1990; Hogan, 1996; McAdams & Pals, 2006). In recent years, there has also been increased interest in understanding evaluative phenomena such as the structure of highly evaluative descriptors (Benet-Martínez & Waller, 2002), moral character and “dark” traits (Cawley et al., 2000; Fleenor et al., 2014; Furnham et al., 2013; Goodwin, 2015; Peterson & Seligman, 2004), traits that characterize a person who is psychologically healthy (Bleidorn et al., 2020) and liked by others (Wortman & Wood, 2011), and how people want to improve their personalities (Baranski et al., 2017; Hudson & Roberts, 2014; Sun & Goodwin, 2020). In other words, people not only have different *levels* of various personality traits but also judge some traits to be more *desirable* than others.

Evaluative self- and other-judgments have clear practical relevance in real-world contexts. For example, “What are your greatest strengths and weaknesses?” is one of the most common job interview questions. Here, we examine how people would answer this question when considering what their best and worst traits are in general, and whether the answer would be different if we asked their friends. We believe that there are at least three reasons to investigate self- and other-perceptions of best and worst traits. First, best and worst traits may shed light on the traits that people consider to be most personally or socially important. Second, mapping best and worst traits onto existing trait taxonomies can provide nuanced insights about the desirability of high or low levels of major trait domains. Third, self–other agreement and asymmetries on best and worst traits can reveal the extent to which the desirability of a trait is in the eye of the beholder. We expand on each of these potential contributions below.

### The Importance of Evaluative Traits

Which traits do people most care about in themselves and most want to know about others? A fundamental assumption that underlies the development of modern trait taxonomies such as the Big Five (John et al., 2021) and HEXACO (Ashton & Lee, 2007) is that socially important personality characteristics become encoded in natural language (i.e., the lexical hypothesis; for reviews, see John et al., 1988; Saucier & Goldberg, 2001). Indeed, these trait taxonomies may be better conceptualized as taxonomies of social perception, rather than personality itself (Srivastava, 2010). Thus, the lexical approach can be used to uncover personality characteristics and social perceptions that are worth studying and explaining (Saucier & Goldberg, 2001). Considering that people judge highly desirable or undesirable traits to be especially important (Leising et al., 2014), asking people about the most desirable and undesirable traits that characterize them or their friends may provide an especially promising path to highlighting the personality characteristics that people most care about.

Three types of studies have addressed related questions about the desirability of different traits. In the first type of study, people have rated the likeability or desirability of hundreds of trait descriptors (e.g., Anderson, 1968; Chandler, 2018; Hampson et al., 1987; Leising et al., 2014; Wood, 2015). However, many extremely desirable and undesirable traits may be rare. Extremely undesirable traits (e.g., “malicious,” “inhumane,” “phony”) may be especially uncommon, given that desirable traits are more

frequently used to describe others (Leising et al., 2014). Thus, the traits that are rated as being most desirable or undesirable in the abstract may not be the traits that people most like and dislike about actual people. However, the finding that the desirability of a trait is almost perfectly correlated (.98) with its relational impact (operationalized as the extent to which perceiving a person as having a given trait would increase or decrease their likelihood of entering into various relationships with them; Wood, 2015) suggests that socially important traits may be especially strongly represented among best and worst traits.

In the second type of study, people described either the traits of people they like or dislike or the positive or negative traits of people they know (Alves et al., 2016, 2017a; Leising et al., 2012). Existing studies focused on the *quantitative* properties of these descriptors, showing, for example, that people have a more differentiated vocabulary for describing disliked persons compared with liked persons (Leising et al., 2012); they tend to use more of the same descriptors to describe liked persons compared with disliked persons (Alves et al., 2016); and positive descriptors are more likely than negative descriptors to be shared across people (Alves et al., 2017a). These studies, however, did not describe the *content* (e.g., trait domain) of liked and disliked traits.

The third type of study examines the traits that are associated with positive impressions of others. For example, Wortman and Wood (2011) examined the personality traits that were correlated with being liked by peers. They found that people who were higher on communal traits that promote others’ interests (e.g., “warm,” “cooperative,” “truthful/honest,” “kind-hearted/caring”) tended to be liked more by others. Goodwin and colleagues (2014) further demonstrated that warmth traits can be separated into warmth and morality dimensions and that moral character information (e.g., “humble,” “kind,” “helpful,” “empathetic,” “responsible,” “honest,” “trustworthy,” “loyal”) most strongly determines the overall positivity of the impressions that people form of others (compared with warmth and competence information). Although the question of which traits are associated with being liked by others is distinct from the question of which traits are liked or disliked by others, these studies imply that other-perceptions of best and worst traits might emphasize traits related to warmth and morality.

### Revisiting the Desirability of Trait Domains

Mapping peoples’ best and worst traits onto existing trait taxonomies such as the Big Five (John et al., 2021) and HEXACO (Ashton & Lee, 2007) models serves two purposes. First, trait taxonomies can be used to parsimoniously summarize the hundreds of different best and worst traits that people might report. Doing so allows inferences to be drawn about which types of traits—broadly speaking—are more or less frequently considered to be a person’s best and worst traits. Second, learning where in these taxonomies the most loaded (i.e., liked and disliked) personality traits can be found can provide nuanced insights into the desirability of having high or low levels of different trait domains.

Considering that the Big Five and HEXACO taxonomies were derived from a set of person descriptors that excluded highly evaluative descriptors (Allport & Odbert, 1936; Ashton et al., 2004; Cattell, 1943; Norman, 1967), one might wonder whether they can adequately capture self- and other-judgments of best and worst

traits. We believe that they can. Very few trait descriptors are evaluatively neutral; instead, most descriptors are rated as being positive or negative to some extent (Anderson, 1968; Chandler, 2018; Leising et al., 2012). Thus, even though the development of modern trait taxonomies excluded highly evaluative terms, this does not imply that the resulting domains are evaluatively neutral. Instead, each of the Big Five and HEXACO domains has a socially desirable or “positive” pole, such that people judge high levels of extraversion, agreeableness, conscientiousness, emotional stability, openness, and honesty-humility to be more desirable than low levels of these traits (Dunlop et al., 2012; John & Robins, 1993; Kim et al., 2019).

People also tend to prefer to use descriptors that are abstract enough to be descriptive of several behaviors (e.g., “kind”) compared with descriptors that only describe specific behaviors (e.g., “charitable”) or that are so abstract that they are devoid of descriptive meaning (e.g., “good”, John et al., 1991). This implies that when describing a person’s best and worst traits, people would be naturally inclined to use descriptors that are concrete enough to be mapped onto a broad trait domain. Finally, when highly evaluative descriptors were included in factor analyses, Tellegen and Waller (1987; as cited in Benet-Martínez & Waller, 2002) found that evaluative terms denoting extreme openness (e.g., “peculiar,” “odd,” “unusual”) emerged as indicators of low conventionality (i.e., high openness). This implies that even though highly evaluative descriptors were traditionally excluded because they were considered to be too ambiguous and empty of descriptive content to be behaviorally informative, many evaluative descriptors in fact contain enough descriptive content that they can be mapped onto major trait domains.

Because each trait domain has a socially desirable pole, we might expect that best traits are more likely to reflect their socially desirable poles, whereas worst traits are more likely to reflect their socially undesirable poles. At the same time, a growing body of research considers the evolutionary (Nettle, 2006), cybernetic (DeYoung, 2015), and practical costs (e.g., for job performance, income, happiness; Grant, 2013; Judge et al., 2012; Smith et al., 2018; Wiese et al., 2018) of having “too much” of the desirable end of various traits. The Aristotelian perspective that virtue is the golden mean implies that both deficient and excessive expressions of traits (relative to what is best for a given situation) can be suboptimal (Grant & Schwartz, 2011; Ng & Tay, 2020). Similarly, people perceive both extremely high and extremely low levels of personality traits as being suboptimal (Bleidorn et al., 2020; Borkekenau et al., 2009; Kuncel & Tellegen, 2009). Indeed, one view of personality dysfunction posits that maladaptive traits reflect the extreme tails of general personality dimensions (e.g., Suzuki et al., 2015; Widiger & Mullins-Sweatt, 2009; cf. Morey et al., 2020). This might explain why a minority of people express desires to be less extraverted, agreeable, conscientious, emotionally stable, open, and honest (Hudson & Roberts, 2014; Sun & Goodwin, 2020). Finally, Peabody’s (1967) “sets of four” approach to separating the descriptive and evaluative aspects of personality judgments suggests that the same “facts” about a person may be interpreted in either a positive or negative light (for a formalized model, see Leising et al., 2015). For example, a person could be described as either “modest” or “self-disparaging” (high honesty-humility), or as either “confident” or “conceited” (low honesty-humility). For these reasons, traits that reflect the socially desirable

poles of broad trait domains may sometimes be considered to be a person’s worst traits.

Which of these broad trait domains are likely to be the most “loaded”? One possibility is that the domains that have previously been found to be particularly evaluative (i.e., especially socially desirable or undesirable) would also be the domains that best and worst traits most frequently reflect. However, different methods of operationalizing evaluativeness have produced dramatically different results. For example, John and Robins (1993) operationalized evaluativeness as the distance between the mean desirability levels of the high pole and low pole of each domain based on undergraduates’ ratings of 76 traits. They found that openness was most evaluatively polarized, followed by agreeableness and conscientiousness, then emotional stability. Extraversion was the least evaluative domain. Kim and colleagues (2019) operationalized evaluativeness as the difference between the midpoint of the scale and the mean desirability rating by experts (who rated 25 facet labels, e.g., “Curiosity,” “Anxiety,” “Sociability”). They found that conscientiousness was the most evaluative domain, followed by agreeableness, emotional stability, and extraversion. Openness was the least evaluative domain (i.e., the complete opposite to what John and Robins found).

These mixed findings show that the “evaluativeness” of a domain depends on which items are used to assess that domain. Because both studies relied on a limited set of items or facets, the evaluativeness of the domains based on these specific operationalizations may not generalize to the true evaluativeness of the domains based on the larger universe of person descriptors (including highly evaluative descriptors). This problem is exacerbated to the extent that the evaluativeness of excluded descriptors differs between trait domains. Moreover, as mentioned above, the traits that are rated to be most desirable or undesirable in the abstract may not be the same traits that people consider to be their own and others’ actual best and worst traits. Mapping best and worst traits onto the high and low poles of major trait domains can provide insight into which of these domains most frequently feature evaluative judgments *in practice*.

### Self–Other Perspectives on Best and Worst Traits

Finally, abstract ratings of the extent to which a trait is desirable do not answer the question of *for whom* a trait is desirable. Research on the links between personality and consequential life outcomes has typically focused on the *personal* benefits of having certain personality traits (e.g., extraverts tend to be happier, Anglim et al., 2020; conscientious people tend to live longer; Jokela et al., 2020). Less research has examined how a person’s personality traits affect *other people* (for a review, see Back & Vazire, 2015). In some cases, the social consequences of personality traits may mirror their personal consequences. For example, the spouses of conscientious people tend to live longer (Roberts et al., 2009) and enjoy greater occupational success (Solomon & Jackson, 2014), over and above the effects of their own levels of conscientiousness. However, it is also possible that some traits have positive consequences for the self but negative consequences for others, or vice versa. For example, interpersonal theory’s principle of complementarity posits that dominant behaviors elicit submissive responses in one’s interaction partners (Leary, 1957; Orford, 1986). This implies that people who have higher levels of

agentic traits such as extraversion may be better at asserting their needs and preferences, but perhaps at the expense of accommodating others' needs and preferences.

Comparing self- and friend-perspectives on best and worst traits can shed light on this issue. We do so by examining (a) whether friends tend to emphasize similar or different best and worst traits compared with the self (i.e., self–other asymmetries) and (b) the extent to which friends agree with targets on what their best and worst traits are (i.e., self–other agreement). To the extent that friends name similar best and worst traits as the self (e.g., if friends and the self both focus equally on high agreeableness best traits), this would imply that these traits have similarly positive or negative consequences for the self and for others. In contrast, to the extent that friends emphasize some best or worst traits more than the self does, or vice versa, this would imply that the personal and social consequences of these traits may diverge, such that (a) the trait affects the self more than it does friends (or vice versa), or (b) the trait is good for the self but bad for friends (or vice versa). Thus, examining the extent to which friends typically report similar kinds of best and worst traits as the self can shed light on the extent to which traits' desirability is in the eye of the beholder.

