
Norman G. Poythress, Jennifer L. Skeem, Scott O. Lilienfeld, Kevin S. Douglas, John F. Edens

Abstract

Newman and Malterer (2009) recently challenged the conclusions reached in our critique of the construct validity of Carver and White’s (1994) Behavioral Inhibition System (BIS) scale for testing Lykken’s (1995) “weak BIS” theory of primary psychopathy, and they argued instead that the problems may lie with Lykken’s model itself. In this rejoinder we respond to specific issues raised by Newman and Malterer. Although we agree with Newman and Malterer that there may well be problems with Lykken’s model, the additional data and rebuttal arguments presented by Newman and Malterer are insufficient to rehabilitate the BIS scale which, because it fails to assess sensitivity to conditioned fear stimuli, is not a valid measure for testing Lykken’s low fear (or “weak BIS”) hypothesis.

1. Introduction

In a recent review (Poythress, Edens, et al., 2008) we questioned the construct validity of Carver and White’s (1994) Behavioral Inhibition System (BIS) scale for testing Lykken’s (1995) theory of primary psychopathy. Lykken hypothesized that primary psychopathy results from an innately fearless temperament. Citing evidence regarding its psychometric properties, content validity, and criterion-related validity, we concluded that the BIS scale inadequately captures the fear sensitivity construct upon which Lykken-based his “weak BIS” hypothesis of primary psychopathy. In a critique of our review, Newman and Malterer (2009) challenged our findings and conclusions. In this rejoinder we explain why Newman and Malterer’s criticisms are less than compelling.

2. Multidimensional structure of the BIS scale

In our review, we (Poythress, Edens, et al., 2008, p. 271) noted that the internal consistency of the BIS scale has generally been satisfactory in non-offender samples and was reported as α = .71 in a female offender sample but only α = .58 in a male offender sample. In their critique, Newman and Malterer reported that the internal consistency for the BIS scale was satisfactory (α = .75) in one of their large offender samples.

This is a helpful addition to the limited literature on the psychometric properties of the BIS scale in offender samples. Still, high internal consistency (which references the level of covariance among items) does not convey unidimensionality (which references the pattern of covariance among items) (see Hattie, 1985); a scale may evidence high internal consistency (α) in the presence of multidimensionality. Factor analysis is better suited for determining whether the BIS scale measures a unidimensional construct. We cited numerous problems from structural analyses of the BIS and Behavioral Activation System (BAS) scales. In particular, the two BIS items that have apparent face validity for indexing fear (BIS-F) sensitivity (“I have very few fears compared to my friends;” “Even if something bad is about to happen to me, I rarely experience fear or nervousness”) have frequently performed poorly (e.g., low factor loadings on the BIS factor; marked secondary loadings on BAS factors). Newman and Malterer ignored our warnings about this problem.

Table 1 summarizes the results of nine studies that used item level CFA to evaluate the BIS/BAS scales. None yielded fit indices that uniformly support the original structure. This is despite the fact that in several studies CFA was applied to data from the same sample used to derive the structure being tested, thereby potentially capitalizing on chance variations in the original dataset. In the only study to use an offender sample, Poythress, Skeem, et al. (2008) found that the two BIS-F items formed a separate factor. Using a college sample, Heym, Ferguson, and Lawrence (2008) similarly found that the BIS scale was best modeled as two separate factors, with the BIS-F items (along with a third item) forming a...
merely an index of NE orioral inhibition system functioning. In response, Newman and with measures of NE than with other putative measures of behav-

studies, the BIS scale is as strongly, or more strongly, correlated trum (Poythress, Edens, et al., 2008, Table 1). In numerous ciations with constructs in the negative emotionality (NE) spec-

3. Association of BIS scale with measures of negative emotionality

Part of our critique of the BIS scale focused on its positive asso-
ciations with constructs in the negative emotionality (NE) spectr um (Poythress, Edens, et al., 2008, Table 1). In numerous studies, the BIS scale is as strongly, or more strongly, correlated with measures of NE than with other putative measures of behavo rial inhibition system functioning. In response, Newman and Malterer (2009) asserted that our “…criticism of the BIS scale as merely an index of NE … appear[s] to reflect Lykken’s peculiar views of anxiety” (p. 674), and that “… the claim that the WAS [Welsh Anxiety Scale] and BIS scales are measuring the same con struct seems overstated” (Newman & Malterer, 2009, p. 675).

In concluding that the BIS scale is primarily an index of NE, we did not reference Lykken’s views – peculiar or otherwise – about anxiety. Instead, our conclusions were based on the empirical literature. Rather than being “overstated,” they echo conclusions reached in other comprehensive reviews of the literature: “… the strength of correlations between [various anxiety measures] and purpose-built BIS scales was no weaker than that observed among the latter … it may be affirmed that anxiety and BIS scales assess roughly the same construct” (Torrubia, Ávila, & Caseras, 2008, p. 218, emphasis added).

