So What Do You Propose We Use Instead? A Reply to Block

Lewis R. Goldberg  
University of Oregon and Oregon Research Institute

Gerard Saucier  
California State University, San Bernardino

Unfortunately, Block's brilliant critique is terribly biased, much like a legal brief that presents only one side of the issues at suit. It does not distinguish between the Big Five model of phenotypic personality attributes from alternative models of the causal underpinnings of personality differences. Ironically, it attempts to explain away the extensive evidence for the Big Five model as largely the result of data prestructuring, with no acknowledgement of the unique contribution of the lexical approach to minimizing such problems. Even more seriously, it omits a good deal of crucial evidence favorable to the Big Five model, including studies of Block's own Q-set and independent investigations of personality-related terms in other languages. Sadly, Block's closing suggestions provide little in the way of specific proposals for alternatives that he would have us use instead.

We have distinguished between the intuitive scientist, who seeks to make the most accurate or otherwise optimal decision, and the intuitive lawyer, who desires to justify a preselected conclusion... The intuitive lawyer builds the strongest possible case for the preferred conclusion by selectively seeking evidence that will be favorable, and by assessing implications in a way that is carefully critical of disagreeable implications while failing to subject supporting evidence to equally critical scrutiny. (Baumeister & Newman, 1994, p. 16, italics added)

Block can be a brilliant barrister, and his article is a remarkable legal brief. As such, it provides a penetrating, but highly jaundiced, analysis of the evidence behind an important rapprochement that is beginning to be achieved among those involved in the scientific study of personality. Like any legal brief, its intent is to convince and persuade—in this case to persuade us to ignore the large body of scientific evidence that has accrued over the years about the structure of phenotypic personality attributes.

Block argues that the five-factor “bandwagon” must “simply be halted” (1995, p. 209). But, science proceeds by using the best tools and models that are currently available, not by halting until ideal ones get developed. The Big Five model is not yet fully developed nor is the evidence for it perfect. It is, however, the most comprehensive and parsimonious model of phenotypic personality descriptors that is now available, and the evidence for its utility is more voluminous and more compelling than for any other such model. As a consequence, to halt its use would only serve to halt the development of its successors.

Indeed, its successors will emerge, as Block acknowledges, from the empirical comparison of alternative scientific models, based on “scientific contests, fairly waged and evaluated” (1995, p. 207). We agree overwhelmingly with Block that “there must be empirical and conceptual competition between alternate dimensional offerings to see which, predictively, best carves nature at its joints” (p. 207). But, scientific contests depend on the availability of viable alternative models.

Just as a good criminal defense attorney will invoke the possibility that a tall bushy-haired stranger was the actual perpetrator of the crime, so Block raises the possibility of an alternative taxonomic framework that might be a viable contender to the five-factor model:

Thus, let us suppose that a quite different set of reliable variables—an alternative personality framework—had been assessed by McCrae and Costa within their ALBSA sample—a taxonomy consisting, perhaps, of measures of ego control (overcontrol versus undercontrol), ego resiliency, agency-communion, introspective ness, energy level, and liberalism-conservatism. These particular variables are highly reliable, are of conceptual interest, and are empirically relatively independent. (p. 204)

But, then, having whetted our appetite for a true scientific contest, Block cancels the match: “I do not offer this set of variables as definitive or as without problems similar to those that beset the FFA” (Footnote 7, p. 204, italics added). So much for the credibility of the argument for the tall bushy-haired stranger!

To his credit, Block relishes difficult cases. What makes the present case more challenging than many that Block has taken on over the years (e.g., Block, 1967, 1971, 1977a, 1977b, 1977c, 1989) is the sheer mass of evidence for the Big Five representation that has now accrued, only a selective portion of which is reviewed in his critique. What makes his case appear more plausible, however, is the fact that the scientific literature has conflated two alternative five-factor models, which, though superficially similar, were developed for different purposes (John & Robins, 1993). The Big Five factor structure has been proposed as a taxonomic framework for phenotypic personality attributes as these have become encoded in human languages (Saucier & Goldberg, in press); the Five-Factor Model of McCrae and Costa (in press), on the other hand, has more ambitious goals, having been developed expressly as a model of genotypic personality traits. Most of Block's critique is focused on the evidence relating to the five-factor model as a representation of the causal underpinnings of individual differences, the quest for...
which unites Cattell, Eysenck, Costa and McCrae, and, of course, Block himself.