We can also examine whether there is self–other agreement on a specific person's best and worst traits. In other words, to what extent do individual targets and friends agree on what the targets' best and traits are? Most research on self–other agreement has focused on statistical correlations between self- and informant-reports on scale-based measures (for a review, see [Connelly & Ones, 2010](#)). The lay concept of agreement is much more straightforward—most people might simply want to know whether the self and their friends use similar terms when describing that person's personality. For example, if Natalie says that her best trait is that she is “kind,” does her friend also say that her best trait is “kind” (or use a similar descriptor, such as “caring”)? In other words, are the qualities that we most value and despise in ourselves the same qualities that our friends value and despise in us? If some traits are good for the self but bad for others, we may even see some cases in which a friend thinks that one of the target's self-reported best traits is actually their worst trait (or vice versa).

## The Present Study

Our primary goal is to provide a rich description and systematic categorization of the traits that people most like and dislike about themselves and their friends. Typically, when people describe themselves and others in real life, they use whichever person descriptors spontaneously come to mind instead of filling out personality questionnaires or selecting descriptors from a constrained set of options. Thus, to maximize ecological validity, we used a free-response design to ask target participants (“targets”) to state, in their own words, their own best (or most liked) and worst (or most disliked) traits. We also asked their friends to tell us which traits they most liked and disliked in the target. These friend-reports allowed us to investigate self–other asymmetries (do friends emphasize different best and worst traits compared with the self?) and self–other agreement (to what extent do friends agree with their targets on what their best and worst traits are?). We examine self–other agreement at the level of (a) broad traits (e.g., whether a target and their friend both name a high agreeableness best trait) and (b) specific traits (e.g., whether a target and

their friend name very similar best traits, such as “kind” and “caring”). Grounded in a tradition that values comprehensive descriptions of naturally-occurring phenomena ([Asch, 1952/1987](#); [Rozin, 2001](#)), we took an exploratory approach, rather than seeking to test specific hypotheses.

## Method

### Ethics and Open Practices Statement

We used data from three of our existing datasets. Data collection and coding procedures for Sample 1 were approved by Institutional Review Boards (IRBs) at Washington University in St. Louis (IRB ID: 201206090; Study Title: Personality and Intimate Relationships Study) and the University of California, Davis (IRB ID: 669518–15; Study Title: Personality and Interpersonal Roles Study). Data collection procedures for Samples 2 and 3 were approved by the IRBs at the University of Pennsylvania (Sample 2; IRB ID: 831767; Study Title: Moral Change Goals) and the University of California, Davis (Sample 3; IRB ID: 1328211-2). Data collection procedures for trait ratings (which we use for supplemental analyses; see the [online supplemental materials](#), Section 7) were approved by the IRB at University of Pennsylvania (IRB ID: 844999; Study Title: Best and Worst Trait Ratings).

For Sample 1, we used data from the first wave of the longitudinal Personality and Interpersonal Roles Study (PAIRS). Other published articles have used the PAIRS dataset (for a full list of citations, see <https://osf.io/3uag4/wiki/home/>). A few articles used the self- and informant-reports of personality traits that we use in supplemental analyses, but none have used the best and worst trait measures included in this study. For Samples 2 and 3, we used data from a study on personality change goals. The previously published article using these samples ([Sun & Goodwin, 2020](#)) used the self- and informant-reports of personality traits that we use in supplemental analyses, but did not use the best and worst trait measures. Codebooks for all measures in these datasets are available at <https://osf.io/jce7k/>. Below, we describe the measures and procedures relevant to the current article.

The codebook, data (posted in a way that prevents targets from finding out what their friends said about them) and R scripts required to reproduce the analyses reported in this article are available at <https://osf.io/jce7k/>. We did not preregister these analyses as we were already familiar with the datasets when we conceptualized this project. Instead, to limit the risk of overinterpreting potentially spurious effects, we highlight the findings that replicate across at least two samples (at a conventional  $p < .05$  threshold) and are therefore more likely to be robust. The effects reported in the results section that met this replication threshold also met an alternative standard of evidence for claims of new discoveries—whether the effects are significant at a  $p < .005$  threshold ([Benjamin et al., 2018](#))—in at least one sample. Note that we coded and analyzed a few additional variables in the Sample 1 data for an undergraduate research project (see the [online supplemental materials](#), Section 1). We later refined the scope of the current paper to the variables that are presented in this paper and coded only these variables in the Sample 2 and 3 data. Apart from the additional variables coded in Sample 1, we report all coded variables.

## Participants and Procedure

We restricted our investigation to college students in the United States, based on convenience and resources available to us. However, we obtained data from students at three different universities across different regions of the United States (West, Midwest, and Northeast) and across public and private universities, thus providing some data regarding the generalizability of our results at least to the broader population of U.S. college students. Sample sizes were determined based on time and budgetary constraints, as well as whether targets and friends provided complete data on the best and worst trait questions. The resulting sample sizes range from 121 to 184 target–friend pairs per sample. Given that participants each reported two best and two worst traits, these sample sizes ensure a minimum of 242 responses per trait type (e.g., self-reported best traits).

### Sample 1

Sample 1 comprised 434 students at Washington University in St. Louis, who were recruited in 2012 and 2013 via flyers and classroom announcements. Targets completed a battery of questionnaires during an initial laboratory-based assessment (\$20 compensation), in which they reported their best and worst traits (using the measures described below). Targets also nominated several types of informants, who reported on the targets' best and worst traits. We chose to only use (and only coded) informant data from the target's best friend in St. Louis, to keep the type and number of informants consistent across targets. Because these measures were added a few months after the study began, a subset of targets ( $n = 28$  after exclusions) and their informants did not complete these measures at the first wave. Instead, these targets and their informants completed these "catch-up measures" at the second wave (four months after the first wave). Thus, we used data from the second wave for 28 participants, and data from the first wave for the remaining participants.

We ended data collection when we reached the end of a semester and had recruited at least 400 participants. We excluded targets who did not have a best friend informant, and only retained targets who had two best and two worst traits for both the self- and friend-reports (i.e., no missing data). After these exclusions, the final subset of 184 targets (132 women, 50 men, 2 not reported) used in the Sample 1 analyses ranged in age from 18 to 29 years ( $M = 19.46$ ,  $SD = 1.99$ ) and identified as White ( $n = 107$ ), Asian ( $n = 40$ ), Black ( $n = 13$ ), American Indian or Alaska Native ( $n = 2$ ), Native Hawaiian or other Pacific Islander ( $n = 1$ ), or other or multiple ( $n = 21$ ). The final subset of 184 friends (135 women, 48 men, 1 not reported) used in the Sample 1 analyses ranged in age from 18 to 43 years ( $M = 19.83$ ,  $SD = 2.73$ ) and reported having known their targets for an average of 3.27 years ( $SD = 3.22$ ).

### Samples 2 and 3

Because the data collection procedures for Samples 2 and 3 were almost identical to each other, we report the methods for these two samples together. Targets for Samples 2 and 3 comprised undergraduate students from the University of Pennsylvania ( $N = 300$ ) and the University of California, Davis ( $N = 500$ ) who were compensated with course credit. Stopping rules were determined based on preregistered criteria (for details, see Sun &

Goodwin, 2020). Targets completed a survey in which they reported their best and worst traits and nominated up to four informants. Of the 1,023 (Sample 2) and 1,464 (Sample 3) nominated informants, 417 (Sample 2) and 541 (Sample 3) informants responded to a survey in which they reported the targets' best and worst traits. Informants were entered into a prize drawing for a 1 in 10 chance of winning a \$20 Amazon.com gift card. To maximize comparability with Sample 1, we selected the first friend informant (i.e., excluded any other informant types, such as romantic partners or family members) who had complete data on the best and worst trait questions. As for Sample 1, we only retained targets who had complete data on the best and worst trait questions for both the self- and friend-reports.

After these exclusions, the final subset of 121 targets (92 women, 28 men, 1 not reported) used in the Sample 2 analyses ranged in age from 18 to 22 years ( $M = 19.41$ ,  $SD = 1.12$ ) and identified as White/Caucasian ( $n = 47$ ), Asian ( $n = 44$ ), Hispanic/Latino ( $n = 13$ ), Black ( $n = 7$ ), Pacific Islander ( $n = 1$ ), other or multiple ( $n = 8$ ), or did not disclose their ethnicity ( $n = 1$ ). The final subset of 121 friends (93 women, 27 men, 1 other) used in the Sample 2 analyses ranged in age from 18 to 27 years ( $M = 19.4$ ,  $SD = 1.38$ ) and reported having known their targets for an average of 4.24 years ( $SD = 4.02$ ).

The final subset of 159 targets (135 women, 24 men) used in the Sample 3 analyses ranged in age from 18 to 47 years ( $M = 19.69$ ,  $SD = 2.78$ ) and identified as White/Caucasian ( $n = 29$ ), Asian ( $n = 81$ ), Hispanic/Latino ( $n = 28$ ), Black ( $n = 1$ ), Pacific Islander ( $n = 1$ ), or other or multiple ( $n = 19$ ). The final subset of 159 friends (121 women, 37 men, 1 not reported) used in the Sample 3 analyses ranged in age from 18 to 63 years ( $M = 20.02$ ,  $SD = 4.15$ ) and reported having known their targets for an average of 4.85 years ( $SD = 4.57$ ).

## Measures

### Open-Ended Responses

Targets described their two best traits and their two worst traits (Sample 1: "My two [best/worst] qualities/traits are . . ."; Samples 2–3: "What do you [like/dislike] most about your personality?"). Targets' friends described the target's two best traits and two worst traits (Sample 1: "[target's name]'s two [best/worst] qualities/traits are . . ."; Samples 2–3: "What do you [like/dislike] most about [target's name's] personality?"). We converted the open-ended responses into lowercase, corrected spelling errors, and removed identifying information, but otherwise retained the full context of the text responses for the two coding tasks described below.

### Broad Trait Categories

Two experts (Jessie Sun and Simine Vazire) coded the responses into broad trait categories.

**Big Five Domains.** For Sample 1, we coded the trait descriptors into one of twelve mutually exclusive categories: high or low levels of each of the Big Five domains (extraversion, agreeableness, conscientiousness, emotional stability, and openness), not a Big Five trait, or not a personality trait.

**Big Six Domains.** The HEXACO model is a six-factor alternative to the Big Five model (Ashton & Lee, 2007). The sixth

factor, honesty-humility, describes tendencies toward interpersonal sincerity (i.e., being unwilling to manipulate others), fairness, greed avoidance, and modesty. This factor has sometimes been labeled “morality” or “integrity” and has been interpreted as reflecting individual differences in selfishness (Diebels et al., 2018). To draw a more fine-grained distinction between two aspects of prosociality—kindness and politeness-related traits (agreeableness) versus integrity-related traits such as honesty, fairness, loyalty, sincerity, selfishness, and morality (honesty-humility)—we decided to additionally code high or low levels of HEXACO honesty-humility in Samples 2 and 3 (for which responses were pooled and coded together).

In other words, we coded the trait descriptors into one of fourteen mutually exclusive categories: high or low levels of each of the “Big Six” domains (Big Five plus honesty-humility), not a Big Six trait, or not a personality trait. As a result, some descriptors that were coded into the agreeableness domain (or not a Big Five trait) in Sample 1 were coded into the honesty-humility domain in Samples 2 and 3. Apart from this modification, to maximize comparability of the Big Five results with Sample 1, we generally retained the Big Five definitions of the agreeableness and emotional stability dimensions (rather than adopting the HEXACO rotational variants of these traits; see Ashton et al., 2014). Thus, the Big Six categories we coded should be thought of as the Big Five domains plus honesty-humility rather than codings based on the HEXACO model (see the online supplemental materials, Section 2 for the definitions we used for the Big Six codings).

