Journal of Research in Personality 52 (2014) 55–67

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

The accuracy of roommate ratings of behaviors versus beliefs

Sampo V. Paunonen [a,*], Chester Kam [b]

[a] Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada
[b] Faculty of Education, University of Macau, Macau

ARTICLE INFO

Article history:
Available online 29 July 2014

Keywords:
Person perception
Peer ratings
Accuracy
Assumed similarity
Observability

ABSTRACT

We evaluated the accuracy of peer ratings of roommates' personality characteristics, against roommate self-ratings, as a function of rating domain observability. Instead of the usual ratings of broad personality traits, however, our domains represented peer ratings of narrow exemplars of personality traits. Specifically, we compared roommate ratings on (a) observable trait-related behaviors with (b) unobservable trait-related attitudes or beliefs. We observed greater self-peer agreement in rating behaviors, in general, than in rating beliefs. We also observed greater tendency of raters to adopt an assumed similarity heuristic when judging their roommates' attitudes and beliefs than their behaviors. We discuss the contribution of these findings to understanding the determinants of accuracy in personality judgments and developing best practices for personality assessment.

1. Introduction

The purpose of this article is to report on a study in which we evaluated some determinants of the accuracy of university students' ratings of each other's personality characteristics. The principal factor of interest here was whether the attribute being considered by a rater referred to (a) an observable behavior of the target, or to (b) an unobservable attitude or belief. We surmised that behaviors would generally be rated more accurately than would attitudes and beliefs (Paunonen & O'Neill, 2010). An attitude, for example, cannot be seen directly, so it must be inferred from certain observable cues emanating from a target's behavior or behavior residue (Funder, 1995; Funder & Colvin, 1988). But the observed cues might be only remotely related to the criterion attitude, leading the observer to judgments of suboptimal validity. In contrast, ratings of a behavior could, in principle, be based on direct observation of concrete instances of that behavior, requiring little or no inference on the part of the rater. As a consequence, such behavior ratings would have less room for error and be more accurate than ratings of inferred attitudes or beliefs.

This area of research is important for at least two reasons. First, finding that certain behavior domains engender more accuracy in judgments of target personality has obvious implications for research in person perception and understanding why some experimental targets are seemingly easier to "know" than are others (Kenny, 1994). Is it possible for differences in rater accuracies to have occurred across studies because judges were asked to rate target attitudes or beliefs in one study but target behaviors in another? Second, recognizing that some types of characteristics are easier to rate than are other characteristics can have an impact on the construction of standardized measures of personality (John & Robins, 1993). If people can report more accurately on other's behaviors than on their beliefs, behavior-referenced personality test items might yield better indices of construct validity. We return to these ideas later on in the General Discussion section.

1.1. Accuracy and observability—personality traits

Studies in the past have looked at the issue of rater accuracy as related to a dimension of observability (e.g., Albright, Kenny, & Malloy, 1988; Beer & Watson, 2008; Borkenau & Liebler, 1992; Cheek, 1982; Vazire, 2010; Watson, Hubbard, & Wiese, 2000). But those studies generally have not evaluated ratings of specific attitudes or beliefs or instances of behavior. The research on accuracy as a function of observability has typically assessed ratings of targets' standing on global traits, such as Big Five personality factors for example, with the more observable traits being posited to engender the greater accuracy (e.g., see Beer & Watson, 2008; Borkenau & Liebler, 1992; Funder & Colvin, 1988; Funder & Colvin, 1997; John & Robins, 1993; Watson, 1989).

Fig. 1 shows one model of how attitudes and behaviors are related to personality traits and personality factors (Big Five Conscientiousness, in this case) in a hierarchical structure, first published by Eysenck in 1947. Because personality traits are hypothetical constructs, they cannot be observed directly and must be

* Corresponding author.
E-mail address: paunonen@uwo.ca (S.V. Paunonen).

http://dx.doi.org/10.1016/j.jrp.2014.07.006
0092-6566/

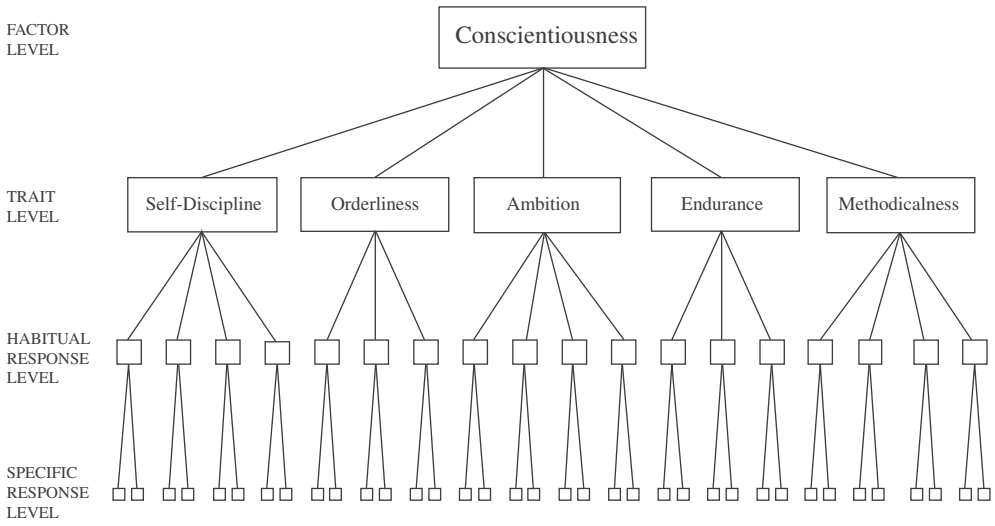

Fig. 1. A hierarchical model of personality organization. Adapted from "Hierarchical Organization of Personality and Prediction of Behavior," by S.V. Paunonen, 1998, *Journal of Personality and Social Psychology*, *74*, p. 539. Copyright 1998 by the American Psychological Association.