In contrast, the formulation of the Big Five model of phenotypic personality attributes has proceeded from the more modest goal of discovering a framework for organizing those thousands of terms in the lexicons of the world’s languages that refer to individual differences in personality attributes. Given this goal, it makes sense to study more languages than English alone in a search for cross-language regularities analogous to those found in descriptions of colors (e.g., Witkowski & Brown, 1977) and facial expressions of emotion (e.g., Ekman & Friesen, 1971). Moreover, it also makes sense to focus primarily on single terms and secondarily on questionnaire statements, and to eschew the investigation of “paragraphs, pages, chapters, and books” (Block, 1995, p. 196) at least until it has been shown that such complex stimuli are worth pursuing.1

The Scientist as Solicitor

The differing methods of lawyer and scientist can be no better illustrated than in the ways that they react to perceived difficulties with a scientific hypothesis: The lawyer prepares a case against the hypothesis, whereas the scientist designs a study to test it. In the case of the lexical hypothesis, it is instructive to compare the reactions of Block with those of Tellegen, both of whom are skeptical about the Big Five factor structure. Indeed, Tellegen has had many of the same concerns as Block: He was critical of the decisions made by Norman, Goldberg, and others to select the sets of terms they studied, and like Block, he worried about the lack of knowledge of the meanings of such terms, even among college students. To solve both problems, he used a dictionary to sample systematically from the total set of adjectives referring to psychological differences (not restricted to common trait descriptors), and then he included with each of the selected terms its dictionary definition.

Tellegen’s findings undermine Block’s conjectures that “the sequence of empirical procedures that repeatedly issued similar five-factor structures may have been constrained to produce the results obtained” (1995, p. 200). For, in a sample of terms that could have no possible subjective constraints, with a methodological procedure that eliminated individual differences in subjects’ understandings of the meanings of the terms, and in both five-factor and seven-factor solutions, five of the factors were identified by Tellegen and Waller (1987) as variants of the Big Five.2

Moreover, the findings of Tellegen and Waller (1987) are only a part of a pattern of contradictory evidence that Block does not even mention, much less discuss. Part of Block’s problem in discerning this pattern may lie in the common human difficulty of separating background from figure or of observing that on a particular night the usually noisy dog has failed to bark. For it is easy to detect differences in the sheer number of factors reported in an investigation, whereas it is far more difficult to track those factors that have proven to be replicable across diverse studies. Many investigators have reported more than five factors; to our knowledge, none of these other factors have replicated across other sets of personality attributes.

For example, Lanning (1994) reported eight factors in his extensive analyses of Block’s California Adult Q-Set, whereas John, Caspi, Robins, Moffitt, and Stouthamer-Loeber (1994) reported seven factors in their analyses of the child version of that instrument. In each of these investigations, five of the factors corresponded to the Big Five. In contrast, none of the additional factors corresponded to each other, and none of them corresponded to either of the two Evaluative factors reported by Tellegen and Waller (1987). Thus, broad factors beyond these five have not (yet) proven to be generalizable.

An old axiom of trial law asserts, “When the case is weak, pound the table!” And, nowhere is Block’s case weaker than in his review of the findings from the German lexical-taxonomy project. The crux of Block’s argument against the Big Five factor structure as a representation of English trait-descriptive adjectives lies in the following suspicion:

It may be that in the sequence of adjective culling, and cluster culling, and cluster forming procedures by Norman and by Goldberg, which eventuated in very much the same five factors, adjectives and cluster-based variables interstitial to these factors were unknowingly de-emphasized. . . . It, therefore, would be helpful if the immense task of adjective and cluster culling and cluster forming were to be replicated by disinterested investigators. Until an independent verification has been established, the proposal that the lexically based five factors be accepted as the conceptual framework for the scientific study of personality would seem premature. (p. 199)

That is fair enough: Independent verification is the hallmark of scientific progress. And it would be hard to think of a more independent verification than one conducted by a completely independent team of scientists, a team that strove to use procedures that were as objective and precisely formulated as possible (Angleitner, Ostendorf, & John, 1990; Ostendorf & Angleitner, 1992), especially an empirical analysis of the lexicon from a