**Coding Procedure.** The expert coders coded all of the best traits first, and then all of the worst traits (pooled across self- and friend-reports). Duplicate responses within each set of traits (best traits or worst traits) were only coded once; the coded categories were then applied to all other instances of that response. If more than one descriptor was mentioned in a given response, we coded for the trait category that seemed most salient, central, or able to capture the multiple descriptors. If multiple descriptors seemed to be weighted equally, we coded the trait category that was mentioned first. If a response seemed to reflect blends of different trait domains, we coded that response into the best-fitting Big Five or Six category. We aimed to use the *not a Big Five/Six trait* category only in cases where the descriptor could not plausibly fit into any of the Big Five or Six categories. As a result, only a small percentage of responses could not be coded into a broad trait category: 1.36% (Sample 1), 1.03% (Sample 2), and 0.79% (Sample 3) of responses were coded as *not a Big Five/Six trait* (e.g., “annoying,” “religious,” “family-oriented”) and 0.68% (Sample 1), 1.03% (Sample 2), and 0.79% (Sample 3) were coded as *not a personality trait* (e.g., “always hungry,” “athletic,” “cute”).

Interrater agreement (Cohen’s  $\kappa$ ) was substantial for Sample 1 ( $\kappa = .70$ ) and the combined Samples 2 and 3 ( $\kappa = 0.74$ ). Coders resolved disagreements through discussion and, for Samples 2 and 3, by defaulting to the decisions we made for Sample 1 (except for traits categorized into the honesty-humility dimension). When discussion did not lead to a resolution, we crowdsourced judgments via Twitter polls (Vazire, 2020). Tables S3–S5 in the online supplemental materials show the most frequent descriptors that were coded into each of the trait categories. People tended to report descriptors that were from two different trait categories (approximately 70% of the time; e.g., reporting two best traits that reflected high extraversion and high agreeableness, respectively), rather

than the same trait category (e.g., reporting two best traits that both reflected high agreeableness; see Table S6 in the online supplemental materials).

### Agreement on Specific Traits

We used the broad trait codings to examine self–other agreement on broad traits. However, we were also interested in whether friends agreed with targets on what their best and worst traits specifically were. Research assistants provided these judgments in a separate coding task.

**Checking for Distinct Traits.** For this coding task, we required that targets and friends each reported two distinct best traits and two distinct worst traits, otherwise we excluded them completely. For example, if a friend reported that the target’s two worst traits were “stressed” and “stressed out,” their responses were not included in this coding task. This was to ensure that self–other agreement would not be inflated due to the target or the friend reporting the same trait twice. Thus, before the agreement coding task, the authors (Sample 1: Becky Neufeld and Paige Snelgrove; Samples 2–3: Jessie Sun and Becky Neufeld) checked whether each pair of best or worst traits for the self-reports and friend-reports were synonyms or near-synonyms. Becky Neufeld and Paige Snelgrove agreed in 97.92% of the cases for Sample 1 (Jessie Sun resolved disagreements), and Jessie Sun and Becky Neufeld agreed in 96.07% and 97.33% of the cases for Samples 2 and 3, respectively (Simine Vazire resolved disagreements). After these exclusions, these analyses involved 182 (Sample 1), 109 (Sample 2), and 146 (Sample 3) targets and their friends.

**Agreement Coding Task.** To code agreement, six research assistants completed coding “trials” to indicate whether targets and friends mentioned the same or very similar best traits or worst traits. In the “best trait” coding trials, coders were presented with one of the two descriptors that a target reported as their best traits, along with both of the descriptors that their friend used to describe their best traits. In the “worst trait” coding trials, coders were presented with one of the two descriptors that a target reported as their worst traits, along with both of the descriptors that their friend used to describe their worst traits. In each trial, the research assistants coded whether or not the self-reported descriptor was the exact same descriptor or a near-synonym of either of the two friend-reported descriptors. For example, for the trial, “Is “KIND” equal to either “CARING” or “CHEERFUL”?” (No or Yes), the correct answer would be Yes, because “kind” (self-reported) is a synonym of “caring” (friend-reported). These trials were repeated until coders made a judgment for each of the targets’ self-reported descriptors. Best trait and worst trait trials were interspersed and presented in a randomized order. Although the trials were not explicitly labeled as being about best or worst traits, this would have been obvious based on the content of the descriptors listed in each trial.

Before starting this task, the research assistants attended a training session in which they saw examples of descriptor sets and were given guidance on which ones would be considered matches. The task was split up into six blocks for Sample 1 and two blocks each for Samples 2 and 3. For Sample 1, all research assistants started with a training block. Then, we staggered which of the remaining blocks each research assistant continued with. For Samples 2 and 3, research assistants completed a training survey that

contained descriptors from Sample 1. Then, we staggered which of the four blocks they started on.

For each trial, if at least half of the coders said the descriptors matched, we coded it as matching; otherwise, we coded it as not matching. We chose to use this dichotomized metric (instead of the continuous percentage of coders who said that the descriptors matched) for ease of interpretation and comparability with self–other agreement at the level of broad trait categories. However, supplemental analyses (see [Table S7](#) in the online supplemental materials) showed that the results were similar regardless of whether the continuous or dichotomous metric was used. Because of a merging error, the descriptors included in the initial coding trials for Sample 1 were incorrect for 22 participants. Because this error was discovered after these research assistants had left the lab, the four authors repeated the coding trials for these participants and also coded 14 additional participants who had previously been excluded from this coding task due to exclusion criteria for the undergraduate research project (see the [online supplemental materials](#), Section 1 for details). Interrater agreement (Fleiss'  $\kappa$ ) was substantial for Sample 1 ( $\kappa = .66$  for the research assistant codings, excluding the incorrect data;  $\kappa = .75$  for the author codings) and moderate for Samples 2 ( $\kappa = .52$ ) and 3 ( $\kappa = .51$ ).

### Data Analyses

All analyses were conducted using the R statistical computing environment (Version 3.5.2; [R Core Team, 2018](#)). We conducted some additional text cleaning for our analyses and visualizations. Specifically, we (a) removed qualifiers (e.g., “sometimes,” “a bit,” “too”), (b) changed different forms of the same descriptor into the adjective form (e.g., “sarcasm” was changed into “sarcastic”) or a noun or short phrase if an adjective was not possible (e.g., “good listener,” “easily influenced”), (c) kept only the first descriptor if two descriptors were mentioned (e.g., “can be inconsiderate/self-centered” was summarized as “inconsiderate”), and (d) summarized long-winded responses as an adjective, noun, or short phrase (e.g., “easily gets down on self after failure” was summarized as “easily discouraged”). These cleaned responses were used to quantify and visualize the frequency of best and worst trait descriptors. None of these procedures altered what coders saw in the tasks described above (i.e., coders always saw participants’ spellchecked, lowercase, and deidentified but otherwise original responses).

## Results

### Size of Vocabulary and Most Frequent Trait Descriptors

T1 As shown in [Table 1](#), people had a larger vocabulary for worst traits, compared with best traits. Pooling across the three samples and self- and friend-reports, the number of unique worst trait descriptors ( $n = 442$ ) was 1.69 times greater than the number of best trait descriptors ( $n = 261$ ). This is consistent with past work which has shown that there are a greater number of negative than positive person descriptors ([Chandler, 2018](#)) and supports the principle that “good is more alike than bad” ([Alves et al., 2017b](#)). Interestingly, friends (compared with targets) also used 1.33 times as many unique descriptors to describe the targets’ worst traits ( $n = 342$  vs.  $n = 258$ , pooling across the three samples). Consistent with Zipf’s law ([Zipf, 1936](#); for a review, see [Piantadosi,](#)

**Table 1**  
*Number of Unique Best and Worst Trait Descriptors*

Sample	Sample size	Best traits	Worst traits	Maximum possible
Sample 1	368	160	239	736
Self	184	121	142	368
Friend	184	103	171	368
Sample 2	242	135	224	484
Self	121	85	115	242
Friend	121	100	158	242
Sample 3	318	150	251	636
Self	159	98	125	318
Friend	159	111	183	318

Note. Values in boldface represent the total number of unique best or worst trait descriptors across self- and friend-reports. Sample size = number of participants; Maximum possible = maximum possible number of best or worst trait descriptors if all participants had each reported two unique descriptors for the respective trait type.

2014), a small number of descriptors were used extremely often (relative to other descriptors), whereas most descriptors were rarely used (see [Figure S1](#) in the online supplemental materials).

[Figure 1](#) provides an at-a-glance impression of the most common best and worst trait descriptors that targets and their friends reported, pooled across all three samples. [Table S2](#) in the online supplemental materials provides a more detailed breakdown for each sample, as well as an initial exploration of self–other differences in the trait descriptors that were reported. These show that there were several consistencies across self- and friend-reports (see [Figure 1](#) and [Table S2](#) in the online supplemental materials) and across the samples (see [Table S2](#) in the online supplemental materials) in the most frequent best trait descriptors. For example, targets and friends both frequently mentioned “caring,” “kind,” “friendly,” “funny,” “hardworking,” “loyal,” “outgoing,” and “honest” as best traits. There were fewer self–other consistencies on worst trait descriptors, but targets and friends both frequently mentioned “lazy,” “stubborn,” and “insecure.”

[Figure 1](#) and [Table S2](#) in the online supplemental materials also highlight the largest differences in the frequencies with which a descriptor was considered to be a best or worst trait by the self or by a friend. Several of these self–other differences replicated across at least two samples. For example, targets (compared with friends) more frequently mentioned “empathetic,” “honest,” “compassionate,” “optimistic,” and “friendly,” whereas friends (compared with targets) more frequently mentioned “supportive,” “funny,” “trustworthy,” “loyal,” and “kind” as best traits. For worst traits, targets (compared with friends) more frequently mentioned “shy,” “sensitive,” “lazy,” “anxious,” “indecisive,” “insecure,” and “emotional,” whereas friends (compared with targets) more frequently mentioned “loud,” “inconsiderate,” “stubborn,” and “competitive.”

### Which Broad Trait Categories Do Best and Worst Traits Reflect?

Our first goal was to examine how frequently people mentioned best and worst traits that reflected the high or low poles of each of the Big Five or Six personality trait domains. As shown in [Figure 2](#), best traits most frequently reflected high agreeableness and extraversion, whereas worst traits most

