Running head: COMMENT ON WILLS AND POTHOS (2012)

The Bayesian Evaluation of Categorization Models: Comment on Wills and Pothos (2012)

Wolf Vanpaemel

Faculty of Psychology and Educational Sciences

University of Leuven

Michael D. Lee

Department of Cognitive Sciences

University of California, Irvine

Address correspondence to:

Wolf Vanpaemel

Faculty of Psychology and Educational Sciences, University of Leuven

3000 Leuven, Belgium.

Telephone: (32) 16 326256.

Facsimile: (32) 16 325993.

Electronic Mail: wolf.vanpaemel@ppw.kuleuven.be

Abstract

Wills and Pothos (2012) review approaches to evaluating formal models of categorization, raising a series of worthwhile issues, challenges, and goals. Unfortunately, in discussing these issues and proposing solutions, Wills and Pothos (2012) do not consider Bayesian methods in any detail. This means not only that their review excludes a major body of current work in the field, but also that it does not consider the body of work that provides the best current answers to the issues raised. In this comment, we argue that Bayesian methods can be—and, in most cases, already have been—applied to all the major model evaluation issues raised by Wills and Pothos (2012). In particular, Bayesian methods address the challenges of avoiding over-fitting, considering qualitative properties of data, reducing dependence on free parameters, and testing empirical breadth.

Introduction

In their review, Wills and Pothos (2012) raise a number of challenges for the evaluation of formal models in psychology, focusing on models of categorization as a case study. In particular, they argue that the evaluation of psychological models should avoid over-fitting, take qualitative properties of data seriously, reduce dependence on free parameters, and test empirical breadth. We fully agree these are worthwhile goals, and suspect that many others do too.

Wills and Pothos (2012) present a series of recommendations to address these challenges. Unfortunately, their proposals fail to give serious consideration to Bayesian methods. This is a clear omission in the context of a review, since Bayesian methods have been advocated prominently as a means of evaluating psychological models in general (e.g., Kruschke, 2010, 2011; Lee, 2004; Lee & Wagenmakers, 2005; I. J. Myung & Pitt, 1997; Pitt, Myung, & Zhang, 2002; Shiffrin, Lee, Kim, & Wagenmakers, 2008; Vanpaemel, 2010), and have been applied to the evaluation of categorization models in particular (Lee, 2008; Lee & Vanpaemel, 2008; Lee & Wetzels, 2010; Vanpaemel, 2011; Vanpaemel & Lee, 2007; Vanpaemel & Storms, 2010; Voorspoels, Storms, & Vanpaemel, 2011).

More importantly, as we argue in this comment, it is a serious omission in the context of making good recommendations. Bayesian methods provide the principles and tools needed to address all four of the major issues discussed by Wills and Pothos (2012). In this comment, we discuss each issue in turn. For each, we explain how Bayesian methods address the problem, cite existing research in psychology that uses Bayesian methods to address the problem, and discuss why we consider the Bayesian approach preferable to the approaches proposed by Wills and Pothos (2012).

Bayesian Approaches to Four Model Evaluation Issues

Bayesian statistics is a complete, coherent, and rational approach to representing uncertainty and information, based on the axioms of probability theory (Cox, 1961; Jaynes, 2003). It is perfectly suited to the problem of drawing inferences from data using formal models, and for making inferences and decisions about the adequacy of models based on data. These abilities mean that Bayesian methods can address all of the major issues raised by Wills and Pothos (2012). The Bayesian approach has advantages on both principled and practical grounds. Some of the issues raised by Wills and Pothos (2012) are automatically solved by adopting Bayesian methods, because of their inherent properties. Other issues raised by Wills and Pothos (2012) can be solved in practice using Bayesian methods, because they allow researchers to implement richer and more general evaluations of psychological models.

Avoiding Over-fitting

Wills and Pothos (2012) are rightly concerned about the issue of over-fitting. This occurs when a model can describe existing data very well, but does not generalize well to new or related data. Over-fitting is usually caused by a model being too complicated, in the sense that it is able to describe many different patterns of data. It is well understood in cognitive modeling that over-fitting should be avoided (e.g., Pitt & Myung, 2002; Nosofsky & Zaki, 2002). Standard quantitative criteria that consider the best possible fit of a model, such as the sum-squared error (SSE) measure highlighted by Wills and Pothos (2012), are highly prone to over-fitting.

Bayesian methods can solve the problem. Bayesian methods for model evaluation directly address the issue of over-fitting. Bayesian methods include prior and posterior predictive Bayesian assessments of model adequacy, such as the Bayes Factor. The key property of these methods is that they evaluate how well a model fits data on average, over a set of parameter values, rather than how well it fits data in the best-case scenario at a single parameter value. Measures of average fit automatically balance goodness-of-fit with model complexity. More complicated models are, in their statistical definition, those that are able to predict more data patterns (I. J. Myung, Balasubramanian, & Pitt, 2000). Since a measure of average fit includes those predictions that poorly fit the observed data, more complicated models are penalized by Bayesian methods. The founding principles of Bayesian statistics thus guarantee that they address the issue of over-fitting.

Bayesian methods have solved the problem. Bayesian model evaluation methods were developed long ago in statistics (e.g. Jeffreys, 1935, 1961), and are now standard textbook material (e.g., Casella & Berger, 2002; Gelman, Carlin, Stern, & Rubin, 2004; Kass & Raftery, 1995; Robert, 1994). In psychology, Bayesian methods feature as early as Edwards, Lindman, and Savage (1963), and have been promoted heavily in the last 15 years, starting with I. J. Myung and

Pitt (1997), and growing from there (e.g., Kruschke, 2010, 2011; Lee, 2004; Pitt et al., 2002; Shiffrin et al., 2008; Vanpaemel, 2010). Many papers in the cognitive modeling literature have now used Bayesian methods to address the issue of over-fitting (e.g., Gallistel, 2009; Kemp & Tenenbaum, 2008; Steyvers, Lee, & Wagenmakers, 2009), including some focused on categorization models (e.g., Lee, 2008; Lee & Wetzels, 2010; Vanpaemel & Storms, 2010; Voorspoels et al., 2011).

