REVIEW

Quo Vadis Forensic Neuropsychological Malingering Determinations? Reply to Drs. Bush, Faust, and Jewsbury

Christoph Leonhard[1]

Abstract

The thoughtful commentaries in this volume of Drs. Bush, Jewsbury, and Faust add to the impact of the two reviews in this volume of statistical and methodological issues in the forensic neuropsychological determination of malingering based on performance and symptom validity tests (PVTs and SVTs). In his commentary, Dr. Bush raises, among others, the important question of whether such malingering determinations can still be considered as meeting the legal Daubert standard which is the basis for neuropsychological expert testimony. Dr. Jewsbury focuses mostly on statistical issues and agrees with two key points of the statistical review: Positive likelihood chaining is not a mathematically tenable method to combine findings of multiple PVTs and SVTs, and the Simple Bayes method is not applicable to malingering determinations. Dr. Faust adds important narrative texture to the implications for forensic neuropsychological practice and points to a need for research into factors other than malingering that may explain PVT and SVT failures. These commentaries put into even sharper focus the serious questions raised in the reviews about the scientific basis of present practices in the forensic neuropsychological determination of malingering.

The thoughtful commentaries of Drs. Bush (2023), Faust (2023), and Jewsbury (2023) outline the impact of these reviews (Leonhard, 2023a, b) on forensic neuropsychological determinations of malingering. I will briefly reflect on each commentary before discussing their combined impact

Questioning What We Thought We Knew: Reply to Dr. Bush

Dr. Bush focusses on important ethical and legal consequences of these reviews: Current neuropsychological practices must now be questioned in view of the serious legal consequences of determinations of malingering in civil and criminal contexts. The key issue is whether related expert testimony will remain admissible under the Daubert standard. I have co-authored a follow-on article (Leonhard & Leonhard, 2023) which explores this in detail.

That article discusses the credibility of claimants as a pivotal issue. In our judicial system, the jury assesses the credibility of claimants. In their expert testimony on malingering, neuropsychologists claim to have a scientific method to determine when someone is malingering and is therefore not credible. So far, courts have admitted such testimony under the Daubert Standard, accepting assertions that PVTs and SVTs are scientifically valid "objective" tests to determine malingering mainly because they are widely accepted and peer reviewed. The findings of these reviews undermine these core assertions. Our follow-on article (Leonhard & Leonhard, 2023) discusses in detail why malingering determinations based on PVTs and SVTs should not be admitted under the Daubert Standard and why the Daubert Standard failed as a gatekeeper.

Our article cautions against admission of related expert testimony because, cloaked with the appearance of scientific validity and objectivity, juries could be unduly influenced to accept experts' conclusions instead of relying on their own assessment of claimants' credibility. Furthermore, use of PVTs and SVTs compels criminal defendants to be witnesses against themselves in violation of their Fifth Amendment right against self-incrimination.

✉ Christoph Leonhard
chleonhard@yahoo.com

1 The Chicago School of Professional Psychology, Xavier University of Louisiana, New Orleans, USA

Regarding implications for the field of neuropsychology, I agree that professional organizations will want to examine how the information in these reviews will affect forensic neuropsychological practice. However, in its recent update of its consensus statement on neuropsychological assessment of effort, response bias, and malingering, the American Academy of Clinical Neuropsychology (Sweet et al., 2021) did not identify any of the issues raised in these reviews, except for problems with too-close-to-call cases where the consensus was that more research was needed.

Untenable Inferences: Reply to Dr. Jewsbury

Following a somewhat different path, Dr. Jewsbury agrees that positive likelihood chaining is not mathematically defensible to combine findings from multiple PVTs and SVTs to determine malingering. His commentary adds the important insight that this erroneous method guarantees that neuropsychological examinees will be found to be malingering if at least four PVTs indicate PVT^+ regardless of the number of PVTs that indicate that the examinee is not malingering. This is analogous to a ratchet tool: When set to tighten the screw, any movement to the right tightens it while movements to the left have no effect. The ratchet effect is echoed in findings that common PVTs are guaranteed to determine an examinee to be either malingering or non-disabled because the cutoff for malingering is higher than the cutoff for disability (Erdodi & Lichtenstein, 2017). A situation where no examinee can be shown to be non-malingering and disabled.