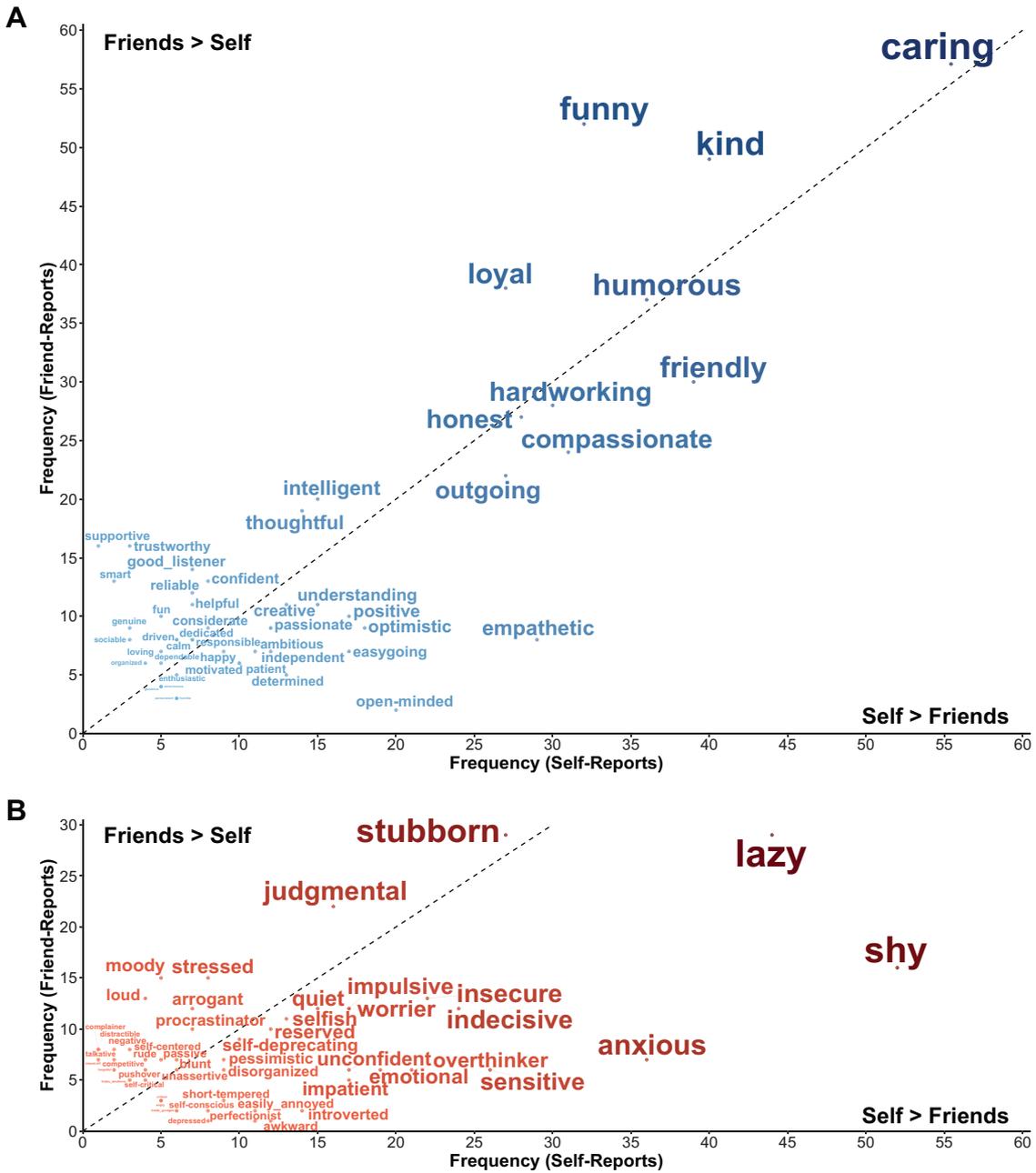
AQ: 16

F1

F2

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Figure 1**  
 Scatterplots of the Top 50 Most Frequent Best (Panel A) and Worst Trait Descriptors (Panel B) Reported by the Self and by Friends, Pooled Across All Three Samples



*Note.*  $x$  and  $y$  axes represent the number of times each descriptor was mentioned by the self or by friends, respectively. Larger and darker words correspond to more frequently mentioned descriptors. Descriptors below the diagonal were more frequently mentioned by the self than by friends; descriptors above the diagonal were more frequently mentioned by friends than by the self. Please zoom in to read less frequently mentioned descriptors. See the online article for the color version of this figure.

frequently reflected low emotional stability—especially for targets’ self-reports.

Notably, whereas best traits almost exclusively reflected the socially desirable poles of the Big Five or Six, worst traits typically reflected their socially undesirable poles but also occasionally reflected the socially desirable poles of these traits. As shown

in Tables S3–S5 in the online supplemental materials, these describe the potential downsides of having high levels of extraversion (e.g., “loud,” “overpowering,” “attention-seeker”; Samples 1–3: 3.8/5.37/3.46%), conscientiousness (e.g., “busy,” “perfectionist,” “overcommitted”; Samples 1–3: 1.63/6.4/4.09%), agreeableness (e.g., “conflict-avoidant,” “pushover,” “people-pleasing”;

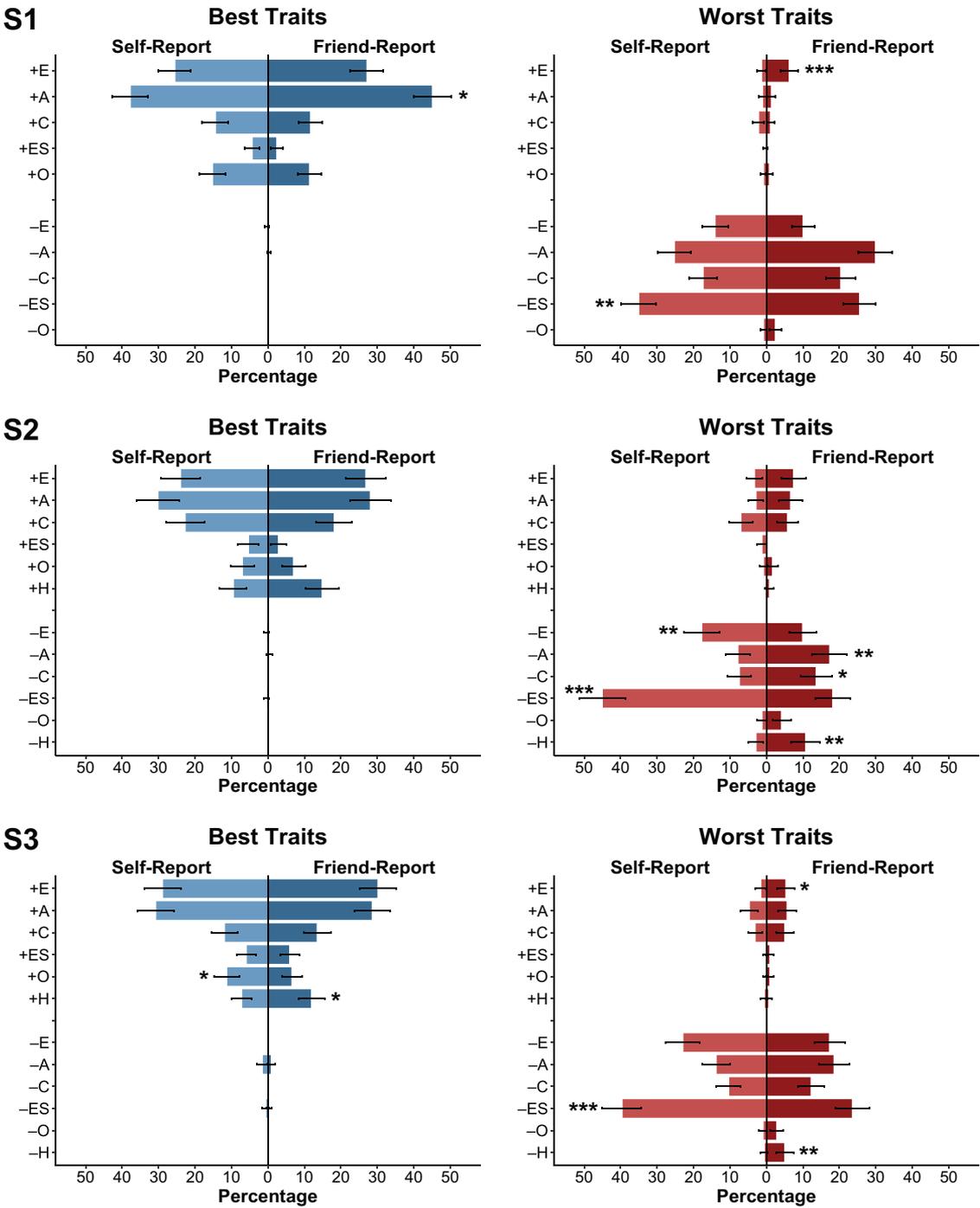
This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

OC  
 LL  
 IO  
 NR  
 E

AQ: 15

BEST AND WORST TRAITS

**Figure 2**  
*The Percentage of Self- or Friend-Reported Best or Worst Traits That Reflected the High (+) or Low (-) Poles of Each of the Big Five (Sample 1, Panel S1) or Big Six traits (Sample 2–3, Panels S2–S3)*



*Note.* Sample sizes were 183 for Sample 1, 121 for Sample 2, and 159 for Sample 3. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$  for the differences between self- and friend-reports. Asterisks are on the side with a larger percentage. The percentages were computed separately for self- and friend-reports, and the denominator included all responses for each perspective (including those that could not be coded as being a Big Five or Six trait). Error bars depict 95% confidence intervals. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

ONLINE

Samples 1–3: 1.22/4.75/5.19%), openness (e.g., “overanalytical,” “eccentric,” “escapist”; Samples 1–3: 0.82/1.24/0.63%), emotional stability (e.g., “unemotional,” “nonchalant,” “robotic”; Samples 1–3: 0.14/0.62/0.63%), and honesty-humility (e.g., “self-righteous,” “honest,” “truthful”; Samples 2–3: 0.41/0.63%).

**Self–Other Asymmetries in Best and Worst Traits’ Broad Trait Categories**

Next, we explored whether targets and their friends differed in their relative emphasis on each of the broad trait categories. We used the McNemar test (with continuity correction)—which accounts for the dependency between the self- and friend-reports—to compare the difference in proportions between self- and friend-reports for each Big Five or Six trait category. As shown in Figure 2, there were relatively few self–other asymmetries in which broad traits were mentioned as best traits, and none of these asymmetries replicated across samples.

There were a greater number of self–other asymmetries in which broad traits were mentioned as worst traits. The most replicable effect was that targets were significantly more likely than friends to mention low emotional stability worst traits across all samples. There was also some replicable evidence that friends were significantly more likely than targets to mention low honesty-humility (Samples 2 and 3) and high extraversion worst traits (Samples 1 and 3). In other words, friends seemed to be more focused on traits that are observable in the context of social interactions and that are more likely to have consequences for them.

**Self–Other Agreement on Best and Worst Traits**

In the section above, we examined whether, on average, the content of best and worst traits (i.e., which Big Five or Six domain best and worst traits reflected) tended to be similar across self- and friend-reports of best and worst traits (i.e., self–other asymmetries). Here, we turn our attention to pairs of targets and friends

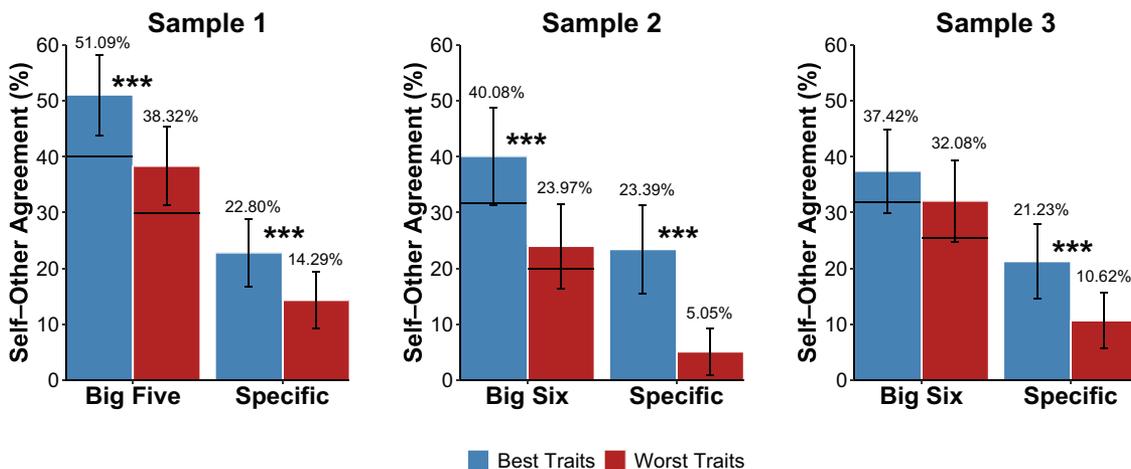
and examine how often targets and their friends agreed on each target’s best and worst traits. We examine agreement on the broad trait (i.e., Big Five or Six) categories the traits reflected, as well as the specific traits that were described by targets and friends. We also examine agreement separately for best traits and worst traits. Thus, each participant had four self–other agreement scores. Each of these scores had possible values of 0 (targets and friends agreed on neither of the two traits), .5 (targets and friends agreed on one of the two traits), and 1 (targets and friends agreed on both of the two traits). We report these results in a percentage metric for ease of interpretation.

As shown in Figure 3, targets and their friends agreed on the Big Five or Six category of targets’ best traits about 40% of the time and worst traits about 30% of the time. This could be considered high or low depending on one’s expectations; 50% agreement on best (worst) traits would imply that on average, friends listed one best (worst) trait that was in the same broad trait category as one of the two best (worst) traits that the target listed. Unsurprisingly, self–other agreement was descriptively lower for specific traits. Here, targets and their friend described the same (or a very similar) trait for 20% of best traits and 10% of worst traits. We think that 20% self–other agreement on specific best traits is quite impressive considering that targets and their friends were each given two blank boxes.

