On the Adequacy of Bayesian Evaluations of Categorization Models: Reply to Vanpaemel & Lee (2012).

Andy J. Wills

University of Exeter

Emmanuel M. Pothos

Swansea University

Author Note

Andy J. Wills, Psychology, University of Exeter; Emmanuel M. Pothos, Psychology, Swansea University.

Correspondence concerning this article should be addressed to Andy Wills, Psychology, University of Exeter, Perry Road, Exeter. EX4 4QG. United Kingdom. Email: andy@willslab.co.uk

Psychological Bulletin

MS 2012-0911

Invited Reply to MS 2011-0818

Final Version

24 April 2012

Abstract

Vanpaemel and Lee (2012) argue, and we agree, that the comparison of formal models can be facilitated by Bayesian methods. However, Bayesian methods neither precede nor supplant our proposals (Wills & Pothos, 2012), as Bayesian methods can be applied both to our proposals and to their polar opposites. Furthermore, the use of Bayesian methods to control for model complexity can be actively misleading when combined with the consideration of narrow data sets, and significant development work is required before Bayesian methods can be applied to some of the leading formal models of categorization. Even where Bayesian methods can be applied, the use of non-Bayesian methods is sometimes preferable due to their computational simplicity (Vanpaemel & Storms, 2010). We also clarify our position on arbitrarily variable parameters, and on the relationship between ordinal properties and overfitting.

Vanpaemel and Lee (2012) argue that the comparison of formal models of categorization can be facilitated by Bayesian methods of model selection. Despite some controversy over Bayesian methods (e.g. Gilboa, 2009), we tend to agree with this general point. Indeed, it is a point we have already made (Wills & Pothos, 2012, pp. 119-120). However, we disagree with many of the specific claims made by Vanpaemal and Lee, as we outline below. In brief, Bayesian methods can be applied in a manner inconsistent with our proposals just as easily as they can be applied in a manner consistent with them.

In the target article (Wills & Pothos, 2012), we argued that model comparisons are most fruitful when relative adequacy is assessed by comparing well-defined models on the basis of the number and proportion of irreversible, ordinal, penetrable successes. The central concepts of this argument are therefore as follows: (1) that models should be well-defined, (2) that models should be penetrable, (3) that ordinal success is the primary goal (although not discounting the importance of quantitative closeness, Wills & Pothos, 2012, p. 112), (4) that models should be compared on the number (and proportion) of ordinal successes, (5) that models' successes should be irreversible, in the sense of the avoidance of arbitrarily variable parameters (Wills & Pothos, 2012, pp. 112-113). Vanpaemel and Lee (2012) recast our proposals as being about (a) avoiding over-fitting, (b) taking qualitative properties of data seriously, (c) reducing dependence on free parameters, and (d) testing empirical breadth.

We note first that Vanpaemel and Lee's (2012) recasting of our proposals is narrower than the original, choosing not to address the proposals that models should be well-defined and penetrable. They also elevate avoidance of overfitting from one of a number of reasons for favoring ordinal properties (in our paper), to a central principle (in their Comment). Putting this elevation to one side, the substantive area of disagreement on over-fitting between us is that we asserted that “adopting ordinal adequacy as the primary measure of success also reduces (but does not necessarily eliminate) the risks of illusory model superiority due to overfitting” (Wills, & Pothos, 2012, p.111), whereas they claimed that a complicated model can “over-fit different ordinal data patterns as easily as it can over-fit the quantitative details” (Vanpaemel & Lee, 2012, p.18). The thinking behind our assertion was that the bulk of empirical investigation of psychological phenomena proceeds through the establishment of robust ordinal phenomena, and thus a model comparison process that concentrates on empirically robust phenomena would be at less risk of fitting noise, than one that attempted to fit the minutiae. Vanpaemel and Lee provide no specific reasons to doubt this viewpoint.