Given the Daubert standard's reliance on peer reviewed publications, as consensus develops that positive likelihood chaining is mathematically erroneous, researchers and journal editors may want to evaluate whether publications that advocate for positive likelihood chaining should have corrigenda issued or be retracted. One of the requirements for retractions is "…clear evidence that the findings are unreliable, … as a result of major error (e.g., miscalculation…)" (COPE Council, 2019, p. 2). Not correcting or retracting such papers will allow expert witnesses to continue to present malingering determinations based on positive likelihood chaining as settled science, and this erroneous method will continue to be promoted in peer reviewed publications (cf. e.g., Roor et al., 2023, p. 17).

Dr. Jewsbury also agrees that Simple Bayes lacks utility in the determination of malingering status. However, except for call for use of Simple Bayes in a book chapter (Bender & Frederick, 2018) apparently partially based on a conference presentation by Dr. Frederick from 2015, there would not appear to be any calls for its use nor does anyone appear to have actually used it. Malingering researchers (Chafetz,

2020; Larrabee et al., 2019, pp. 1357 & 1368) have also specifically rejected Frederick's call for its use. Discussion of Simple Bayes is therefore of limited applicability.

The discussion of the concept of "correlation" in the evaluation of conditional and unconditional independence of PVTs and SVTs is problematic. PVT scores are always highly skewed (Leonhard, 2023a, p. 26) and are dichotomized to predict malingering. Therefore, (in)dependence of malingering predictors should be evaluated with the χ^2 test, not based on correlations. If significant lack of independence is found, tetrachoric correlation can estimate the magnitude of the association.

The many narrative definitions of conditional dependence in the commentary are difficult to reconcile. However, there is a mathematical definition of conditional independence as shown in the statistics paper (Leonhard, 2023a, Supplemental Appendix, p. 9). Using this formula, I concluded that PVTs and SVTs with the operating characteristics reported in the malingering literature are neither conditionally nor unconditionally independent. The commentary would appear to eventually agree with this (Jewsbury, 2023, pp. 12–13). It is difficult, however, to reconcile the many claims in the commentary that conditional dependence cannot be deduced from unconditional dependence, with writing (page 12) that *unconditional* correlations > .35 should be taken as evidence that the *conditional* independence assumption cannot be met—note the average PVT and SVT correlation found in the review was .92.

The attempted extension of Bayes' theorem to prediction from more than one predictor is unfortunately a mathematical impossibility, and the reason why Bayesian statistics have not been widely adopted (cf. Bolstad & Curran, 2016, pp. 434–435). As stated in the commentary (page 3), Bayes' theorem with three predictors requires knowledge of 14 unknown constants to calculate the posterior. The problem is later elegantly skirted by specifying uninformative priors (prevalence = 0.5) and unrealistic sensitivity and specificity values (both at .85 or .93), a situation where Bayesian updating is irrelevant and reverts to simple conditional probability. Regarding musings whether Bayesian computations may be modified to solve this problem (page 15). As discussed in the statistics review (Leonhard, 2023a, Footnote 10), Markov chain Monte Carlo algorithms do exactly that (cf. Al-Khairullah & Al-Baldawi, 2021). However, these still require *unconditionally* non-collinear predictors (Bayman & Dexter, 2021, pp. 362 & 364) and their diagnostic accuracy is no better than logistic regression (Witteveen et al., 2018).

Another unfortunate claim is that high correlation (collinearity) among predictors does not present an obstacle to their use to improve prediction of outcomes (Jewsbury, 2023, pp. 10 & 11). Highly correlated predictors cannot