To provide a statistical test of these patterns, we ran a linear mixed effect model predicting self–other agreement from the breadth (broad vs. specific) and valence (best vs. worst) of the trait, with random intercepts for the target. Across all three samples, self–other agreement was significantly and substantially higher for broad traits compared with specific traits (Sample 1:  $b = 26.18\%$ , 95% CI [22.03%, 30.32%]; Sample 2:  $b = 17.93\%$ , 95% CI [12.93%, 22.92%]; Sample 3:  $b = 19.03\%$ , 95% CI [14.6%, 23.45%]; all  $ps < .001$ ). Self–other agreement was also significantly higher for best traits compared with worst traits, and these differences were small-to-medium in magnitude (Sample 1:  $b =$

F3

**Figure 3**  
*Self–Other Agreement for Best and Worst Traits, Separately for Broad Trait Categories and Specific Traits*



*Note.* Error bars depict 95% confidence intervals. Horizontal lines indicate the extent of agreement by chance estimated using randomization tests (only conducted for the broad trait categories). \*\*\*  $p < 0.001$  for the difference in self–other agreement between best and worst traits. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

ONLINE

10.66%, 95% CI [6.52%, 14.79%]; Sample 2:  $b = 17.17%$ , 95% CI [12.22%, 22.13%]; Sample 3:  $b = 7.87%$ , 95% CI [3.47%, 12.27%]; all  $ps < .001$ ).

Next, we conducted paired-samples  $t$ -tests to examine differences in self–other agreement between best and worst traits, separately for broad traits or specific traits. As shown in Figure 3, for broad traits, self–other agreement was significantly higher for best traits than worst traits in Samples 1 and 2 (Sample 1:  $M_{\text{dif}} = 12.77%$ , 95% CI [6.17%, 19.38%]; Sample 2:  $M_{\text{dif}} = 16.12%$ , 95% CI [7.96%, 24.27%]; both  $ps < .001$ ), but the difference was smaller and not statistically significant in Sample 3 ( $M_{\text{dif}} = 5.35%$ , 95% CI [–2.11%, 12.8%],  $p = .159$ ). For specific traits, self–other agreement was significantly higher for best traits than worst traits across all three samples (Sample 1:  $M_{\text{dif}} = 8.52%$ , 95% CI [2.57%, 14.46%]; Sample 2:  $M_{\text{dif}} = 18.35%$ , 95% CI [11.55%, 25.14%]; Sample 3:  $M_{\text{dif}} = 10.62%$ , 95% CI [4.68%, 16.55%]; all  $ps < .001$ ).

### Randomization Tests for Broad Traits

One possibility is that the relatively high level of self–other agreement on the broad trait categories is driven not by friends uniquely agreeing with targets on what their best and worst traits are, but rather by base rates (i.e., normativeness effects; see Furr, 2008). For example, agreement could be high by chance because both targets and friends more frequently mention high agreeableness best traits than high openness best traits. To examine the magnitude of such normativeness effects, we conducted randomization tests (Block, 1960; Sherman & Serfass, 2015). We created 1,000 pseudosamples in which we matched each target with the friend of a *different* randomly-selected target. Then, we computed the self–other agreement on best and worst traits from each of the 1,000 pseudosamples. The proportion of self–other agreement estimates that exceeded the level of self–other agreement observed in the real data can be interpreted as a  $p$ -value (i.e., the probability of observing, in random data, a level of agreement that was at least as strong as was observed in the real data).

As shown in Figure 3, the average level of self–other agreement observed across 1,000 pseudosamples (i.e., agreement driven purely by normativeness) was substantial; however, in all cases, the extent of self–other agreement observed in the real data was significantly greater than that observed by chance. Specifically, for best traits, the average self–other agreement observed by chance was 40.01% for Sample 1 (vs. 51.09% in the real data,  $p < .001$ ), 31.61% for Sample 2 (vs. 40.08% in the real data,  $p < .001$ ), and 31.82% for Sample 3 (vs. 37.42% in the real data,  $p = .005$ ). For worst traits, the average self–other agreement observed by chance was 29.96% for Sample 1 (vs. 38.32% in the real data,  $p < .001$ ), 19.99% for Sample 1 (vs. 23.97% in the real data,  $p = .021$ ), and 25.42% for Sample 3 (vs. 32.08% in the real data,  $p < .001$ ). In other words, even though agreement due to normativeness was quite substantial, the observed self–other agreement was still 6% to 11% greater for best traits and 4% to 8% greater for worst traits compared with what we would expect from base rates alone.

These results also explain the higher levels of self–other agreement on best traits compared with worst traits. Because best traits almost exclusively reflected the socially desirable poles of the broad trait domains, there were essentially only five or six possible types of answers (with some types of answers being more likely

than others). In contrast, worst traits were far more diverse; because worst traits reflected both poles of each trait domain, this essentially doubles the number of possible types of answers and substantially reduces the likelihood of agreement due to chance. This raises the possibility that the greater self–other agreement on best traits versus worst traits may simply be explained by the greater diversity of worst traits. Indeed, after taking normativeness effects into account, the differences in self–other agreement on best versus worst traits (for the broad trait categories) were only 2.72%, 95% CI [–2.99%, 8.15%] (Sample 1), 4.49%, 95% CI [–2.07%, 10.33%] (Sample 2), and –1.06%, 95% CI [–7.23%, 4.41%] (Sample 3), and none of these differences were detectable. Thus, these results suggest that the higher levels of self–other agreement on best traits' broad trait categories can be entirely explained by the fact that worst traits were less predictable from base rates.

We did not conduct randomization tests for self–other agreement on the specific traits because we used human coders to judge agreement for each pairwise comparison for the specific traits. Conducting the same randomization tests would require repeating the coding task 1,000 times. Because people reported a large number of trait descriptors (see Table 1), it is unlikely that self–other agreement on the specific traits can be completely explained by base rate effects. However, because people used a greater number of unique descriptors to describe worst traits than best traits (see Table 1), it is likely that as for the broad traits, the differences in self–other agreement between best versus worst traits for specific traits are at least in part due to differences in base rates. Indeed, supplemental analyses showed that the relative frequency of the descriptor explained 20% to 37% of the greater self–other agreement on specific best traits in Samples 1 and 3 (see Table S13 in the online supplemental materials); however, even after accounting for these base rate effects, there was still greater self–other agreement on best traits than worst traits. One limitation of this analysis is that the relative frequency of a specific descriptor is a noisy estimate of the relative frequency of the underlying trait (e.g., someone who is humorous can be described as either “funny” or “hilarious,” among other possibilities). However, this provides some evidence to suggest that the greater self–other agreement on specific best traits cannot be entirely explained by differences in base rates.

We also collected trait rating data to examine whether the greater self–other agreement on best traits than worst traits could be explained by differences in relative personal importance, breadth, observability, or perceived base rates, and found little evidence for these possibilities (see Table S13 in the online supplemental materials).

## Discussion

What do people think their best and worst personality traits are? Do their friends agree? To answer these questions, we asked U.S. college student targets and their friends to tell us, in their own words, what the targets' best and worst traits are. The key findings, along with their replicability across three samples, are summarized in Table 2. Targets and friends most frequently reported agreeableness- and extraversion-related best traits. People had more variable concepts of worst traits, which spanned both the socially undesirable and socially desirable poles of the Big Five or Six traits—especially low emotional stability. Overall, the kinds of traits that people most liked in themselves were similar to those

T2

**Table 2**  
*Summary and Replicability of Key Findings*

Finding	Sample 1	Sample 2	Sample 3
Descriptively most prevalent broad trait category			
Best traits	+A	+A	+A
Worst traits	-ES	-ES	-ES
Self-other asymmetries			
Best traits			
+A: Friends > Self	✓	?	?
+H: Friends > Self		?	✓
+O: Self > Friends	?	?	✓
Worst traits			
-ES: Self > Friends	✓	✓	✓
-H: Friends > Self		✓	✓
+E: Friends > Self	✓	?	✓
-E: Self > Friends	?	✓	?
-A: Friends > Self	?	✓	?
-C: Friends > Self	?	✓	?
Self-other agreement			
Broad trait categories > specific traits	✓	✓	✓
Best traits > worst traits			
For broad trait categories			
Overall agreement	✓	✓	?
Above chance agreement	?	?	?
For specific traits			
Overall agreement	✓	✓	✓
Agreement controlling for relative frequency	✓	✓	✓

*Note.* Effects are ordered from most to least replicable, within each group of effects. ✓ indicates that the effect was detectable at  $p < .05$  in the respective sample. ? indicates that there was no detectable evidence for the effect in the respective sample. Blank cells indicate effects that are not applicable (specifically, honesty-humility was not coded in Sample 1).

traits that their friends most liked in them, but there were some notable self-other asymmetries in perceptions of worst traits. These results complement existing studies of trait desirability in three ways. First, they clarify which broad trait domains tend to be most evaluatively polarized in practice (i.e., when considering the best and worst traits of actual people), rather than in the abstract. Second, by examining both self- and friend-perspectives, these results clarify for whom a given trait is more desirable. Finally, we show that even the “desirable” pole of each domain has some downsides, such that these generally desirable poles are not “purely” positive. We discuss each of these implications in turn.

**Evaluativeness in the Real World**

Previous studies of the evaluativeness of major trait domains have asked raters to judge the desirability of a predefined set of traits in the abstract, and have produced conflicting conclusions about which of the Big Five domains are particularly evaluative (John & Robins, 1993; Kim et al., 2019). However, the traits that are judged to be highly evaluative in the abstract may not be the same traits that people like or dislike the most about themselves or their friends. This could be because the traits rated as most evaluative in the abstract are rarely present in typical populations, or because their manifestation is so rare that they do not spontaneously come to mind when describing oneself or one’s friends. Moreover, predefined sets of researcher-selected descriptors may exclude many descriptors that people would naturally use to

describe their own and their friends’ best and worst traits. By soliciting open-ended self- and friend-perceptions of best and worst traits and mapping them onto the high and low poles of the Big Six, our results reveal which of these major trait domains most frequently capture the descriptors that are used to evaluate real people. Of course, by only including self-reports and reports by close friends, we are likely missing some of the most undesirable characteristics that might only be mentioned by enemies (for example). Thus, these findings complement rather than overrule the findings based on abstract ratings of traits by showing how evaluativeness plays out in two important real-world contexts: self-perceptions and close friendships.

To explore this issue, we conducted supplemental analyses to summarize which of the Big Six domains best capture the open-ended evaluative descriptions of best and worst traits in college students and their friends. We operationalized the evaluativeness of a domain as the overall percentage of desirable traits (i.e., the percentage of best traits minus the percentage of worst traits) that fell within the high pole of that domain plus the overall percentage of undesirable traits (i.e., the percentage of worst traits minus the percentage of best traits) that fell within the low pole of that domain. As shown in Figure 4, agreeableness captured the most evaluatively polarized content in our participants’ responses (consistent with John & Robins, 1993), whereas openness was the least evaluative domain (consistent with Kim et al., 2019; but very different from John & Robins, 1993). Extraversion and emotional stability were the next two most evaluative domains after agreeableness. This was in contrast to previous studies, which tended to find that extraversion and emotional stability were less evaluative than other trait domains (John & Robins, 1993; Kim et al., 2019). Finally, conscientiousness and honesty-humility were less-frequently mentioned as best and worst traits.

Interestingly, however, when evaluativeness is conceptualized in line with previous research as the difference in desirability ratings between the high and low poles of each domain (see the online supplemental materials, Section 7, for details of the trait rating study), different conclusions emerge that are more consistent with results from previous research. As shown in Figure 4, extraversion and emotional stability are the least evaluative domains (in line with John & Robins, 1993). The prosocial traits of agreeableness and especially honesty-humility (Samples 2 and 3) were the most evaluative domains, followed by openness. In other words, although openness and honesty-humility were the two domains that were least represented among the best and worst traits, when such traits were mentioned, they tended to be more evaluatively extreme (as rated by a separate group of raters) than traits within more frequently represented domains (e.g., extraversion and emotional stability). This points to the disconnect between the trait domains that are most desirable or undesirable in the abstract and the trait domains that best capture peoples’ real-world evaluations of their own and their friends’ best and worst qualities.