---

1 Block invokes as evidence that single terms provide insufficient descriptive potential a number of examples of expert-defined personality syndromes that supposedly cannot be encoded as single adjectives. Such examples are misleading, because even experts’ concepts tend to fall into single-word summary labels once they become familiar and important enough (e.g., masochistic, sadistic, passive-aggressive, histrionic, authoritarian, Type-A, schizoid, self-actualizing). As yet, however, there is no clear evidence that the structure of expert personality concepts differs markedly from the structure of lay personality concepts, at the broadest level. Indeed, evidence summarized in a recent volume edited by Costa and Widiger (1994) suggests substantial congruence between the Big Five and the structure of expert-defined conceptions of personality disorders. Obviously, there is a virtually unlimited number of personality concepts definable in sentences or paragraphs, but most of these will not be used with any frequency in the discourse of either experts or laypersons. If Block can produce a personality concept that has not yet been summarized by a single-work label, we expect that his concept is not yet well accepted in either the science or the common language of personality. The lexical hypothesis makes no claim to cover all possible personality concepts, only those that have proven useful enough to need summarizing in a single-word label. In general, then, single personality-descriptive terms function as summary concepts, and like all such concept labels, they make use of a large body of background information to unpack their total meanings. That is, trait-descriptive labels function in exactly the same way as does the label “computer,” which does not specify all “the complexities and the complications” (p. 196) of those helpful machines.

2 Tellegen and Waller’s finding of two additional Evaluation factors using a single subset of personality adjectives and a single sample of subjects now needs careful replication with diverse subsets of terms in other subject samples.
WHAT SHOULD WE USE INSTEAD?

completely different language! Indeed, it could easily be argued that the strongest evidence in the case for the Big Five factor structure can be found in the report of Ostendorf (1990) on the factors obtained in analyses of German trait-descriptive terms. Block relegates this extraordinarily important independent confirmation to a footnote:

A recent monograph by Ostendorf (1990) warrants mention. He reports, based on a complicated, arduous, and methodologically sophisticated study of lay self- and other-ratings of 430 German single-word descriptors, that five factors similar to the English lexical Big Five emerge. I agree but note that he obtained at least eight highly replicable factors, according to the well-regarded criterion of Everett (1983) (see Ostendorf’s Tables 46 and 47). (Footnote 4, p. 199)

This Table-banging serves to obscure a crucial point: Everett’s (1983, p. 200) criterion is based on split-half factor comparability coefficients of .90 or higher, corresponding to angles between the corresponding factors that are less than 26°. In Table 46 (peer ratings) of Ostendorf (1990), the actual values of the replicability coefficients in the five-factor solution are .99, .99, .99, .98, and .93, as compared with a value of only .76 for the sixth factor in the six-factor solution. The comparable values in Ostendorf’s Table 47 (self ratings) are .99, .99, .98, .97, and .93 for the five-factor solution, as compared with a value of .14 for the sixth factor in the six-factor solution. Of far greater significance, however, Ostendorf (1990) provides clear evidence that the first five factors in the German analyses are nearly identical to those reported in previous analyses of English trait-adjecitives (e.g., Goldberg, 1990). That is, even if Block had been correct in his interpretation of Ostendorf’s tables and it had turned out that there were eight large orthogonal factors in the German personality lexicon, it would still ruin Block’s constraint hypothesis if five of these German factors corresponded to the English Big Five! Here again, Block argues his case against the lexical hypothesis in complete disregard of critical data that support it.

Straw Figures and Imaginary Lions

U.S. Supreme Court Justice David Souter recently wrote of a dissenting opinion by fellow Justice Antonin Scalia that it was “certainly the work of a gladiator, but he thrusts at lions of his own imagining” (Savage, 1994). Certainly one way to build one’s case is to attribute an extreme position to one’s adversary and then to denigrate that imaginary position (e.g., Dawes, Singer, & Leemons, 1972). Block erects and whips several such straw figures.