Newman and Malterer (2009) reported new correlations from two studies with male offenders in which BIS associations with a measure of NE were relatively weak (r = .217 and r = .159). Although these data seem to be relevant counter-points to our general findings (Poythress, Edens, et al., 2008, Table 1), they are far less compelling than Newman and Malterer imply, at least insofar as they may underestimate the associations of the BIS scale with anxiety. The measure of NE Newman used is a higher-order scale of the Multidimensional Personality Questionnaire (MPQ; Tellegen & Waller, in press). This higher-order scale subsumes not only a trait anxiety subscale (Stress Reaction) but also two additional subscales (Aggression and Alienation) that are weaker markers of NE. It is probable that the inclusion of these subscales resulted in a lower correlation than would have been observed between BIS and Stress Reaction, as an index of trait anxiety, alone. Indeed, in our own large (N = 1515) sample of offenders (Poythress, Skeem, et al., 2008), the correlation between BIS and a measure of trait anxiety was substantially stronger, r = .449.

Newman and Malterer also reported regression analyses in two studies showing that associations for the BIS scale with another purpose-built BIS measure remained significant even after controlling for measures of anxiety/NE. They concluded: “Clearly, the BIS scale captures unique variance beyond that associated with anxiety, neuroticism, and NE” (Newman & Malterer, 2009, p. 674). Relatedly, they reported that in the study by Newman, MacCoon, Vaughn, and Sadeh (2005) “the BIS scale scores of primary psychopathologies were lower than those of the other groups even controlling for the effects of WAS anxiety – a finding that refutes the possibility that the association reflected nothing more than shared variance” (p. 675).

Although these findings are intriguing, their meaning is unclear. They do not indicate whether, or the extent to which, the unique variance captured by the BIS scale assesses fear sensitivity – the specific behavioral inhibition construct central to Lykken’s “weak BIS” theory. Instead, the unique variance may capture measure- ment error, non-overlapping aspects of NE, other BIS functions (e.g., increasing arousal; initiating the evaluation of risk), or some combination thereof. Newman and Malterer’s own data and our own cast doubt on the notion that the unique variance captured by the BIS scale assesses fear sensitivity. Newman and Malterer (p. 676) reported that the BIS scale correlated weakly (r = .190), or was uncorrelated (r = .057) with MPQ Harmavoidance (HA), a measure of fear sensitivity. In our study of 1515 offenders, the

### Table 1

Summary of studies examining structure of Carver and White’s (1994) BIS/BAS scales using CFA.

<table>
<thead>
<tr>
<th>Study</th>
<th>Samplea</th>
<th>N</th>
<th>%Female</th>
<th>Modelb</th>
<th>GFI</th>
<th>AGFI</th>
<th>CFI</th>
<th>NNFI</th>
<th>RMSEA</th>
<th>IFI</th>
<th>TLI</th>
<th>RMSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heubeck, Wilkinson, and Cologon (1998)</td>
<td>S</td>
<td>336</td>
<td>68</td>
<td>2U</td>
<td>0.68</td>
<td>0.61</td>
<td>0.60</td>
<td>–</td>
<td>–</td>
<td>0.56</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Ross, Millis, Bonebright, and Bailey (2002)</td>
<td>S</td>
<td>476</td>
<td>69</td>
<td>2F</td>
<td>0.82</td>
<td>0.78</td>
<td>0.68</td>
<td>0.64</td>
<td>0.10</td>
<td>0.69</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Heym et al. (2003)</td>
<td>S</td>
<td>303</td>
<td>78</td>
<td>2C</td>
<td>0.82</td>
<td>0.78</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Heym et al. (2004)</td>
<td>S and C</td>
<td>345</td>
<td>73</td>
<td>4C</td>
<td>0.92</td>
<td>0.89</td>
<td>0.87</td>
<td>0.83</td>
<td>0.062</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Muller and Wytykowska (2005)</td>
<td>S</td>
<td>303</td>
<td>78</td>
<td>2</td>
<td>0.82</td>
<td>0.78</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Saha and Sperniec (2006)</td>
<td>S</td>
<td>345</td>
<td>69</td>
<td>2U</td>
<td>0.76</td>
<td>0.53</td>
<td>0.11</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Caci, Deschaux, and Bayle (2007)</td>
<td>S</td>
<td>144</td>
<td>76</td>
<td>2C</td>
<td>0.66</td>
<td>0.65</td>
<td>0.63</td>
<td>0.12</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Poythress, Skeem, et al. (2008)</td>
<td>S</td>
<td>1515</td>
<td>16</td>
<td>2U</td>
<td>0.89</td>
<td>0.83</td>
<td>0.07</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*a S = student sample, S and C = student and community sample combined, O = offender sample.

