Factors are Still Fictions

WILLIAM REVELLE and LORIEN G. ELLEMAN

Northwestern University
revelle@northwestern.edu

Abstract: Mõttus considers the causal relationship between traits as they relate to outcomes. We applaud his efforts and add that all latent traits identified by factor analysis are convenient mathematical fictions. Traits are the formative results of the (perhaps) non-linear sums of basic biological and social mechanisms. We suggest that personality is fractal and has an equally complex structure (is self-similar) at any level of analysis. Traits are useful fictions for relating the myriad of fundamental causes with the seemingly infinite types of behavioral observations we may make.

At least once a decade, it is time to remind personality researchers that factors are indeed fictions (Loevinger, 1957; Revelle, 1983), and that we should not reify the factors known as 'The Big Five' (Block, 1995, 2010). Mõttus does this, and does it well. He focuses on the supposed causal relationship between traits (as exemplified in the 'Big Five') and outcomes. The argument could equally be applied to causal sources of these traits. Just as some theorists explain individual differences in health or longevity in terms of individual differences in conscientiousness or neuroticism (e.g., Weston & Jackson, 2015), many theorists with biological bent like to 'explain' individual differences in traits such as extraversion with individual differences in strength of biological mechanisms such as the Behavioral Activation System (Corr, 2008; Gray & McNaughton, 2000; Smillie, 2008) or in terms of interactions of the six relatively autonomous systems (sensory, motor, cognitive, affective, value and style) of Royce (1983).

In his use of traits, Mõttus refers to the many who take the perspective termed 'realist' by Borsboom et al. (2003)[2] and claim that traits are real psychological attributes. This is reminiscent of the earlier claim by Cattell (1943, 1945) that factors are source traits that can be used to explain the observed correlations between surface traits of items or of behavior clusters. It is also reminiscent of Royce's use of hierarchical factor analysis to identify 'invariant dimensions of individuality' (Royce, 1983, p. 684). To use the terminology of Bollen and Lennox (1991), implicit in this explanation of outcomes is the reflective latent trait model where traits are common causes of items or behaviors. The items are locally independent; that is, when controlling level of the trait, the items do not correlate. This might well be, but these traits are themselves presumably formative results of the (perhaps non-linear) sums of basic biological and social mechanisms.

That reflective source traits can be used as explanations of covariances of observed items, and behaviors is a convenient mathematical fiction. That five (or three, or six, or ten) factors can be extracted from the matrix of intercorrelations of 67 or 171 paragraph descriptors (Cattell, 1945), 100 (Goldberg, 1992) or 540 adjectives (Hofstee, de Raad, & Goldberg, 1992), or 696 short stemmed items (Condon, 2014), is merely a way of representing (modeling) covariances with factors that can, with a suitable choice of items, lead to near local independence of the items. Unfortunately, any particular exploratory factor solution may be subjected to an infinite number of alternative rotations, all of which are mathematically identical in fitting the covariances. The well-known debate between Eysenck (1967) and Gray (1981) as to whether to rotate towards Extraversion-Neuroticism or Impulsivity-Anxiety cannot be resolved on psychometric grounds.

Furthermore, such factor analytic solutions, although near approximations, are just that: approximations. The optimal number of factors to describe any particular covariance matrix is a tradeoff between parsimony (few factors) and goodness of fit (more factors). But most measures of goodness of fit vary as functions of sample size. As sample sizes increase beyond the 100 of Cattell (1945), or the 800–1000 participants of the Eugene-Springfield data set of Goldberg and Saucier (2016), to the sample sizes available through web-based data collection, e.g. $>$ 24 000 in Condon and Revelle (2015),$>$ 65 000 in Revelle, Wilt, and Rosenthal (2010) or $>$ 200 000 in Revelle et al. (2016), the number of interpretable factors increases. Indeed, it appears that the factorial structure of personality is fractal, that is, it is equally complex (self-similar) at all levels of analysis.

Each of three to five high-level factors shows horizontal as well as vertical structure (Goldberg, 1993) and can be subdivided into three to five lower-level factors which in turn yield three to five homogeneous item composites. As is usually the case, Lew Goldberg has made this point before:

'Because one always loses specific variance as one amalgamates measures, the optimal level of prediction is completely a function of statistical power and thus of sample size. In the population (i.e. samples of unlimited size) optimal prediction by regression analysis will always be at the level of individual items; that is, for huge samples it would be silly even to amalgamate the items into scales because one would inevitably lose some specific variance at the item

[2]Although Borsboom et al. (2003) credit Spearman (1904) as originator of the concept of latent variables as theoretical constructs, the concept of unobserved (latent) cause of observations goes back at least 2400 years to the Allegory of the Cave in Plato's *The Republic*.

