Common Factors and Causal Networks

James J. Lee

Vision Lab, Department of Psychology, Harvard University

Laboratory of Biological Modeling, NIDDK, NIH

Cognitive Genomics Lab, BGI-Shenzhen

Abstract

The target articles touches upon some of the most difficult and essential questions in personality psychology. Questioning the notion of a common factor as an as-yet unobserved common cause of a behavior domain's exemplars, the authors propose using graphical representations as sources of hypotheses positing more complex causal structures. I do not find the case for the deemphasis of the common factor model to be compelling for those behavior domains (cognitive abilities) with which I am most familiar. It behoves all personality psychologists, however, to question the foundations of their favored tools.

Common Factors and Causal Networks

The target article brings before us, once again, the question of whether psychometric common factors should play any role in the science of individual differences. Notwithstanding the qualifiers in their conclusion, the authors clearly lean toward *No.*

The alternative advocated by the authors encompasses a weaker and stronger formulation. The weaker formulation is that we may represent the correlations among exemplars of a behavior domain as an undirected, weighted graph. The description of the algorithm for node placement suggests that (the picture of) the resulting graph is similar to a multidimensional scaling plot (Marshalek, Lohman, & Snow, 1983; Guttman & Levy, 1991). As a stimulus toward further investigation, a less "lossy" visualization technique than multidimensional scaling can of course raise no objections.

A stronger formulation, although never succinctly stated, is evident in the vivid informal accounts of how different behavioral variables might be causally related. It is here that the authors wade into murkier waters. They first claim that the common factor model—the foundation of mainstream psychometric methodology (McDonald, 1999)—is a *causal* model. That is, the common factor of a behavior domain is supposed to be a hypothetical, as-yet unobserved quantity; if an individual's amount of this quantity could be experimentally manipulated, then as a consequence the individual's scores on the indicators should increase by amounts proportional to their factor loadings. Rightfully questioning the plausibility and logical soundness of this conception, the authors go on to propose replacing the common factor model with systems of structural equations (generally nonrecursive, generally nonlinear) whose graphical representations allow any two nodes to be adjacent. The authors may object to this characterization, but I do not see how their informal examples ("changes in terms of state" leading adjacent nodes to "become active") permit a weaker position.

An attempt to elucidate the causal graph containing a set of variables can also raise no objections. But we might question whether this ambitious enterprise is truly in competition with the common factor model. Many writers have found factor analysis to be a powerful tool even while disavowing the causal interpretation of the common factor model that the authors set up as their foil (Lord & Novick, 1968; Messick, 1989; McDonald, 2003; Bartholomew, 2004). In essence, if we wish to measure a psychological attribute such as "mathematical ability," we can use the common factor model toward this end without supposing that mathematical ability is a hidden "lever in the brain" (Ryle, 1949). I suspect that the authors would consider attempts to measure a folk-psychological trait such as mathematical ability to be bad philosophy and worse science. Nevertheless such attempts make up the daily business of the psychometricians at ETS, ACT, and elsewhere, and I doubt that the target article would dissuade them from engaging in the myriad applications enabled by factor-analytic theory: computing errors of measurement, altering the length of a test to meet a desired reliability, constructing alternate forms, computer adaptive testing, detecting biased items, and so forth.

In fact, longstanding operational tests point to certain domains of personality psychology where factor analysis may be more than merely complementary to the authors' preferred approach. Consider the fact that ETS has written thousands of items for its various tests of aptitude and achievement; two successive administrations of a given test must have no (scored) items in common. Is it sensible to search for the causal graph connecting *all* of the items that have been written for a particular test over the years? This is a deeply puzzling question. It suggests that an abstraction of a potentially infinite set—a common factor—may sometimes be more fundamental than a micro-casual account relating the set members.

It is noteworthy that my commentary appeals to abilities rather than non-cognitive personality traits. The proffered alternative to the common factor model may seem less attractive in the domain of abilities for at least two reasons. First, ability tests yield data that approach Stout's (1990) ideal of "essential low-dimensionality" more closely than do personality questionnaires. To the extent that the poor fit of personality questionnaires to low-dimensional factor models provides motivation for a different approach (Vassend & Skrondal, 1997), this motivation is correspondingly reduced in the domain of abilities. Second, a single item in an ability test often seems less psychologically interesting ("more exchangeable") than a single item in a scale such as the NEO or HEXACO. A political scientist might object to treating the item stem *I get chores done right away* as congeneric with *I take voting and other duties as a citizen very seriously*, since the behavior probed by the latter may be judged to be important in its own right. Contrast this with the following pair of items:

- If $4x - 5 = 11$, then $3x = ?$
- If the average of 4 consecutive integers is 18, what is the sum of the least and greatest of the integers?

The necessary judgments here concern what Meehl (1978) called the *situation-taxonomy* and *response class problems*: parsing the raw streams of stimuli and responses into meaningful units. A possible resolution in the domain of non-cognitive personality traits may be to improve the fit of questionnaire data to common factor models through more careful item writing and leaving substantively important individual differences that are either peripheral or "factorially complex" as nodes to be connected to the personality traits by further investigation.

What exactly is a trait? This is a question that has challenged scientists in all fields where it has arisen (Wagner, 2001). The target article's orientation toward single indicators does not dispose of the question—as the authors themselves recognize in their acknowledgment of "semantically logical dependencies." If we slightly reword an item stem such as *we can never do too much for the poor and elderly*, it will order the respondents somewhat differently. How should we conceive of the “error-free” attribute that the stem is intended to measure? It behoves personality psychologists to consider the difficult and essential questions such as this that lie at the foundation of our field.

References

Bartholomew, D. J. (2004). *Measuring intelligence: Facts and fallacies.* Cambridge, UK: Cambridge University Press.

Guttman, L., & Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence*, *15*, 79–103.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*, 107–127.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Erlbaum.

McDonald, R. P. (2003). Behavior domains in theory and in practice. *Alberta Journal of Educational Research*, *49*, 212–230.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806–834.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). New York, NY: Macmillan.

Ryle, G. (1949). *The concept of mind.* London, UK: Hutchinson.

Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, *55*, 293–325.

Vassend, O., & Skrondal, A. (1997). Validation of the NEO Personality Inventory and the five-factor model: Can findings from exploratory and confirmatory factor analysis be reconciled? *European Journal of Personality*, *11*, 147–166.

Wagner, G. P. (2001). *The character concept in evolutionary biology.* San Diego, CA: Academic Press.