Full Length Article

Evaluating the predictive validity of personality trait judgments using a naturalistic behavioral criterion: A preliminary test of the self-other knowledge asymmetry model

Andrew Beer [a],*, Simine Vazire [b]

[a] *University of South Carolina Upstate, United States*
[b] *University of California Davis, United States*

ABSTRACT

We tracked 87 participants over two days using the Electronically Activated Recorder (EAR). Coded variables included expressions of mood, amount of talking in various situations (e.g., with one other person, with a friend, etc.), locations, and behavioral markers of the Big Five. Collection of self-, informant-, and stranger-ratings on markers of the Big Five allowed for a unique test of the Self-Other Knowledge Asymmetry (SOKA) model. Although effect sizes were modest, there was evidence for the validity of both self- and informant-ratings across most trait dimensions. Stranger-ratings showed evidence of validity in the domain of Extraversion. Predictions derived from the SOKA model were partially supported, though more research with larger samples is needed to provide stronger tests of SOKA.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Personality traits are invisible – putative latent constructs that drive consistent individual differences in thought, feeling, and behavior. Thus, from a purely objective standpoint, they must generally be inferred rather than directly observed. Despite this theoretical thorniness, attempts to systematically catalog and effectively assess broad dimensions of individual differences have spanned 80 years—much of the time that psychology, as a formal discipline, has existed. The most common method of personality trait assessment has been self-report (Vazire, 2006), founded on the beliefs that individuals (a) have near complete access to their own behavior, and (b) have unique access to their mental states, including their motivations, intentions, and internal emotional states. Indeed, these advantages frequently produce circumstances under which self-report measures of personality traits predict meaningful life outcomes (Ozer & Benet-Martínez, 2006) and everyday behaviors (Mehl, Gosling, & Pennebaker, 2006).

However, the use of self-reports as a primary source of information about personality certainly leaves some things to be desired (for an extensive review of concerns with self-assessments of various sorts, see Dunning, Heath, & Suls, 2004; for a review of

self-reports of personality in particular, see Back & Vazire, 2012; Paulhus & Vazire, 2007). Although individuals have access to most of their behavior, some more automatic or unconscious aspects of behavior may go unnoticed by the actor but be quite impactful on his or her environment and thus an important aspect of one's personality. For example, evading eye contact in personal interactions is an act about which the actor could easily be unaware but may influence interaction partners quite a bit. Moreover, when forming impressions of one's own behavior, people may be overly sensitive to the thoughts and feelings they were having, and place too little emphasis on their overt actions (Andersen & Ross, 1984). Indeed, much has been written about limitations of the actor's perspective for observing his/her own behavior (Jones & Nisbett, 1971; Malle, 2006; Robins, Spranca, & Mendelsohn, 1996; Watson, 1982).

People may also have biases that distort the accuracy of their self-views, even when they have perfect access to the information they would need to form accurate self-views. There is now a fairly large body of evidence suggesting that self-enhancement is quite common (Alicke, 1985; Kwan et al., 2011; Taylor & Brown, 1988), but that there are also important individual differences in self-bias (Paulhus, 1984; Paulhus & John, 1998). Importantly, the fact that people do not all share the same direction and level of bias about themselves is even more problematic for the accuracy of self-perceptions (Vazire, 2010). If everyone self-enhanced, and did so to more or less the same degree, this would inflate the

absolute level of people's self-ratings, but the between-person rank-order accuracy of self-ratings would be intact. Instead, the fact that some people self-enhance, others self-deprecate, and still others are relatively unbiased (Bollich, Rogers, & Vazire, 2015), means that the between-person rank-ordering of people's self-views no longer matches up with the rank-ordering of their actual standing on a trait. Additionally, there are some circumstances in which individuals may have perfectly accurate self-views but may be motivated to willfully and knowingly misrepresent themselves if they believe something may be gained by doing so (e.g., assessments by current or prospective employers, dating website profiles, etc.).

For these and other reasons, psychologists will frequently turn to others for personality information regarding a given target individual. The rationale is that knowledgeable informants (e.g., friends, spouses, co-workers, roommates, etc.) have access to a wide variety of behaviors over time and across situations. Moreover, while close others certainly generate biased personality perceptions their biases tend to be more uniformly positive than self-biases and thus less disruptive of the between-person rank ordering on a given trait (Leising, Erbs, & Fritz, 2010; Leising, Gallrein, & Dufner, 2014). Indeed, informant judgments of personality do predict behavior (Connelly & Ones, 2010; Hofstee, 1994; Kolar, Funder, & Colvin, 1996). However, informants are not without their blind spots either. For example, Vazire (2010) found that friends' ratings of personality were less predictive of behavior than self-ratings for traits low in observability (e.g., neuroticism).