Of more concern is that Vanpaemel and Lee (2012) appear to have misconstrued the concept of irreversible success and its achievement through the avoidance of *arbitrarily variable* free parameters. Specifically, despite what is implied by their Comment, we did not endorse the idea that “parameter values are meant to change across experimental conditions” (Vanpaemel & Lee, 2012, p. 23), nor did we advocate “fixing each to a single value before seeing the data”(p. 23). The former approach, as we argued in the original paper, negates most of the advantages of formal models over informal ones (Wills & Pothos, 2012, p. 121). The latter approach is unrealistic, as Vanpaemel and Lee state and we agree. We did discuss an example where parameters could be removed from the Generalized Context Model (GCM) through the assumption that attention maximizes categorization accuracy (Nosofsky, 1984). In discussing this example, we did not state, and did not intend to imply, that it was necessarily possible to reduce the number of free parameters to zero.

To use the terminology of the Comment, our proposal is that parameters should be *global* – in other words, they should be determined at the level of the domain of phenomena that the model is intended to address, not at the level of individual experiments. Determination of global parameters by parameter estimation is both reasonable and likely to be necessary. In the original article, we noted the potential of hierarchical Bayesian methods for fitting individual and group average data simultaneously (Wills & Pothos, 2012, p.119). In their Comment, Vanpaemel and Lee (2012) state that the same methods can also be used for the estimation of global parameters. We thank them for highlighting this important point.

We do not, however, endorse their statement that Bayesian methods provide "the best current answers" (Vanpaemel & Lee, 2012, p. 14) to the issues we had raised. Nor do we agree with their stronger expressions of this view, which might be taken to imply that Bayesian methods precede and replace the proposals we have made (in particular the final paragraph of their Comment seems to make this point). To say that a method provides answers to these sorts of issues is much like saying a car tells you where to go on vacation. The driver, not the car, dictates the direction of travel. As noted, we accept that the Bayesian framework, in principle, provides tools for doing some of the things we believe are important – in particular, hierarchical Bayesian methods provide, in principle, a method for taking empirical breadth seriously. But, as Vanpaemel and Lee seem to accept (at least at one point, p. 27), Bayesian methods provide only a framework. One can apply Bayesian methods to the fitting of narrow data sets with arbitrarily variable parameters just as easily as one can apply Bayesian methods in a way consistent with our proposals.

Also, like all tools, Bayesian methods have the potential to be used inappropriately. We therefore disagree that "the correct application of Bayesian inference automatically controls for model complexity" (Vanpaemel & Lee, 2012, p. 27). The use of methods that compensate for model complexity can be actively misleading, when combined with the currently prevalent approach of considering narrow data sets. It is not difficult to devise situations where a model is more complex than it needs to be to accommodate a narrow data set even where, in the context of a broader data set, the more complex model is the more adequate one.

A recurring theme of the Comment is that Bayesian methods not only have the potential to facilitate our proposals but that they have also already been used to do so. The latter statement, at least within categorization research, is largely inaccurate. It is clear from the Comment (Vanpaemel & Lee, 2012, p. 27), and from other writings (e.g. Lee, 2008) that any learning model poses significant challenges for a Bayesian framework; challenges that, as far as we are aware, no one has yet surmounted within categorization. The class of learning models includes prominent categorization models such as ALCOVE (Kruschke, 1992), COVIS (Ashby, Paul & Maddox, 2011) and SUSTAIN (Love, Medin & Gureckis, 2004). Thus it seems that there are significant technical challenges to overcome before Bayesian methods can be applied to comparison problems in the formal modeling of categorization. In practice, we are not aware of any extensive model comparisons in categorization using Bayesian methods. Let us hasten to add that we do hope that situation changes because, as we have already accepted, the Bayesian framework has clear theoretical advantages.

Are there any situations where, if both Bayesian and non-Bayesian methods were available, one would favor non-Bayesian methods? According to Vanpaemel and Storms (2010), the answer is "yes." Vanpaemel and Storms (2010) compared Bayesian and non-Bayesian methods of parameter estimation within the Varying Abstraction Model of categorization (VAM, Vanpaemel & Storms, 2008). They concluded that the Bayesian method was "uncalled for" and that the non-Bayesian method was "justified" (p. 421), largely due to the computational simplicity of the non-Bayesian method—and because of the fact that the variants of the VAM model were of comparable complexity to each other. At least at present, using Bayesian methods is simply more technically challenging (indeed, often requiring very specialized expertise) and time consuming than using non-Bayesian methods, and even advocates of Bayesian methods sometimes counsel against their use for this reason (as in Vanpaemel and Storms, 2010). We agree.