DOI: 10.1002/per

level that could serve to increase predictive accuracy.' (Goldberg, 1993, pp. 181–182).

Mõttus gave one example of the need to consider item-level data when examining trait-behavior correlations. When predicting variation in Body Mass Index (BMI), rather than the very broad trait of Neuroticism, or the facet of Impulsivity, it is at the item level, e.g., 'When I am having my favorite food, I tend to eat too much' that the best predictors of BMI are found (Terracciano et al., 2009). Support for this also comes from our finding using the SAPA methodology (Revelle et al., 2016) with N > 50 000 where the best measures predicting BMI are 'I ate too much' (r = 0.25) or 'used public transportation' (r = 0.20).

Traits, facets, or items are useful links between fundamental causes (genes, enzymes, transmitters, brain structures, environmental experiences) and behavior. They are useful because they summarize broad patterns of relationships with observed regularities of behavior. They lead to appropriate levels of specificity within the broad framework of symmetry of predictor and criterion (Wittmann, 1988). Although useful, they are fictions created for the purpose of telling a coherent story relating the myriad of fundamental causes with the seemingly infinite types of behavioral observations we may make.

Partially supported by National Science Foundation grant: SMA-1419324 to William Revelle.

Establishing that Estimated Trait-Outcome Associations Aren't Artefactual, Inflated, or Attenuated by Specific Indicators

KENNETH J. SHER

University of Missouri
sherk@missouri.edu

Abstract: Mõttus' suggestions of ways to establish the validity of observed trait-outcome associations are well reasoned. I consider one general issue, the extent to which some observed associations could be artefactual due to criterion contamination of specific trait indicators, and briefly review alternative ways of assessing this problem. Although optimal ways are not clear-cut, approaches Mõttus and I suggest could be viable candidates. Moreover, these same approaches could be used to identify trait indicators that attenuate observed trait-outcome associations artefactually. Copyright © 2016 European Association of Personality Psychology

Mõttus highlights several important issues often overlooked in everyday practice, focusing on importance of insuring validity of estimated trait-outcome associations. Among other points, he highlights renewed recognition of the importance of treating experimental conditions as random rather than fixed effects, a position long-espoused in experimental and clinical research (e.g., Brunswik, 1955; Hammond, 1954), and that the same logic could be extended to conceptualizing personality tests items.

The question of 'whether the associations of traits with outcomes are independent of which indicators are employed rather than being specific to particular indicators' (p. 19) should logically be related to the generalizability coefficient of the scale measuring the trait. In principle, this coefficient represents how well the employed set of indicators generalizes to the universe of suitable indicators. If generalizability is high, the average inter-item correlation is of moderate magnitude, and the number of indicators is large, the extent to which a given indicator is artefactually inducing a trait-outcome association should be, *a priori*, very low. However, in common practice, there is often opportunity for indicator-specific contamination to create appearance of association at the trait level.

The issue of predictor-criterion overlap or 'criterion contamination' has long been recognized in research in personality and substance use (e.g., Darkes, Greenbaum, & Goldman, 1998) and failure to recognize this potential problem can probably be blamed on complacency or lack of due diligence in knowing the item content of personality scales employed in one's research. In my work as a reviewer and editor, I have been surprised at how often this issue has figured prominently in an editorial decision, suggesting its importance has not been sufficientlyly recognized. Moreover, the prototypic exemplars noted by Mõttus and Darkes et al. (1998) and others may represent only the tip of an iceberg in that it seems likely there can be considerable predictor-criterion overlap that is much subtler and more implicit than explicit. For example, if liking 'wild parties' is an indicator of extraversion, its association with drinking alcohol may not be due to blatant criterion contamination but rather implied contamination (i.e., 'wild parties' are often accompanied by alcohol excess). Scrutiny of personality trait indicators and outcome measurements can reveal clear instances of likely contamination, instances of unlikely contamination, but also instances in a 'gray zone' and subject to interpretation and judgment. Due diligence in considering this issue in all work and attempting to address it empirically would appear to be foundational to good research practice.

Although logical and semantic analysis of indicators and outcomes is a reasonable place to start, statistical approaches such as examining the residual correlation between an

DOI: 10.1002/per