Since it appears that personality ratings made by the self or knowledgeable informants cannot be taken as completely valid on their own, the general temptation has been to simply aggregate across sources of data to achieve the most valid estimate of an individual's standing on a given trait domain (see Letzring, Wells, & Funder, 2006, for an extensive discussion). This logic is generally sound, and in a perfect world, we would simply collect massive amounts of data from multiple sources. However, this is a laborious process for those interested in practically applying the research findings, and in terms of theory, it seems clear that aggregation only buys predictive validity in some circumstances—in others, one source or another predicts a relevant outcome just as well on its own (Vazire & Mehl, 2008). In addition, to the extent that some of the variation across judges is non-random (e.g., a positive evaluative bias), these errors would be compounded or sustained—not eliminated—in an aggregate.

Considerations such as these require a more careful analysis of sources of personality data. Recently, Vazire (2010) put forth a general model to aid in determining which data source might be the most valid for a given trait assessment, based on previous work by John and Robins (1993) and Luft and Ingham (1955). The Self-Other Knowledge Asymmetry (SOKA) model employs two primary dimensions—which can be considered as properties of the traits themselves—to explicate the issue of source validity. The first consideration is the observability of the trait. Traits with clear, frequent, and publicly available behavioral manifestations should be judged quite accurately by knowledgeable informants. In extreme cases, such as extraversion, it is reasonable to assume that aggregating a few independent judgments from someone almost entirely unacquainted with the target individual may yield a fairly accurate estimate. On the other hand, traits defined more by internal affective or cognitive aspects, such as neuroticism, should be judged more accurately by the target herself than by others. However, these predictions must be qualified based on the second factor in the SOKA model, evaluativeness, or the extent to which the trait in question has a clearly socially desirable pole. For example, high agreeableness (warmth, compassion) is a quality generally admired by others and sought after in social relationships (Cuddy, Fiske, & Glick, 2007; Graziano & Eisenberg, 1997) and thus would be considered highly evaluative. Traits such as this should be especially susceptible to both positive and negative self-biases (i.e., by individual differences in self-enhancement/self-deprecation) and thus accuracy for self-judgments, in particular, should be impaired. Traits for which there is no clear polar preference in the population should be less susceptible to such biases, leaving self-judgments largely unaffected. Thus, knowledgeable informants should be more accurate than the self for highly evaluative traits. Overall, self-reports should be more accurate than other-reports for traits low in observability (especially if they are also low in evaluativeness) and other-reports should be more accurate than self-reports for traits high in evaluativeness (especially if they are also high in observability).

How, then, should one test these predictions? The first step is to determine where traits lie on the observability and evaluativeness continua. Happily, some data exist about this for the Big Five traits (John & Robins, 1993). These data suggest that Extraversion and Neuroticism are generally lower in evaluativeness (relative to Agreeableness, Conscientiousness, and Openness), and that Extraversion is considerably more observable than the other domains. If one were to plot the major dimensions of the Big Five along the observability and evaluativeness continua, it might resemble something like Fig. 1 (the placement of each dimension is based on Figs. 4 and 5 in John & Robins, 1993), which yields predictions for accuracy across sources as seen in Table 1. Readers may question the description of Neuroticism as low in evaluativeness, but John and Robins's data show that the two poles of this dimension are not especially far apart in social desirability. This suggests that people should not be especially afraid of being judged harshly for being high (or low) on Neuroticism, and thus motivated reasoning should not be a threat to self-reports. Moreover, Neuroticism's low level of observability is a threat to the validity of informant reports.

In order to test these predictions, Vazire (2010) presented results based on a series of laboratory tasks relevant to three primary domains (extraversion, intellect, and neuroticism) and found general support for the SOKA model in that (a) self-ratings of Neuroticism (low observability, low evaluativeness) predicted neuroticism-relevant behaviors (e.g., nervous hand movements during a speech) better than did ratings of Neuroticism by knowledgeable informants or strangers, (b) informant-rated intellect (high observability, high evaluativeness) predicted intellect/creativity-related behaviors (e.g., performance on a creativity test) better than did ratings of intellect by the self or strangers, and (c) self-, informant-, and stranger-rated Extraversion (high observability, low evaluativeness) predicted extraversion-relevant behaviors (e.g., talking) equally well.

These initial findings are certainly interesting, but as is the case with any new theory, further tests of its generalizability are warranted. Specifically, there has been a recent push to take personality and social psychology back outside the laboratory (Baumeister, Vohs, & Funder, 2007; Furr, 2009; Wilson & Vazire, 2015). Personality psychology, in particular, is a science concerned with consistent, everyday thoughts, feelings, and behavior. The move into the laboratory and away from the field has been driven primarily by practicality: observing behavior as it naturally unfolds is difficult and time-consuming. However, recent innovations in technology and experience sampling have facilitated naturalistic observation. One particularly useful method for personality research has been the Electronically Activated Recorder (EAR; Mehl & Holleran, 2007; Mehl, Pennebaker, Crow, Dabbs, & Price, 2001; Mehl & Robbins, 2012), which systematically samples ambient sounds in a given individual's natural environment. This method has already helped to shed light on several topics, including narcissism (Holtzman, Vazire, & Mehl, 2010), gender (Mehl, Vazire, Ramírez-Esparza, Slatcher, & Pennebaker, 2007) and ethnic