improve prediction under any method, Bayesian or otherwise. Many multivariate statistics texts offer detailed related mathematical analysis, including Chatterjee and Simonoff (2012, pp. 26–28), Hocking (2013, pp. 142–143), and Fox (1991, pp. 10–11). Let me illustrate this key point with an analogy. One study found, a man weighing ≤ 63.6 kg lives an average of 7.72 years longer than a man weighing ≥ 90.9 kg (Samaras & Storms, 1992, p. 258). Yet, there are many other important predictors of longevity including blood pressure, cholesterol, smoking status, gender, etc. (Risk Assessment Workgroup, 2013). A body weight estimate obtained from a bathroom scale thus does predict longevity. But does prediction improve if additional weight estimates from other scales, say at a gym or a health clinic are also considered? Despite the weight estimate becoming marginally more accurate, the longevity prediction will not improve unless factors that are not collinear with weight, such as smoking or blood pressure, are also considered. Prediction of malingering from PVT scores is analogous: a single PVT score is usually an insufficient predictor of malingering (cf. Sherman et al., 2020). A man weighing ≥ 90.9 kg does not lose an additional 7.72 years of life expectancy each time a new scale confirms the weight. Analogously, indications of malingering based on more than one PVT do not increase the likelihood of malingering.

When discussing the conditional independence requirement for Simple Bayes (e.g., Jewsbury, 2023, p. 5), artificial intelligence (AI) applications of Simple Bayes are conflated with computation of posteriors. In AI, independence is unimportant because the aim is not to compute posteriors but to obtain classifications (cf. Domingos & Pazzani, 1997). The numerical values obtained through Simple Bayes in AI are, however, inaccurate overestimations of the posteriors. Conditional independence is therefore not just convenient, it is key to accurate calculations of posterior probabilities (Hand & Yu, 2001, p. 388; Zadora et al., 2014, p. 209).

Advancing the Discussion: Reply to Dr. Faust

Dr. Faust adds important narrative texture to several implications for determinations of malingering:

Some PVTs and SVTs, despite their names, may well be more appropriately conceptualized as effort tests or response set scales. However when making forensic determinations of malingering following, Larrabee et al. (2007) and many others, 47.8% of forensic neuropsychologists rely exclusively on PVT and SVT scores to determine malingering (Schroeder et al., 2016, p. 526) and 99% consider their use mandatory (Schroeder et al., 2016, p. 748). If this remains unexamined in the peer reviewed literature, it will continue to meet the Daubert standard.

Dr. Faust's calls for additional research on factors other than malingering that may explain PVT and SVT failure. This issue has received little attention in the malingering literature. Exceptions include one study (Henry et al., 2018, p. 740) which shows that cogniphobia, common among forensic examinees, predicts PVT performance and another (Batt et al., 2008) which found 45 to 75% of patients with various brain injuries fail PVTs when they are distracted.

There is also much relevant work in neighboring fields. For example, the cognitive load on forensic examinees' working memory may be an issue. Because of the high stakes, forensic examinees face a dual tracking task as they track cognitively challenging tests while also monitoring the effect they are having on the examiner. However, dual tracking suppresses cognitive performance (Chen & Bailey, 2020; Heyselaar & Segaert, 2019), particularly if the second task is effect monitoring (Wirth et al., 2018). Affective arousal may be another factor because it is likely greater among forensic examinees than among clinical patients. The Yerkes and Dodson Law (1908) regarding the inverted U relationship between arousal and cognitive performance explains why affective arousal significantly impacts memory performance (cf. Hidalgo et al., 2019). A high-stakes forensic neuropsychological examination may well be a pertinent acute psychosocial stressor causing such arousal. Converging evidence comes from the physiological synchronization literature which found that negative affectivity suppresses cognitively complex performance (Bevilacqua et al., 2019; Stuldreher et al., 2020). Facing a forensic examiner is an adversarial situation that may well engender such negative affectivity in a forensic examinee.

Dr. Faust asked me to clarify my note of caution against putting too much stock into any specific number in these reviews and to consider them mostly for their probative value. This is because numbers derived from such flawed research methods and often also based on erroneous calculations, should be interpreted with great caution. Therefore, for example, the true base rate of malingering is unknown as is the true validity of any malingering detection method. The conclusions of the reviews stand, even if the numbers used to reach them were only assumed arguendo.

Finally, let me address my use of the term construct validity. In research methodology, this term is used in two different contexts. It may refer to whether mechanisms of action or processes that relate predictors to outcomes are well understood (Kazdin, 2017, p. 51). In measurement theory, exploration of construct validity often begins with an examination of the convergence and divergence of an index test with scores from other tests purporting to measure similar vs. dissimilar constructs (Kazdin, 2017, p. 251). In my review, construct validity is defined in the latter sense and contrasted with criterion referenced validity.