inferred from trait-defining specific responses or habitual responses at the lower levels of the hierarchy. Trait orderliness might be inferred from a person's behavior at a specific point in space and time ("He threw his dirty socks on his dresser this morning"), or from the person's characteristic behavior ("He always throws his dirty socks on his dresser"). But those behaviors might not always be observable, representing instead specific or habitual private cognitions ("He enjoyed throwing his dirty socks on his dresser this morning," or "He always enjoys throwing his dirty socks on his dresser," respectively). Note that personality traits, such as orderliness, and factors, such as Conscientiousness, are considered to be bipolar dimensions of personality (see Paunonen & Hong, 2014), a point we return to below.

There are two potential problems with a study that evaluates rater accuracy as a function of the estimated observability of global traits, such as Big Five Conscientiousness or Extraversion (e.g. Borkenau & Liebler, 1995; Cheek, 1982; Vazire, 2010; Watson et al., 2000). First, a particular personality trait might, in general, be high in observability, leading to the expectation that peer ratings on that trait should be relatively accurate. However, a particular person's expression of that trait might not be so observable and not so accurately rated (Cheek, 1982; Kenrick & Stringfield, 1980; Paunonen, 1988). Paunonen and O'Neill (2010) referred to the example of Big Five Conscientiousness, where two targets could be high on that dimension for different reasons: one by virtue of extreme mental self-discipline, and the other because of extreme orderliness of conduct (see Fig. 1). Presumably the level of Conscientiousness would be more observable, and hence more accurately rated, in the orderly target than in the self-disciplined target.[1]

The second issue with rating the observability of personality variables at the level of the global trait is that both poles of a bipolar trait dimension might not be equally observable (Paunonen, 1989). An example might be the dimension introversion–extraversion. Extraverted behaviors are publicly observable by definition. Introverted behaviors, on the other hand, are more difficult to see directly, and they might have to be inferred from the absence of extraverted behaviors. But not seeing something is not the same as seeing its opposite. Thus, inferred introversion would likely engender some error in perceiving such targets, leading to lower accuracy in judging introverts as compared to extraverts. But both introversion and extraversion refer to the same bipolar domain of behavior; in which case one might ask, is that domain highly observable or not?

As described above, there are problems with using measures of a global trait or factor dimensions if the research goal is to evaluate accuracy of personality judgments as a function of the rated domain's observability. (For other problems with using global traits, see Hayes & Dunning, 1997; Paunonen & Hong, 2014; Paunonen & Jackson, 1985; Paunonen & O'Neill, 2010.) A better option, as we describe in the next section, might be to use behavioral, affective, or cognitive exemplars of the trait in question.

1.2. Accuracy and observability—personality behaviors

In our opinion, evaluations of rater accuracy as a function of observability is an important research question (see also Gosling, John, Craik, & Robins, 1998; Human & Biesanz, 2011; Kenrick & Stringfield, 1980; Koestner, Bernieri, & Zuckerman, 1994; Watson et al., 2000; Zuckerman, Bernieri, Koestner, & Rosenthal, 1989; Zuckerman et al., 1988). But observability "should be measured with respect to individual acts of behavior, or even general behavior tendencies, but not with respect to global traits" (Paunonen & O'Neill, 2010, p. 195). Thus, instead of obtaining peer ratings of, say, a target person's level of extraversion or introversion, one might ask the judge about the target's attendance at parties, nervousness while talking to strangers, or liking for solitary board games (cf. Fig. 1). A major advantage here is that reference to such concrete behaviors avoids the problems with measuring trait observability, as alluded to in the previous section. Related to this, behavior exemplars are arguably less subject to interpretive ambiguities (Hayes & Dunning, 1997) than are global trait descriptors, possibly resulting in less response distortion due to desirability biases or semantic misconstruals (Paunonen & Jackson, 1979).

We do not mean to imply in this section that researchers have not looked at the validity of observer ratings of concrete behaviors as a function of observability. For example, a study by Borkenau and Liebler (1992) assessed rater accuracy as a function of the observability of videotaped and audiotaped target behaviors, such as loud voice or fast movements. Those behaviors were then evaluated as cues to ratings of broad Big Five factors as the judgment criteria (see also Borkenau & Liebler, 1995). Gosling et al.

[1] As with many studies in this area of research, we use the term accuracy loosely in referring to the correlation between the target self-ratings of personality and target peer ratings. This usage assumes that the self-reports in question have demonstrable levels of construct validity (Paunonen, 1991). To the extent that this assumption is not viable, a better term than accuracy would be self-peer agreement.