**Sources of Self-Other Agreement on Best and Worst Traits**

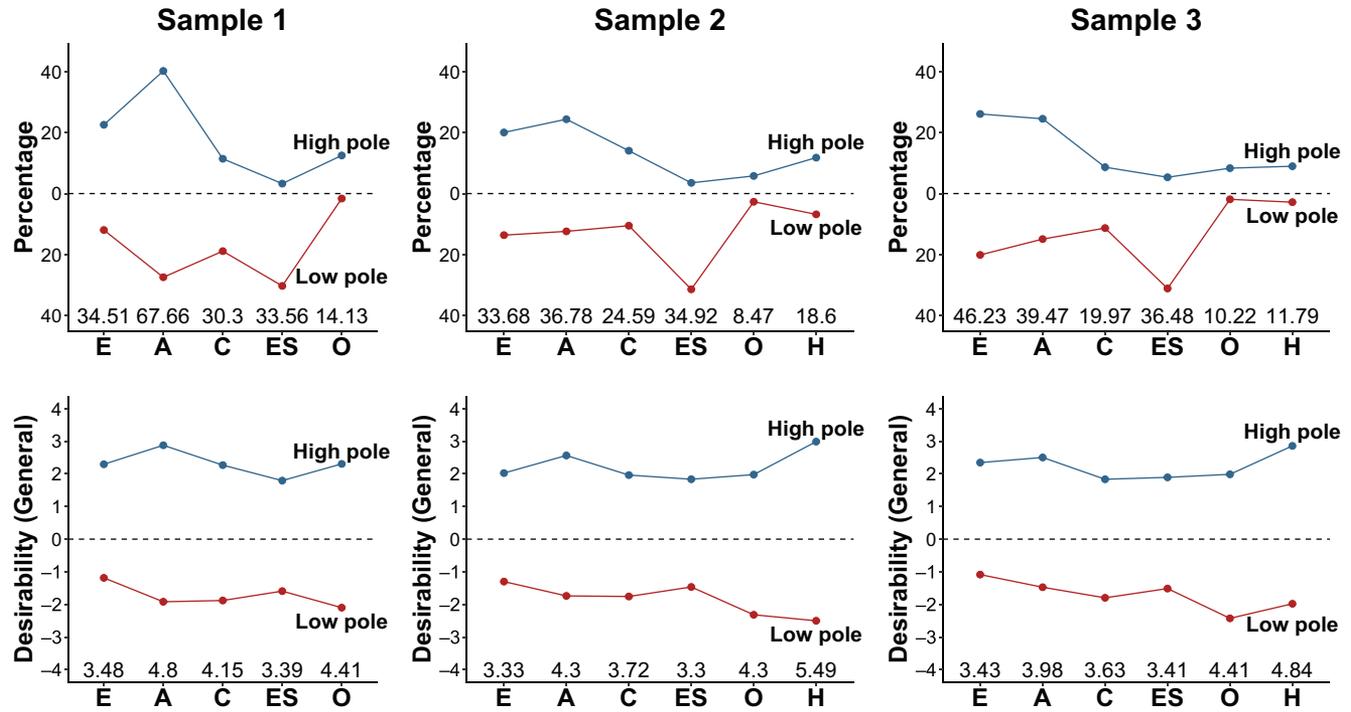
The current findings also provide insight into the extent to which the desirability of a trait is in the eye of the beholder. Targets and their friends agreed about which broad trait categories targets’ best and worst traits reflected about one third of the time, but

F4

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Figure 4**

*Evaluativeness of the Big Five (Sample 1) or Big Six (Samples 2–3) Domains Based on Which Trait Categories Most Frequently Captured People’s Best or Worst Traits (Top Row) and the Mean General Desirability Ratings of Traits in Each Category (Bottom Row)*



*Note.* Top row: “High pole” denotes the overall percentage of desirable traits (i.e., the percentage of best traits minus the percentage of worst traits) that fell within the high pole of each domain; “Low pole” denotes the overall percentage of undesirable traits (i.e., the percentage of worst traits minus the percentage of best traits) that fell within the low pole of each domain. Traits were pooled across self- and friend-reports. Bottom row: “High pole” and “Low pole” denote the mean general desirability ratings of traits that fell within the high or low pole of each domain, respectively. Traits were pooled across best and worst traits and self- and friend-reports. Both rows: Trait domains that have a greater distance along the y-axis between the high and low poles (see overlaid numbers) are more evaluative. See the online article for the color version of this figure.

much (though not all) of this agreement was driven by base rates (both targets and friends tended to mention best or worst traits in some Big Six categories more often than others). Agreement about which specific traits were the targets’ best or worst traits was naturally lower, at around 20% for best traits and 10% for worst traits. Given that targets and their friends had free rein to describe the target’s best and worst traits however they wanted, we find this level of agreement at the specific trait level quite impressive, particularly for best traits. Overall, targets and friends tended to report similar kinds of best traits. However, targets consistently focused more on low emotional stability worst traits than did friends, whereas friends focused more on low prosociality worst traits than did targets (in Samples 2 and 3).

Why might your friends agree or disagree with you about what your best and worst traits are? Conceptually, best and worst traits involve both (a) a *descriptive* judgment that the person has that trait and (b) an *evaluative* judgment of whether the trait can be considered a “best” or “worst” trait. Thus, self–other agreement versus asymmetries occur to the extent that the self and others agree versus disagree on (a) whether the target has a given trait or (b) whether that trait can be considered to be a best or worst trait.

The self–other knowledge asymmetry (SOKA) model (Vazire, 2010) provides a framework for understanding self–other asymmetries in descriptive judgments. This model posits that the self has more knowledge about less observable traits (e.g., emotional stability), whereas well-acquainted others have more knowledge about highly evaluative traits (e.g., prosocial traits) because other-ratings are less susceptible to ego-protective biases. These differences in perspectives may explain why people are more likely to describe themselves as having low emotional stability worst traits (which tend to be less observable to others), whereas friends are more likely to describe others as having prosocial best traits or antisocial worst traits (which tend to be more evaluative but also fairly observable). These findings parallel the lower levels of self–other agreement on emotional stability and agreeableness observed in other studies (Connelly & Ones, 2010; John & Robins, 1993; Sun & Vazire, 2019).

What makes a trait stand out as being a “best” or “worst” trait, out of the many traits that a person has? We suspect that the answers may be somewhat different for best traits and worst traits. One initial possibility is that people simply consider which traits the target has especially high or low levels of, compared with other people. Supplemental analyses predicting best and worst traits

from questionnaire measures of the Big Five or Six suggest that this is only part of the picture (see the [online supplemental materials](#), Section 8). For example, although people who reported having higher levels of extraversion were more likely to report a high extraversion best trait (e.g., “outgoing”), many extreme extraverts did not mention a high extraversion best trait. Similarly, although friends who judged their target as having lower levels of extraversion were more likely to report a low extraversion worst trait (e.g., “quiet”), many extreme introverts’ friends did not mention a low extraversion worst trait. This suggests that best and worst traits do not merely capture trait standings.

Relatedly, friends might be somewhat more likely to report best and worst traits that reflect their own traits. For example, if friends are high in agreeableness, they may be especially likely to report a high agreeableness best trait or low agreeableness worst trait in their target. If so, to the extent that best friends (Sample 1) or first-listed friends (Samples 2 and 3) have systematically higher levels of agreeableness or honesty-humility than their targets, this could help to explain the self–other asymmetries on prosociality-related traits.

Beyond trait standings, we suspect that when people are judging a person’s best traits, they likely consider the qualities that they most value in themselves or admire in others. For example, even if a person is exceptionally creative, they (or their friends) may still say that their best trait is kindness because they value kindness more than creativity. In many cases, people may specifically be thinking about the traits that make themselves or their friends “good people.” Supporting this interpretation, many of the most frequently-mentioned best traits—“caring,” “compassionate,” “honest,” “kind,” “loyal,” “trustworthy” (see [Table S2](#) in the [online supplemental materials](#))—are clear indicators of moral character ([Goodwin et al., 2014](#)). Thus, the self–other similarities in best traits, which heavily emphasize sociability and prosociality, are consistent with the well-established finding that communal or moral traits are more valued than agentic traits for both the self and for others ([Abele & Wojciszke, 2007](#); [Goodwin et al., 2014](#); [Wortman & Wood, 2011](#)). The reverse may also be true to some extent—that worst traits capture the ways in which people most feel like they are failing to live up to their deepest values or are “bad people.”

A further possibility is that targets and friends may both be evaluating which of the targets’ traits are instrumentally best or worst for *themselves* (i.e., the judge). That is, targets may focus more on which of their traits are beneficial or detrimental for their own well-being, whereas friends may focus more on which of the targets’ traits are most helpful/endeoring or harmful/annoying in the context of their relationship with the target. Supporting this possibility, supplemental analyses showed that on average, the traits that were mentioned by targets (vs. friends) were rated as being more desirable or undesirable for the self than for a friend (see [Table S11](#) in the [online supplemental materials](#)).

Such a perspective-dependent evaluative standard may explain both the self–other similarities on best traits and the self–asymmetries on worst traits. That is, the emphasis by both the self and friends on sociable and prosocial best traits may reflect the fact that these traits are beneficial for both personal well-being and friendships ([Anglim et al., 2020](#); [Sun et al., 2018](#); [Wilson et al., 2015](#)). Partially supporting this idea, supplemental analyses (see [Figure S2](#) and [Table S14](#) in the [online supplemental materials](#))

showed that high extraversion and agreeableness traits were generally rated as being similarly desirable for the self and for a friend. Consistent with the finding that friends more frequently mentioned high honesty-humility traits than did the self (in Sample 3), high honesty-humility traits were also rated as being more desirable for a friend than for the self (in both Samples 2 and 3).

Likewise, the various self–other asymmetries in worst traits might arise from these traits’ divergent consequences for the self compared with others. For example, low emotional stability has a bigger impact on one’s well-being than on one’s friendships, whereas low prosociality might have more negative consequences for others than for the self. Supporting this view, supplemental analyses (see [Figure S2](#) and [Table S14](#)) showed that low agreeableness (Samples 1 and 2) and low honesty-humility (Samples 2 and 3) traits were rated as being less desirable for a friend than were low emotional stability traits, whereas low emotional stability traits were rated as being less desirable for the self than were low agreeableness traits (Samples 1 and 3). These self–friend asymmetries in the relative undesirability of low emotional stability and low prosociality traits may help explain targets’ and friends’ respective emphases on worst traits that reflected these domains. More generally, these results also suggest that the desirability of a trait is at least partly in the eye of the beholder; thus, instead of asking whether a trait is desirable or undesirable in the abstract, we may want to ask, for *whom* is this trait desirable or undesirable?

### Mixed Blessings of Best and Worst Traits

Our results also contribute to a deeper understanding of the mixed personal and social consequences of each of the Big Six traits. On the one hand, our results continue to show that each of the Big Six has a generally desirable pole. Across all samples, best traits almost exclusively reflected high levels of extraversion, agreeableness, conscientiousness, emotional stability, openness, and honesty-humility. Although people occasionally mentioned worst traits that reflected high levels of the Big Six, they much more frequently mentioned worst traits that reflected insufficient levels of these trait domains. For example, being disagreeable seems to be a much more common problem than being too agreeable.

However, our finding that a nontrivial percentage of worst traits also reflected the desirable poles of these traits supports the views that (a) having too little *or* too much of a good thing can both be seen as weaknesses ([Grant & Schwartz, 2011](#); [Kuncel & Tellegen, 2009](#)) and (b) the same underlying trait can be interpreted in either a favorable or problematic light ([Peabody, 1967](#)). To further explore the mixed blessings that students and their friends spontaneously express about various personality traits, we conducted supplemental analyses of cases in which a broad trait is expressed as both a best trait and a worst trait within the same person (see [Table S6](#)). Approximately 8% of targets self-reported a worst trait that was in the same broad trait category as one of their best traits. For example, one target said that they were “perseverant” (best trait) but also “unspontaneous” (worst trait), which are both manifestations of high conscientiousness; another target said that they were “kind” (best trait) but also a “pushover” (worst trait), which are both manifestations of high agreeableness. This suggests that

people recognize that their own best traits can be double-edged swords.