In conclusion, these reviews (Leonhard, 2023a, b) and the commentaries (Bush, 2023; Faust, 2023; Jewsbury, 2023) raise serious questions about the scientific basis of present practices in the forensic neuropsychological determination of malingering. Let me end with another analogy: PVTs and SVTs are to neuropsychological exams as the control (C) line is to lateral flow rapid antigen COVID-19 tests. When the C line does not appear, the test cannot validly diagnose COVID regardless of what the test (T) line shows. But it remains an open question whether the C line fails to appear because the patient was malingering COVID or for some other reason.

Author Contribution Not applicable because this manuscript has only a single author.

Data Availability Not applicable.

Declarations

Ethics Approval Not applicable because this is a reply to commentaries and not a study where human or animal data were collected.

Competing Interests The author declares no competing interests.

References

Al-Khairullah, N. A., & Al-Baldawi, T. H. K. (2021). Bayesian computational methods of the logistic regression model. *Journal of Physics: Conference Series, 1804*(1), 012073. https://doi.org/10.1088/1742-6596/1804/1/012073

Batt, K., Shores, E. A., & Chekaluk, E. (2008). The effect of distraction on the Word Memory Test and Test of Memory Malingering performance in patients with a severe brain injury. *Journal of the International Neuropsychological Society : JINS, 14*(6), 1074–1080. https://doi.org/10.1017/S135561770808137X

Bayman, E. O., & Dexter, F. (2021). Multicollinearity in logistic regression models. *Anesthesia and Analgesia, 133*(2), 362–365. https://doi.org/10.1213/ANE.0000000000005593

Bender, S. D., & Frederick, R. (2018). Neuropsychological models of feigned cognitive deficits. In S. D. Bender & R. Rogers (Eds.), *Clinical assessment of malingering and deception* (Fourth edition., pp. 42–60). The Guilford Press.

Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., Poeppel, D., & Dikker, S. (2019). Brain-to-brain synchrony and learning outcomes vary by student–teacher dynamics: Evidence from a real-world classroom electroencephalography study. *Journal of Cognitive Neuroscience, 31*(3), 401–411. https://doi.org/10.1162/jocn_a_01274

Bolstad, W. M., & Curran, J. M. (2016). *Introduction to Bayesian statistics* (Third edition.). Wiley Blackwell.

Bush, S. S. (2023). Questioning what we thought we knew: Commentary on Leonhard's performance validity assessment articles. *Neuropsychology Review*.

Chafetz, M. D. (2020). Deception is different: Negative validity test findings do not provide "evidence" for "good effort." *The Clinical Neuropsychologist*, 1–37. https://doi.org/10.1080/13854046.2020.1840633

Chatterjee, S., & Simonoff, J. S. (2012). *Handbook of regression analysis*. John Wiley & Sons.

Chen, E. H., & Bailey, D. H. (2020). Dual-task studies of working memory and arithmetic performance: A meta-analysis. *Journal of Experimental Psychology. Learning, Memory, and Cognition.* https://doi.org/10.1037/xlm0000822

COPE Council. (2019). *COPE retraction guidelines—English*. Committee on Publication Ethics. https://doi.org/10.24318/cope.2019.1.4

Domingos, P., & Pazzani, M. (1997). On the optimality of the Simple Bayesian Classifier under zero-one loss. *Machine Learning, 29*, 103–130.

Erdodi, L. A., & Lichtenstein, J. D. (2017). Invalid before impaired: An emerging paradox of embedded validity indicators. *The Clinical Neuropsychologist, 31*(6–7), 1029–1046. https://doi.org/10.1080/13854046.2017.1323119

Faust, D. (2023). Invited Commentary: Advancing but not yet advanced: Assessment of effort/malingering in forensic and clinical settings. *Neuropsychology Review*.