For example, Block argues that the recurring finding of five factors stems from prestructuring lexical data “from the top down” in terms of the Big Five, by way of selective cluster-sampling and “iterative judgmental procedures it is understandingly impossible to truly specify” (1995, p. 199). Thus, presumably to find a lexical Big Five, one must use subjective cluster-sampling. Reality is dangerous to this fiction: The Big Five have been isolated without any subjective clustering and with clearly specified and replicable procedures (e.g., Digman & Inouye, 1986; Ostendorf, 1990; Peabody, 1987). Moreover, as we have already noted, the Big Five have been clearly identified in analyses of Block’s own Q-sets (e.g., John et al., 1994; Lanning, 1994; McCrae, Costa, & Busch, 1986; Robins, John, & Caspi, 1994).

Of even greater import, the lexical approach is surely the single approach in personality psychology that best minimizes prestructuring, because the all-important step of variable selection can be taken out of the researcher’s hands and delegated instead to dictionaries and aggregated raters (e.g., Angleitner, Ostendorf, & John, 1990).

In this light, Block’s contrived example of a factor analysis based on 99 measures of shoe-tying competence and one measure of intelligence turns out to be another straw figure. Not only does Block cite no actual personality researcher so eccentric as to select 99 measures of the same trivial variable, but, as we already have argued, the lexical approach provides some assurance that no trivial variables will be over-sampled. For example, there are no terms in English that refer to competence in tying shoes, but there are hundreds that refer to aspects of intellect. As a result, trivial lexical factors will generally not appear, whereas any factor latent in a large number of “redundant” lexical variables (e.g., intellect) is bound to be of some psychological importance.

Although it is reasonable to oppose the uncritical use of factor analysis as a mechanical truth generator, Block’s critique provides only a gross caricature of contemporary factor recommendations (e.g., Goldberg & Digman, 1994). In practice, “upstream influences” (Block, 1995, p. 190) and bloated specifics can be averted by a well-informed and well-rationalized selection of variables (e.g., Peabody & Goldberg, 1989). The problem of finding factorial invariance across differing subject samples and different methods of factor extraction and rotation can be obviated by the time-honored scientific principle of replication. The most consequential factor solutions are those that prove to be replicable across diverse samples of subjects, methods, and variables. To date, the Big Five factor structure is the only taxonomic model to have survived the rigors of this endless process.

Finally, Block’s warnings about the imminent takeover of personality psychology by Big Five zealots represent further thrusts at imaginary lions. The Big Five model is a useful framework for organizing personality attributes, but neither the lexical perspective nor the Big Five model is intended to provide a sufficient and exhaustive model of personality. That is, the lexical approach can identify variables too important to ignore, but it will not necessarily identify all variables of importance (Saucier & Goldberg, in press). The Big Five factor structure is less a threat to other researchers than it is important information for them to utilize, and it is complementary to models of causes, motivations, and intraindividual dynamics, including those developed by experts (John & Robins, 1994). Block’s review fails to take into account the limited purpose and purview of the lexical approach and the Big Five model.

So What’s the Bottom Line?

Just as attorneys can win a seemingly lost case by the cogency of their closing arguments, so Block could have overcome his unbalanced account of the evidence for the Big Five model with a scientific breakthrough made in his closing suggestions. As a consequence, we should examine each of these conclusions as carefully as our allotted space permits. We have already commented on his first suggestion, namely that work replicating and extending the Big Five model be “halted” while the field pauses
"for wider reflection on its conceptual and empirical requirements and on its past attainments and deficiencies" (1995, p. 209). This nihilistic suggestion gets us nowhere, except perhaps back into the 1970s when the field was virtually paralyzed in agonizing reflection on its conceptual and empirical requirements and its past attainments and deficiencies—particularly the latter (Goldberg, in press). Instead, we would argue for the continued evaluation and improvement of the model, including the exploration of its generality in diverse languages (e.g., de Raad & Szirnák, 1994) and the refinement of its facets (e.g., Hofstee, de Raad, & Goldberg, 1992). In addition, we applaud all attempts to develop a viable alternative to this still provisional taxonomic structure.

Block’s second suggestion is that it be more generally recognized that the extraordinarily useful method of factor analysis by itself cannot be empowered to make paramount and controlling decisions regarding the concepts to be used in the field of personality assessment. . . We shall also have to use other and diverse psychological resources involving close conceptualizing, perceptive observation, and unconfounding empiricism. (p. 209)

But, which “resources”? Shouldn’t we expect that after such an extensive discussion of the purported limitations of factor analysis Block would specify some better procedures to be used as alternatives or supplements?