Vanpaemel and Lee (2012) also make the more specific claim that Bayesian methods have already been used to reduce dependence on free parameters in the evaluation of formal models of categorization, and they cite work on the VAM in support of this claim. However, this work is characterized by the independent estimation of model parameters for each of up to thirty different experiments (e.g. Lee & Vanpaemel, 2008), so it seems to have achieved little thus far in reducing reliance on *arbitrarily variable* parameters. The supportable aspect of Vanpaemel and Lee's claim is that Bayesian methods provide a way of specifying prior assumptions about the distribution from which parameters are drawn. In that sense, the parameters are not entirely “free” - there are some constraints on the relative likelihood with which values of arbitrarily variable parameters are selected. As the example of Vanpaemel and Lee (2008) illustrates, this does not prevent the parameters from being arbitrarily variable. And, of course, the parameters were never really “free” in the first place, as non-Bayesian methods have priors too. Bayesian methods just force one to be explicit about the priors. This is, of course, commendable.

One aspect upon which we entirely agree with Vanpaemel and Lee (2012) is the importance of taking the ordinal properties of data seriously. They are critical of our proposal in this regard mainly because we neglected to provide a specific formal method for doing so. The only specific suggestion made by Vanpaemel and Lee was to use parameter-space partitioning (Pitt, Myung, Montenegro, & Pooley, 2008), a procedure that was also briefly mentioned in our original article for other reasons (Wills & Pothos, 2012, p. 120). Parameter-space partitioning is a way of assessing model flexibility through an examination of the extent to which varying the model’s parameters changes the qualitative (typically ordinal) predictions that it makes. Hence, parameter-space partitioning takes the ordinal properties of data seriously, but it does not appear to directly provide a method of comparing formal models on the basis of the number of irreversible ordinal successes.

In summary, we agree that Bayesian methods provide some tools that might be useful in the pursuance of the proposals made in Wills and Pothos (2012). However, Bayesian methods are general tools that could also be used in the pursuance of proposals largely opposite to the ones we set out. Furthermore, there are important challenges in the application of Bayesian methods, compared to non-Bayesian ones, both regarding technical feasibility and practicality, and significant development work is required before they could be applied to some of the leading formal categorization models discussed by Wills and Pothos (2012). We hope that such development work is carried out in due course and that it results in comparison tools that are available and accessible to a wide range of researchers.

References

Ashby, F., Paul, E., & Maddox, W. T. (2011). COVIS. In E. M. Pothos & A. J. Wills (Eds.). *Formal approaches in categorization* (pp. 65-87). Cambridge: Cambridge University Press.

Gilboa, I. (2009). *Theory of decision under uncertainty (Econometric Society Monographs)*. Cambridge University Press.

Kruschke, J.K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44. doi: 10.1037/0033-295X.99.1.22

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, *15*(1), 1-15. doi:10.3758/PBR.15.1.1

Lee, M. D., & Vanpaemel, W. (2008). Exemplars, prototypes, similarities, and rules in category representation: an example of hierarchical Bayesian analysis. *Cognitive Science*, *32*, 1403-24. doi: 10.1080/ 03640210802073697

Love, B.C., Medin, D.L., & Gureckis, T.M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*, 309-332. doi: 10.1037/0033-295X.111.2.309

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *10*, 104-14. doi:10.1037/0278-7393.10.1.104

Pitt, M.A., Myung, J. I., Montenegro, M., & Pooley, J. (2008). Measuring model flexibility with parameter space partitioning: an introduction and application example. *Cognitive Science*, *32*, 1285-1303. doi:10.1080/03640210802477534.

Vanpaemel, W., & Lee, (2012). The Bayesian evaluation of categorization models: Comment on Wills and Pothos (2012). *Psychological Bulletin*.

Vanpaemel, W., & Storms, G. (2008). In search of abstraction: The varying abstraction model of categorization. *Psychonomic Bulletin & Review*, *15*(4), 732-749. doi:10.3758/PBR.15.4.732

Vanpaemel, W., & Storms, G. (2010). Abstraction and model evaluation in category learning. *Behavior Research Methods*, *42*, 421-37. doi:10.3758/BRM.42.2.421

Wills, A. J., & Pothos, E. M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychological Bulletin*, *138*, 102-125. doi:10.1037/a0025715