These mixed blessings were also expressed *between* self- versus other-perspectives. Approximately 10% of friends reported a worst trait that was in the same broad trait category as a best trait that the target reported, and approximately 5% of friends reported a best trait that was in the same broad trait category as a worst trait that the target reported. For example, one target reported that they were “spunky” (best trait) and their friend said they were “loud” (worst trait), which are both manifestations of high extraversion; another target said that they were an “overachiever” (worst trait) and their friend said they were “driven” (best trait), which are both manifestations of high conscientiousness. These represent extreme cases of self–other disagreement, in which the self and friends disagree on whether the same broad trait is good or bad *in the same person*. By using a free-response format instead of focusing on a predefined set of traits, these results generate new hypotheses about how the same broad trait can be expressed in ways that are simultaneously good and bad.

### Limitations and Future Directions

The main limitation of this study is that we cannot be sure that participants were being completely honest when reporting on their own or their friends’ best and worst traits. Because of self-presentational and humility concerns, people may not be willing to admit to having especially negative worst traits or especially positive best traits. Despite assurances of confidentiality, friends might also refrain from reporting overly-negative traits because they feel that it would be mean or disloyal. Indeed, we noticed that friends would frequently preface their descriptions of the targets’ worst traits with qualifiers (e.g., “sometimes,” “occasionally,” “can be”). Although friends were willing to attribute rather negative qualities to their targets (e.g., “selfish,” “arrogant,” “judgmental”), we cannot know whether they might have reported even more negative traits if they were completely candid. This is an inherent limitation of studying self- and other-perceptions, which can only be measured using self- and informant-reports.

Our findings also raise several open questions that future research might address. Just as narrower traits within broad trait domains (e.g., facets and aspects; DeYoung et al., 2007; Soto & John, 2017) have divergent consequences across a range of life domains (e.g., Kaufman et al., 2016; Sun et al., 2018; Zhao et al., 2017), best and worst traits may more frequently reflect some facets or aspects than others. For example, considering that the sociability and positive emotions facets of extraversion tend to be rated as more desirable than the dominance facet (Kim et al., 2019), it might be the case that high extraversion best traits more frequently reflect sociability and positive emotion traits, whereas high extraversion worst traits more frequently reflect dominance-related traits. Future research could shed light on this matter by coding best and worst traits into facets or aspects. Because this would involve a larger number of categories (e.g., 10 aspects or 30 facets), larger sample sizes would be needed to achieve precise estimates of the percentages of traits within each category and to detect self–other asymmetries.

Future research should also examine the real-world social consequences of disclosing and being judged as having various best and worst traits. In many job interviews, people are asked to

describe their greatest strengths and weaknesses. Similarly, in dating contexts, people are often on the lookout for green flags and red flags in their potential partners. Perceptions of best and worst traits may also inform decisions to become friends with a new acquaintance. Thus, future research should examine what kinds of best traits are associated with reputational benefits (e.g., being liked and respected) and consequential decisions in different social domains (e.g., work, romance, friendship). Such research should also explore which kinds of worst traits are most likely to be deal-breakers (e.g., predict future relationship dissolution) in different types of relationships.

### Constraints on Generality

The following findings seem likely to be broadly generalizable: (a) people have more variable concepts of worst traits compared with best traits, (b) people themselves (compared with informants) are more likely to mention low emotional stability worst traits, and (c) well-acquainted others agree with targets to some extent on the targets’ best and worst traits. In contrast, our use of U.S. college student samples and friend informants, both of which were disproportionately female, represent a constraint on generality for conclusions about relative emphases on different broad traits and self–other asymmetries represented in Figure 2. Indeed, supplemental analyses (see Figure S6 in the online supplemental materials) revealed that although men and women’s best traits were practically identical, men and women’s worst traits featured different relative emphases on various broad trait categories and different self–other asymmetries. For example, women appeared to place greater emphasis on low emotional stability worst traits in themselves than did men, whereas men placed greater emphasis on low extraversion worst traits in themselves than did women. However, both men and women more frequently mentioned low emotional stability worst traits in themselves than did their friends, providing support for the generalizability of this particular self–other asymmetry.

The content of best and worst traits may also depend on the age and culture of the targets. For example, considering that midlife adults place greater emphasis on goals to become more energetic than to become less anxious or depressed (compared with college students; Sun & Goodwin, 2020), midlife adults might also place greater emphasis on low extraversion and less emphasis on low emotional stability worst traits (compared with college students). High extraversion best traits might also be less frequently mentioned in cultures in which extraversion is less valued. Finally, the kinds of best and worst traits reported by others may especially depend on who the “other” is. For example, coworkers might place greater emphasis on conscientiousness-related traits. Perhaps most importantly, we are missing the perspective of someone who knows the target well but does not especially like them (except in cases of targets with low self-esteem). Thus, the picture that our results paint of worst traits, in particular, is incomplete; a different picture would likely emerge if we had ratings from others who dislike or are no longer close with the target (e.g., ex-friends, ex-romantic partners; see Leising et al., 2010, for a nomination method for obtaining informant reports from those who dislike a target).

## Conclusion

What are your best and worst traits? Would your friends agree? Building on the lexical hypothesis that natural person descriptors provide a path to discovering important personality characteristics, we gave college students free rein to describe their best and worst traits in their own words and compared their responses to what their friends said about them. Across three samples, friends agreed with targets to a surprising extent on what their specific best traits were. At the same time, targets and friends also showed theoretically notable self–other asymmetries in which worst traits they emphasized (low emotional stability vs. low prosociality, respectively). More broadly, our results show that the desirability of a trait may be partially in the eye of the beholder and that people intuitively recognize the mixed blessings of many traits.

## References

- Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93*(5), 751–763. <https://doi.org/10.1037/0022-3514.93.5.751>
- Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47*(1), i–171. <https://doi.org/10.1037/h0093360>
- Alves, H., Koch, A., & Unkelbach, C. (2016). My friends are all alike — the relation between liking and perceived similarity in person perception. *Journal of Experimental Social Psychology, 62*, 103–117. <https://doi.org/10.1016/j.jesp.2015.10.011>
- Alves, H., Koch, A., & Unkelbach, C. (2017a). The “common good” phenomenon: Why similarities are positive and differences are negative. *Journal of Experimental Psychology: General, 146*(4), 512–528. <https://doi.org/10.1037/xge0000276>
- Alves, H., Koch, A., & Unkelbach, C. (2017b). Why good is more alike than bad: Processing implications. *Trends in Cognitive Sciences, 21*(2), 69–79. <https://doi.org/10.1016/j.tics.2016.12.006>
- 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>
- Anglim, J., Horwood, S., Smillie, L. D., Marrero, R. J., & Wood, J. K. (2020). Predicting psychological and subjective well-being from personality: A meta-analysis. *Psychological Bulletin, 146*(4), 279–323. <https://doi.org/10.1037/bul0000226>
- Asch, S. E. (1987). *Social psychology*. Oxford University Press. (Original work published 1952)
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., Boies, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology, 86*(2), 356–366. <https://doi.org/10.1037/0022-3514.86.2.356>
- 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>
- Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150–166. <https://doi.org/10.1177/1088868306294907>
- Back, M. D., & Vazire, S. (2015). The social consequences of personality: Six suggestions for future research. *European Journal of Personality, 29*(2), 296–307. <https://doi.org/10.1002/per.1998>
- Baranski, E. N., Morse, P. J., & Dunlop, W. L. (2017). Lay conceptions of volitional personality change: From strategies pursued to stories told. *Journal of Personality, 85*(3), 285–299. <https://doi.org/10.1111/jopy.12240>
- Benet-Martínez, V., & Waller, N. G. (2002). From adorable to worthless: Implicit and self-report structure of highly evaluative personality descriptors. *European Journal of Personality, 16*(1), 1–41. <https://doi.org/10.1002/per.431>
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., . . . Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour, 2*(1), 6–10. <https://doi.org/10.1038/s41562-017-0189-z>
- Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61*(1), 27–41. <https://doi.org/10.1037/0003-066X.61.1.27>
- Bleidorn, W., Hopwood, C. J., Ackerman, R. A., Witt, E. A., Kandler, C., Riemann, R., Samuel, D. B., & Donnellan, M. B. (2020). The healthy personality from a basic trait perspective. *Journal of Personality and Social Psychology, 118*(6), 1207–1225. <https://doi.org/10.1037/pspp0000231>
- Block, J. (1960). On the number of significant findings to be expected by chance. *Psychometrika, 25*(4), 369–380. <https://doi.org/10.1007/BF02289754>
- 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., Zaltauskas, K., & Leising, D. (2009). More may be better but there may be too much: Optimal trait level and self-enhancement bias. *Journal of Personality, 77*(3), 825–858. <https://doi.org/10.1111/j.1467-6494.2009.00566.x>
- Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what’s the mechanism? (don’t expect an easy answer). *Journal of Personality and Social Psychology, 98*(4), 550–558. <https://doi.org/10.1037/a0018933>
- Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology, 38*(4), 476–506. <https://doi.org/10.1037/h0054116>
- Cawley, M. J., Martin, J. E., & Johnson, J. A. (2000). A virtues approach to personality. *Personality and Individual Differences, 28*(5), 997–1013. [https://doi.org/10.1016/S0191-8869\(99\)00207-X](https://doi.org/10.1016/S0191-8869(99)00207-X)
- Chandler, J. (2018). Likeableness and meaningfulness ratings of 555 (+487) person-descriptive words. *Journal of Research in Personality, 72*, 50–57. <https://doi.org/10.1016/j.jrp.2016.07.005>
- 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>
- DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality, 56*, 33–58. <https://doi.org/10.1016/j.jrp.2014.07.004>
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*(5), 880–896. <https://doi.org/10.1037/0022-3514.93.5.880>
- Diebels, K. J., Leary, M. R., & Chon, D. (2018). Individual differences in selfishness as a major dimension of personality: A reinterpretation of the sixth personality factor. *Review of General Psychology, 22*(4), 367–376. <https://doi.org/10.1037/gpr0000155>
- Dunlop, P. D., Telford, A. D., & Morrison, D. L. (2012). Not too little, but not too much: The perceived desirability of responses to personality items. *Journal of Research in Personality, 46*(1), 8–18. <https://doi.org/10.1016/j.jrp.2011.10.004>
- Fleeson, W., Furr, R. M., Jayawickreme, E., Meindl, P., & Helzer, E. G. (2014). Character: The prospects for a personality-based perspective on morality. *Social and Personality Psychology Compass, 8*(4), 178–191. <https://doi.org/10.1111/spc3.12094>