*b 2- and 4-Factor model results (where specified: U = uncorrelated, C = correlated).

c Fit indices are from analyses in which six poor performing items were dropped from the measure prior to conducting the Confirmatory Factor Analysis (CFA).

GFI = Goodness of Fit Index; AGFI = Adjusted Goodness of Fit Index; CFI = Comparative Fit Index; NNFI = Non-normed Fit Index; RMSEA = Root Mean Square Error of Approximation; IFI = Incremental Fit Index; TLI = Tucker-Lewis Index; RMSR = Root Mean Square Residuals.
BIS scale correlated \( r = .449 \) with a measure of anxiety but only \( r = .084 \) with HA (Poythress, Skeem, et al., 2008). Because of the large sample size, both correlations are statistically significant; however, the BIS scale shares less than 1% of variance with HA. Even the separate two-item BIS “Fear” scale identified in our study shared less than 2% of its variance with HA \( (r = .140) \).

4. Use of the BIS scale to test Lykken’s “Weak BIS hypothesis”

In our review, we used the study by Newman et al. (2005) to illustrate potential problems with using the BIS scale to test Lykken’s theory. Newman et al. used the BIS/BAS scales “to examine the extent to which primary and secondary psychopathy identified with the PCL-R and WAS satisfy Lykken’s (1995) criteria with respect to Gray’s (1987) BIS and BAS constructs” (p. 320). The WAS is a measure of anxiety. Calling the logic of this study “extremely straight-forward” (p. 674), Newman and Malterer (2009) note that Newman et al. (2005) “followed Lykken’s prescription quite literally by selecting two widely used and well validated measures of Gray’s BIS and BAS constructs to evaluate this use of the WAS” (p. 674). In our critique of Newman et al. (2005), we concluded that anxiety related substantially to both the independent variable (low WAS scores for primary psychopathy) and dependent variable (BIS). We suggested that criterion contamination at least partially explained the “finding” that, as Lykken predicted, primary psychopathy was associated with lower BIS scores. We also noted the absence of any index of fear sensitivity in their study.

Newman and Malterer (2009) characterize our critique as “unjust” (p. 674). They imply that the threat of criterion contamination can be ignored in this study, which was designed to test a specific hypothesis derived from Lykken’s theory. We respectfully disagree; in this study, criterion contamination (i.e., of the BIS scale with anxiety) and criterion deficiency (i.e., of the BIS scale for fearlessness) undermine the hypothesis test.

In their chief substantive response to our critique of Newman et al. (2005), Newman and Malterer (2009) assert that a test of Lykken’s hypothesis does not require a specific measure of fear sensitivity. We agree with Newman and Malterer that a test of Lykken’s hypothesis does not require a particular measure of fear sensitivity (like the Harmavoidance scale). We must be careful here, however, to note what we are not agreeing to. Specifically, we do not agree that Lykken’s hypothesis can be tested without some sound measure of fear sensitivity. Under the old reinforement sensitivity theory (RST; Gray, 1982), the assessment of sensitivity to conditioned fear stimuli may be the most critical component of a valid measure of behavioral inhibition system functioning. Indeed, it is the recognition of a fear stimulus that activates the remaining functions of the BIS (i.e., suspending prepotent approach and avoidance behavior; increasing arousal; initiating the evaluation of risk). If one has a measure of behavioral inhibition system functioning that does adequately assess fear sensitivity, then a separate measure of fear sensitivity would not be needed to test Lykken’s low fear hypothesis. However, the evidence reviewed above suggests that the BIS scale is virtually uncorrelated with fear sensitivity (at least as measured by HA), and thus lacks the critical component needed to test Lykken’s low fear hypothesis. Thus, we respectfully disagree with Newman et al.’s (2005) characterization of the BIS scale as “well validated” for the purpose for which they used it.

The BIS scale almost surely has its valid uses, but the research we have reviewed strongly suggests that it is not valid for testing Lykken’s model of primary psychopathy. Hence the basis for our recommended moratorium on that particular use, which Newman and Malterer’s arguments do nothing to dissuade us against.