Block’s third suggestion is that “personality psychology use a conceptual language suitable for experts rather than one provided by and comfortable for novices” (1995, p. 209). Here Block is confusing scientific concepts and their labels with folk concepts and their labels (Tellegen, 1993). The lexical hypothesis deals only with the latter, and in analyses of the natural language of personality it is appropriate to use lay speakers of the language. Most research in the lexical tradition has used college students as subjects—like most research on virtually any topic in psychology; for lexical analyses, however, college students may be more suitable than random samples of the adult population because of their higher verbal facility. Moreover, studies using even more expert speakers of the language have been conducted, and their findings are quite concordant with those based on college students (e.g., John, 1990; Lanning, 1994). If Block truly believes that a replication of any of the classic studies of the Big Five, using a large sample of clinical experts as raters, would lead to different findings, we challenge him to carry out such an investigation and test this hypothesis. Our bet is that it will be shown to be incorrect.

Block’s fourth suggestion is that “personality psychologists not limit their thinking and research by considering only what is or will be convenient to index via simple self-report or lay-person-report measures” and when necessary use “more complicated and complex ways of studying persons” such as behavioral observations and psychophysiological measures (1995, p. 209). This straw figure argument implies that there are hordes of investigators who advocate limiting our thinking and research to simple self-report or lay-person-report measures. We know of no one. Different methods are more suitable for different scientific problems; for research in the lexical tradition, users of the language will normally provide the most appropriate type of scientific data.

Block’s fifth suggestion is for the field “to resolutely confront its severe, even crippling, terminological problems” (1995, p. 209). “We might fruitfully recommit by deconstructing the myriad, often disparate, meanings that have been ascribed to the Big Five” (p. 210). Block is probably lamenting the fact that over the years different investigators have used different labels for the Big Five domains (John, 1990). Many of these historical and/or stylistic differences are disappearing, as most investigators have come to adopt either the Roman-numeral designations proposed by Norman (1963) or the initial-letter labels proposed by McCrae and Costa (1985). Moreover, as investigators begin to differentiate more clearly between the phenotypic and the genotypic models, the use of one of the two sets of labels can be used to signal whether the investigator is referring to the Big Five structure of phenotypic personality attributes (Factors I to V) or the five-factor model of genotypic personality traits (N, E, O, A, and C).

Block’s “final, most ambitious suggestion is that, in the conceptual/empirical arguments to be made for specific dimensions, of whatever number, these constructs be situated within a coherent, intraindividual theoretical framework” (1995, p. 210). Block laments the fact that differences between individuals, there is no end. An infinite number of sets of descriptive variables can be formulated, each being preferred by its progenitor and contestable otherwise. What is needed is a basis for choosing among these alternative sets. Efforts to study or conceptualize the dynamics underlying intraindividual functioning might well move the study of personality toward such a basis. (p. 210)

Maybe. Although it has been over 25 years since Norman (1967) provided the methodology for analyzing relations among variables at the intraindividual level, few investigators have used these procedures, and Norman’s classic article is not even referenced in Block’s critique. More importantly, Block seemingly fails to grasp the significance of a taxonomic model as an overarching framework in which the “infinite number of sets of descriptive variables” can be embedded and thus understood. Nor does he seem to appreciate the power of such a framework as an aid for conceptualizing intraindividual dynamic processes (e.g., John & Robins, 1994; Zevon & Tellegen, 1982).

In closing, we heartily agree with Block that “the scientific understanding of personality functioning and personality development is intellectually fascinating and of fundamental importance for general psychology” (1995, p. 210). But, how can this understanding best be gained? Surely not by reviews that select only one side of a large body of relevant evidence while omitting whatever is unfavorable to the reviewer’s point of view. This litigious aspect of Block’s review conflicts with the most crucial features of scientific evaluation: “... the intuitive scientist further the goal of reaching the best decision by seeking thoroughness in the collection of evidence, by emphasizing the detection of bias during the assessment of evidence, and by adjusting or recomputing decisions so as to correct for the effects of any biasing or distorting factors” (Baumeister & Newman, 1994, p. 16, italics added). How we wish that, rather than merely playing a contrarian role, Block had provided a more thorough, balanced, and unbiased appraisal of the evidence.

References
WHAT SHOULD WE USE INSTEAD?