AQ: 9

AQ: 10

- Furnham, A., Richards, S. C., & Paulhus, D. (2013). The Dark Triad of personality: A 10 year review. *Social and Personality Psychology Compass*, 7(3), 199–216. <https://doi.org/10.1111/spc3.12018>
- 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>
- Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*, 150(4), 700–709. <https://doi.org/10.1037/xge0000920>
- Goodwin, G. P. (2015). Moral character in person perception. *Current Directions in Psychological Science*, 24(1), 38–44. <https://doi.org/10.1177/0963721414550709>
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106(1), 148–168. <https://doi.org/10.1037/a0034726>
- Grant, A. M. (2013). Rethinking the extraverted sales ideal: The ambivert advantage. *Psychological Science*, 24(6), 1024–1030. <https://doi.org/10.1177/0956797612463706>
- Grant, A. M., & Schwartz, B. (2011). Too much of a good thing: The challenge and opportunity of the inverted U. *Perspectives on Psychological Science*, 6(1), 61–76. <https://doi.org/10.1177/1745691610393523>
- Hampson, S. E., Goldberg, L. R., & John, O. P. (1987). Category-breadth and social-desirability values for 573 personality terms. *European Journal of Personality*, 1(4), 241–258. <https://doi.org/10.1002/per.2410010405>
- Hogan, R. (1996). A socioanalytic perspective on the Five-Factor Model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 163–180). Guilford Press.
- Hudson, N. W., & Roberts, B. W. (2014). Goals to change personality traits: Concurrent links between personality traits, daily behavior, and goals to change oneself. *Journal of Research in Personality*, 53, 68–83. <https://doi.org/10.1016/j.jrp.2014.08.008>
- John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality*, 2(3), 171–203. <https://doi.org/10.1002/per.2410020302>
- John, O. P., Hampson, S. E., & Goldberg, L. R. (1991). The basic level in personality-trait hierarchies: Studies of trait use and accessibility in different contexts. *Journal of Personality and Social Psychology*, 60(3), 348–361. <https://doi.org/10.1037/0022-3514.60.3.348>
- John, O. P., & Soto, C. J. (2021). History, measurement, and conceptual elaboration of the Big-Five trait taxonomy: The paradigm matures. In O. P. John, & R. W. Robins (Eds.), *Handbook of personality: Theory and research* (4th ed., pp. 35–82). Guilford Press.
- John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, 61(4), 521–551. <https://doi.org/10.1111/j.1467-6494.1993.tb00781.x>
- 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.
- Jokela, M., Airaksinen, J., Virtanen, M., Batty, G. D., Kivimäki, M., & Hakulinen, C. (2020). Personality, disability-free life years, and life expectancy: Individual participant meta-analysis of 131,195 individuals from 10 cohort studies. *Journal of Personality*, 88(3), 596–605. <https://doi.org/10.1111/jopy.12513>
- Judge, T. A., Livingston, B. A., & Hurst, C. (2012). Do nice guys—and gals—really finish last? The joint effects of sex and agreeableness on income. *Journal of Personality and Social Psychology*, 102(2), 390–407. <https://doi.org/10.1037/a0026021>
- Kaufman, S. B., Quilty, L. C., Grazioplene, R. G., Hirsh, J. B., Gray, J. R., Peterson, J. B., & DeYoung, C. G. (2016). Openness to experience and intellect differentially predict creative achievement in the arts and sciences. *Journal of Personality*, 84(2), 248–258. <https://doi.org/10.1111/jopy.12156>
- Kim, H., Di Domenico, S. I., & Connelly, B. S. (2019). Self–other agreement in personality reports: A meta-analytic comparison of self- and informant-report means. *Psychological Science*, 30(1), 129–138. <https://doi.org/10.1177/0956797618810000>
- Kuncel, N., & Tellegen, A. (2009). A conceptual and empirical reexamination of the measurement of the social desirability of items. *Personnel Psychology*, 62(2), 201–228. <https://doi.org/10.1111/j.1744-6570.2009.01136.x>
- Leary, T. (1957). *The interpersonal diagnosis of personality*. Ronald Press.
- Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment*, 25(5), 543–556. <https://doi.org/10.1177/1073191116659134>
- Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, 98(4), 668–682. <https://doi.org/10.1037/a0018771>
- Leising, D., Ostrovski, O., & Borkenau, P. (2012). Vocabulary for describing disliked persons is more differentiated than vocabulary for describing liked persons. *Journal of Research in Personality*, 46(4), 393–396. <https://doi.org/10.1016/j.jrp.2012.03.006>
- Leising, D., Scharloth, J., Lohse, O., & Wood, D. (2014). What types of terms do people use when describing an individual's personality? *Psychological Science*, 25(9), 1787–1794. <https://doi.org/10.1177/0956797614541285>
- Leising, D., Scherbaum, S., Locke, K. D., & Zimmermann, J. (2015). A model of “substance” and “evaluation” in person judgments. *Journal of Research in Personality*, 57, 61–71. <https://doi.org/10.1016/j.jrp.2015.04.002>
- McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. *American Psychologist*, 61(3), 204–217. <https://doi.org/10.1037/0003-066X.61.3.204>
- Meindl, P., & Graham, J. (2014). Know thy participant: The trouble with nomothetic assumptions in moral psychology. In H. Sarkissian & J. C. Wright (Eds.), *Advances in experimental moral psychology* (pp. 233–252). Bloomsbury. <https://doi.org/10.5040/9781472594150.ch-012>
- Morey, L. C., Good, E. W., & Hopwood, C. J. (2020). Global personality dysfunction and the relationship of pathological and normal trait domains in the DSM-5 alternative model for personality disorders. *Journal of Personality*. Advance online publication. <https://doi.org/10.1111/jopy.12560>
- Muthén, L., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Author.
- Nettle, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist*, 61(6), 622–631. <https://doi.org/10.1037/0003-066X.61.6.622>
- Ng, V., & Tay, L. (2020). Lost in translation: The construct representation of character virtues. *Perspectives on Psychological Science*, 15(2), 309–326. <https://doi.org/10.1177/1745691619886014>
- Norman, W. T. (1967). *2,800 personality trait descriptors: Normative operating characteristics for a university population*. Department of Psychology, University of Michigan.
- Orford, J. (1986). The rules of interpersonal complementarity: Does hostility beget hostility and dominance, submission? *Psychological Review*, 93(3), 365–377. <https://doi.org/10.1037/0033-295X.93.3.365>
- Peabody, D. (1967). Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology*, 7(4, Pt.2), 1–18. <https://doi.org/10.1037/h0025230>
- Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. Oxford University Press.
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21(5), 1112–1130. <https://doi.org/10.3758/s13423-014-0585-6>
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

AQ: 11

AQ: 12

AQ: 17

- Roberts, B. W., Smith, J., Jackson, J. J., & Edmonds, G. (2009). Compensatory conscientiousness and health in older couples. *Psychological Science*, 20(5), 553–559. <https://doi.org/10.1111/j.1467-9280.2009.02339.x>
- Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, 5(1), 2–14. [https://doi.org/10.1207/S15327957PSPR0501\\_1](https://doi.org/10.1207/S15327957PSPR0501_1)
- Saucier, G., & Goldberg, L. R. (2001). Lexical studies of indigenous personality factors: Premises, products, and prospects. *Journal of Personality*, 69(6), 847–879. <https://doi.org/10.1111/1467-6494.696167>
- Sherman, R. A., & Serfass, D. G. (2015). The comprehensive approach to analyzing multivariate constructs. *Journal of Research in Personality*, 54, 40–50. <https://doi.org/10.1016/j.jrp.2014.05.002>
- Smith, M. B., Hill, A. D., Wallace, J. C., Recendes, T., & Judge, T. A. (2018). Upsides to dark and downsides to bright personality: A multidomain review and future research agenda. *Journal of Management*, 44(1), 191–217. <https://doi.org/10.1177/0149206317733511>
- Solomon, B. C., & Jackson, J. J. (2014). The long reach of one's spouse: Spouses' personality influences occupational success. *Psychological Science*, 25(12), 2189–2198. <https://doi.org/10.1177/0956797614551370>
- Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117–143. <https://doi.org/10.1037/pspp0000096>
- Srivastava, S. (2010). The Five-Factor Model describes the structure of social perceptions. *Psychological Inquiry*, 21(1), 69–75. <https://doi.org/10.1080/10478401003648815>
- Sun, J., Kaufman, S. B., & Smillie, L. D. (2018). Unique associations between Big Five personality aspects and multiple dimensions of well-being. *Journal of Personality*, 86(2), 158–172. <https://doi.org/10.1111/jopy.12301>
- Sun, J., & Goodwin, G. P. (2020). Do people want to be more moral? *Psychological Science*, 31(3), 243–257. <https://doi.org/10.1177/0956797619893078>
- Sun, J., & Vazire, S. (2019). Do people know what they're like in the moment? *Psychological Science*, 30(3), 405–414. <https://doi.org/10.1177/0956797618818476>
- Suzuki, T., Samuel, D. B., Pahlen, S., & Krueger, R. F. (2015). DSM-5 alternative personality disorder model traits as maladaptive extreme variants of the Five-Factor Model: An item-response theory analysis. *Journal of Abnormal Psychology*, 124(2), 343–354. <https://doi.org/10.1037/abn0000035>
- Tay, L., & Jebb, A. T. (2018). Establishing construct continua in construct validation: The process of continuum specification. *Advances in Methods and Practices in Psychological Science*, 1(3), 375–388. <https://doi.org/10.1177/2515245918775707>
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, 59(5), 1–38. <https://doi.org/10.18637/jss.v059.i05>
- 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. (2020, April 20). [Tweet] Hi personality psych twitter! Help settle a debate (personality psychologists only please!) Is the personality characteristic "doesn't stand up for." <https://twitter.com/siminevazire/status/1252398298096230402>
- Widiger, T. A., & Mullins-Sweatt, S. N. (2009). Five-factor model of personality disorder: A proposal for DSM-V. *Annual Review of Clinical Psychology*, 5(1), 197–220. <https://doi.org/10.1146/annurev.clinpsy.032408.153542>
- Wiese, C. W., Tay, L., Duckworth, A. L., D'Mello, S., Kuykendall, L., Hofmann, W., Baumeister, R. F., & Vohs, K. D. (2018). Too much of a good thing? Exploring the inverted-U relationship between self-control and happiness. *Journal of Personality*, 86(3), 380–396. <https://doi.org/10.1111/jopy.12322>
- Wilson, R. E., Harris, K., & Vazire, S. (2015). Personality and friendship satisfaction in daily life: Do everyday social interactions account for individual differences in friendship satisfaction? *European Journal of Personality*, 29(2), 173–186. <https://doi.org/10.1002/per.1996>
- Wood, D. (2015). Testing the lexical hypothesis: Are socially important traits more densely reflected in the English lexicon? *Journal of Personality and Social Psychology*, 108(2), 317–335. <https://doi.org/10.1037/a0038343>
- Wortman, J., & Wood, D. (2011). The personality traits of liked people. *Journal of Research in Personality*, 45(6), 519–528. <https://doi.org/10.1016/j.jrp.2011.06.006>
- Zhao, K., Ferguson, E., & Smillie, L. D. (2017). Individual differences in good manners rather than compassion predict fair allocations of wealth in the dictator game. *Journal of Personality*, 85(2), 244–256. <https://doi.org/10.1111/jopy.12237>
- Zipf, G. (1936). *The psychobiology of language*. Routledge.

AQ: 13

AQ: 14

Received November 18, 2020

Revision received April 12, 2021

Accepted April 23, 2021 ■

# AUTHOR QUERIES

## AUTHOR PLEASE ANSWER ALL QUERIES

**1**

AQau— Please confirm the given-names and surnames are identified properly by the colours.

■=Given-Name, ■= Surname

The colours are for proofing purposes only. The colours will not appear online or in print.

AQ1—Author: Please provide the citation for the Ref. John & Srivastava 1999

AQ2—Author: Please provide the citation for the Ref. Meindl & Graham 2014

AQ3—Author: Please provide the citation for the Ref. Muthén & B O 1998–2017

AQ4—Author: This article has been lightly edited for grammar, style, and usage. Please compare against your original document and make changes on these paged proofs. If no change is required in response to a question, please write “OK as set” in the margin.

AQ5—Author: Please be sure to provide the name of the department(s) with which you and your coauthors are affiliated at your respective institutes if you have not already done so. If you are affiliated with a governmental department, business, hospital, clinic, VA center, or other nonuniversity-based institute, please provide the city and U.S. state (or the city, province, and country) in which the institute is based. Departments should be listed in the author footnote only, not the byline. If you or your coauthors have changed affiliations since the article was written, please include a separate note indicating the new department/affiliation: [author's name] is now at [affiliation].

AQ6—Author: Borkenau et al. (2009) is not listed in your references. Please add complete publication information to the References list or delete the citation.

AQ7—Author: Leising et al. (2010) is not listed in your references. Please add complete publication information to the References list or delete the citation.

AQ8—Author: Per journal style, all journal references should include an issue number; please supply any that are missing.

AQ9—Author: Please provide the citation for the Ref. Blanton & Jaccard 2006.

AQ10—Author: Please provide the citation for the Ref. Bullock et al. 2010.

AQ11—Author: Please provide the citation for the Ref. Gomila 2020.

AQ12—Author: Please provide the citation for the Ref. Lee & Ashton 2018.

AQ13—Author: Please provide the citation for the Ref. Tay & Jebb 2018.

AQ14—Author: Please provide the citation for the Ref. Tingley et al. 2014.

AQ15—Author: Please confirm figure(s) and legend(s) as set.

AQ16—Author: Please review the typeset table(s) carefully against your original table(s) to verify accuracy of editing and typesetting.

AQ17—Author: Please provide the citation for the Ref. John et al. 1999.