During this moratorium, what measure of fear sensitivity might be used to supplement the BIS scale and adequately investigate Lykken’s theory? We tentatively recommended the Harmavoidance (HA) scale in our original article. Newman and Malterer question the HA scale as “a pure measure of fearlessness” (p. 676), and note “…Gray never endorsed the Harmavoidance scale as a superior measure of the BIS construct” (p. 674). As Newman and Malterer acknowledged (p. 676), in recommending the HA scale, we were cautious, suggesting only that it “may be the best available measure of fearful temperament” (Poythress, Edens, et al., 2008, p. 273, emphasis added). We did not imply that HA is a “gold standard” for assessing fear-proneness; in fact, we believe that use of the term “gold standard” is unwarranted in virtually all personality and psychopathology research, which inevitably involves “open concepts” (Lilienfeld & Fowler, 2006). We noted that some HA items may confute preference for risky activities with avoidance of boring ones. Newman and Malterer’s further analyses delineating possible limitations of the HA scale do not refute our suggestion, but instead extend our cautionary note.

One potential alternative to the HA scales is Wolpe and Lang’s (1977) Fear Survey Schedule (FSS). In two studies, Perkins and colleagues reported moderate positive associations for the BIS scale with FSS: \( r = .301 \) (Perkins & Corr, 2006) and \( r = .517 \) (Perkins, Kemp, & Corr, 2007). Still, investigators must proceed with caution using this measure because, despite its name, the FSS was designed to assess sensitivity to a variety of unconditioned fear stimuli as well as an array of anxiety inducing stimuli. Thus, shared assessment of anxiety may partially explain the association between BIS and FSS in these studies.

We hope that Newman and Malterer (2009) are correct that “Gray never endorsed the Harmavoidance scale as a superior measure of the BIS construct” (p. 674). To our knowledge, no one else has advocated the HA scale as a measure of BIS, so why should Gray have done so? HA is a plausible measure of only one BIS function in the old RST, fear sensitivity, but not of the behavioral inhibition construct as a whole.

The relations and distinctions between fear and anxiety stand to be clarified as the field begins testing the revised RST (Gray & McNaughton, 2000), in which sensitivity to conditioned fear stimuli has been realllocated from the BIS to the flight/flight/freeze system (FFFS). In this new framework “Lykken’s theory of primary psychopathy arguably would be described … as a ‘weak FFFS’ phenomenon” (Poythress, Skeem, et al., 2008, p. 733) rather than as a ‘weak behavioral inhibition system’ phenomenon. There is a growing awareness among RST investigators that purpose-built measures for testing the old RST, including Carver and White’s (1994) BIS/BAS scales, may require replacement or revision (e.g., Corr & McNaughton, 2008; Heym et al., 2008). In this context the MPQ indices of HA and NE have been identified by leading RST theorists as “capable of distinguishing between Fear and Anxiety,” respectively (Smillie, Pickering, & Jackson, 2006).

5. Problems with the BIS scale and potential problems with Lykken’s model

More broadly, Newman and Malterer (2009) suggest that the problems we identify with the BIS scale “are as much a function of Lykken’s model as Carver and White’s measure” (p. 673). We agree that there may well be problems with Lykken’s hypothesis that primary psychopathy is marked by fearlessness. Indeed, there is some evidence for this proposition. Based on a sample of 96 male prisoners with high PCL-R scores, Hicks, Markon, Patrick, Krueger, and Newman (2004) applied model-based cluster analysis to participants’ MPQ scores and identified two variants that largely are consistent with primary (“Emotionally Stable”) and secondary
(“Aggressive”) subtypes. Contrary to hypotheses that would be derived from Lykken’s model, the primary and secondary subtypes obtained highly similar HA scores. Similarly, in our study of 691 male offenders who met diagnostic criteria for antisocial personality disorder (Poythress et al., 2007), model-based cluster analysis of variables that theoretically related to putative etiological differences among subtypes did not yield psychopathy variants that clearly conformed to Lykken’s hypothesized subtypes. Although HA scores did differentiate between the primary and secondary psychopathic variants in expected ways, our “primary” variant was marked not only by low HA, but also low anxiety and high scores on Carver and White’s BAS scales. The latter finding is inconsistent with Lykken’s hypothesis that BAS functioning for primary psychopathy would be average. Together, these results are consistent with Fowles and Dindo’s (2006) observation that “...it is not completely clear that the deficit in psychopathy is specifically one of low fear rather than a combination of low fear and low anxiety” (p. 29).

These apparent problems with Lykken’s model of primary psychopathy do nothing to salvage (a) the psychometric characteristics of the BIS scale, or (b) the appropriateness of the BIS scale for testing Lykken’s model. The title of Newman and Malterer’s critique (2009) – “Problems with the BIS/BAS Scales or Lykken’s Model of Primary Psychopathy?” – implies that these two possibilities are mutually exclusive, which is not necessarily the case. Independent of potential problems with Lykken’s model, available data indicate that the BIS scale may be multidimensional, is heavily loaded with negative emotionalism, insufficiently assesses fearlessness, and, ultimately, is an invalid “stand alone” measure for testing Lykken’s low fear hypothesis.

References