Fox, J. (1991). *Regression Diagnostics*. SAGE Publications, Inc. https://doi.org/10.4135/9781412985604

Hand, D. J., & Yu, K. (2001). Idiot's Bayes: Not so stupid after all? *International Statistical Review / Revue Internationale De Statistique, 69*(3), 385–398. https://doi.org/10.2307/1403452

Henry, G. K., Heilbronner, R. L., Suhr, G., & J., Wagner, E., & Drane, D. L. (2018). Illness perceptions predict cognitive performance validity. *Journal of the International Neuropsychological Society: JINS, 24*(7), 735–745. https://doi.org/10.1017/S1355617718000218

Heyselaar, E., & Segaert, K. (2019). Memory encoding of syntactic information involves domain-general attentional resources: Evidence from dual-task studies. *Quarterly Journal of Experimental Psychology, 72*(6), 1285–1296. https://doi.org/10.1177/1747021818801249

Hidalgo, V., Pulopulos, M. M., & Salvador, A. (2019). Acute psychosocial stress effects on memory performance: Relevance of age and sex. *Neurobiology of Learning and Memory, 157*, 48–60. https://doi.org/10.1016/j.nlm.2018.11.013

Hocking, R. R. (2013). *Methods and applications of linear models: Regression and the analysis of variance*. John Wiley & Sons.

Jewsbury, P. (2023). Inference with multiple tests: Commentary to Leonhard. *Neuropsychology Review*.

Kazdin, A. E. (2017). *Research Design in Clinical Psychology* (5th ed.). Pearson.

Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Redefining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits*. Oxford University Press.

Larrabee, G. J., Rohling, M. L., & Meyers, J. E. (2019). Use of multiple performance and symptom validity measures: Determining the optimal per test cutoff for determination of invalidity, analysis of skew, and inter-test correlations in valid and invalid performance groups. *The Clinical Neuropsychologist, 33*(8), 1354–1372. https://doi.org/10.1080/13854046.2019.1614227

Leonhard, C. (2023a). Review of statistical and methodological issues in the forensic prediction of malingering from validity tests: Part I: Statistical issues. *Neuropsychology Review*.

Leonhard, C. (2023b). Review of statistical and methodological issues in the forensic prediction of malingering from validity tests: Part II: Methodological issues. *Neuropsychology Review*.

Leonhard, C., & Leonhard, C. (2023). Neuropsychological malingering determinations: Science or fiction of lie detection? *Georgia Law Review*, 58(2).

Risk Assessment Workgroup. (2013). *Assessing Cardiovascular Risk* (p. 139). National Heart, Lung, and Blood Institute. Retrieved May 5, 2023, from https://www.nhlbi.nih.gov/sites/default/files/media/docs/risk-assessment.pdf

Roor, J. J., Peters, M. J. V., Dandachi-FitzGerald, B., & Ponds, R. W. H. M. (2023). Performance validity test failure in the clinical population: A systematic review and meta-analysis of prevalence rates. *Neuropsychology Review*. https://doi.org/10.1007/s11065-023-09582-7

Samaras, T. T., & Storms, L. H. (1992). Impact of height and weight on life span. *Bulletin of the World Health Organization, 70*(2), 259–267.

Schroeder, R. W., Martin, P. K., & Odland, A. P. (2016). Expert beliefs and practices regarding neuropsychological validity testing. *Clinical Neuropsychologist, 30*(4), 515–535. https://doi.org/10.1080/13854046.2016.1177118

Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology*, acaa019. https://doi.org/10.1093/arclin/acaa019

Stuldreher, I. V., Thammasan, N., Erp, J. B. F. van, & Brouwer, A.-M. (2020). Physiological synchrony in EEG, electrodermal activity and heart rate reflects shared selective auditory attention. *Journal of Neural Engineering*, *17*(4), 046028. https://doi.org/10.1088/1741-2552/aba87d

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., & Suhr, J. A. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 35*(6), 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

Wirth, R., Janczyk, M., & Kunde, W. (2018). Effect monitoring in dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(4), 553–571. https://doi.org/10.1037/xlm0000474

Witteveen, A., Nane, G. F., Vliegen, I. M. H., Siesling, S., & IJzerman, M. J. (2018). Comparison of logistic regression and Bayesian networks for risk prediction of breast cancer recurrence. *Medical Decision Making, 38*(7), 822–833. https://doi.org/10.1177/0272989X18790963

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology, 18*, 459–482. https://doi.org/10.1002/cne.920180503

Zadora, G., Martyna, A., Ramos, D., & Aitken, C. (2014). *Statistical analysis in forensic science: Evidential value of multivariate physicochemical data*. John Wiley & Sons Inc.