Limitations of Wills and Pothos' (2012) proposals. Wills and Pothos (2012) briefly mention Bayesian model evaluation measures, but do not acknowledge they solve the problem of over-fitting. Instead, Wills and Pothos (2012, p. 111) advocate evaluating models in terms of their ability to capture ordinal properties of data as a remedy to over-fitting. A complicated model, however, can over-fit different ordinal data patterns as easily as it can over-fit the quantitative details of the data. Thus, adopting Bayesian methods appears a better solution to the problem of over-fitting than the proposal made by Wills and Pothos (2012).

Taking Qualitative Properties of Data Seriously

Wills and Pothos (2012) highlight the importance of qualitative properties of the data, such as ordinal properties, for evaluating psychological models (see also Nosofsky & Palmeri, 1997a; Pitt, Kim, Navarro, & Myung, 2006). Although we are critical of the argument that ordinal property evaluation should be preferred to reduce the risk of over-fitting, we do agree that going beyond the quantitative minutiae of the data is a worthwhile goal. In particular, it is sensible to place strong evaluative emphasis on the ability of models to describe those features of data that are the most scientifically important. These may be ordinal constraints, as focused on by Wills and Pothos (2012), or may be more general. It may be important, for example, that a model is able to capture the crossing of two learning curves, a heavy tail in a response time distribution, or a non-monotonicity in performance. Standard quantitative criteria like SSE are not naturally sensitive to these sorts of qualitative or phenomenological features of data.

Bayesian methods can solve the problem. Emphasizing some aspects of data over others is naturally done within a Bayesian framework for evaluation. The basic approach is to overlay utilities on the probabilities produced by Bayesian inference. Augmenting probabilities with utilities affords the ability to evaluate models relative to specific properties of the data, rather than strictly in terms of their likelihood given data. So, for example, if an important aspect of evaluation is that two learning curves cross, a large utility can be placed on that property, and a model will be positively evaluated to the extent that it is inferred to satisfy that property. The advantage of imposing utilities within the Bayesian framework is that the uncertainty about model parameters is automatically and completely incorporated into the utility comparisons. Thus, while Bayesian methods do not automatically solve the problem, they provide the right tools to address the problem.

Bayesian methods have solved the problem. Incorporating utility functions is the cornerstone of Bayesian decision theory, which is textbook material in statistics (e.g., Berger, 1985; Bernado & Smith, 2000). To be fair, applications of the Bayesian decision theoretic framework are hard to find in psychology. Exceptions include J. Myung and Pitt (2009) and Zhang and Lee (2010), but these authors have slightly different goals than the immediate evaluation of cognitive models. More general and flexible decision criteria are desirable in many cases of model evaluation, including in categorization, and the Bayesian solution is not used as widely as it should be. One approach that is closely related to the Bayesian approach, Parameter Space Partitioning, focuses on evaluating ordinal criteria, and has been applied to the evaluation of a range of psychological models (Pitt et al., 2006; Pitt, Myung, Montenegro, & Pooley, 2008).

Limitations of Wills and Pothos' (2012) proposals. Wills and Pothos (2012) discuss examples where it seems that primacy is given to ordinal properties of the data, but, in fact, is not. Both Nosofsky and Palmeri (1997b) and Love, Medin, and Gureckis (2004), which Wills and Pothos (2012, p. 111, p. 113, p. 119) portray as evaluating models based on their ordinal predictions, first fit the models quantitatively, using a measure related to SSE. Based on these quantitative fits, the qualitative model behavior is assessed. Hence, these studies do not directly compare the model's ordinal predictions to the ordinal properties of the data. Thus, while their discussion reinforces the point that qualitative properties are important, it does not provide methods for incorporating them directly into formal evaluation.

Reducing Dependence on Free Parameters

Wills and Pothos (2012) are concerned about free parameters in psychological models (see also Nosofsky, 1998). We agree that minimizing the dependence of successful models on free parameters is often a worthwhile goal. Free parameters tend to reduce the empirical content or assertive power of a theory (Popper, 1959). Theories with high empirical content tell us more about the world than theories low in assertive power. For models that quantify substantive theory, free parameters correspond to meaningful psychological variables and, therefore, their existence often reflects important incompleteness or immaturity in theorizing. Treating parameters as entirely free to vary, and fitting the same model parameters for multiple data sets separately, which is a standard practice, corresponds to admitting a lack of the theory needed to constrain possible values, and ignores the collective empirical information that related experiments provide about parameters.[1]

Bayesian methods can solve the problem. Bayesian methods allow—in fact, require—information about parameters to be expressed formally in a model by using prior distributions on model parameters. Some approaches to specifying parameters priors are designed to represent a default state of knowledge (Kass & Wasserman, 1996), whereas informative priors represent theoretical assumptions, or other relevant knowledge, about the psychological variables they represent. As better theories are developed, and more information about psychological variables becomes available, Bayesian methods naturally incorporate this knowledge. Informative priors automatically reduce the dependence of a model on free parameters, by making the parameters less "free", and so simultaneously decrease the flexibility and increase the empirical content of the model. Thus, Bayesian methods provide the right tools to address the problem, but do not automatically solve the problem, as many Bayesian model evaluations do not use informative priors, but rather rely on uniform, flat, or other priors intended to be "uninformative" or "weakly informative".

Bayesian methods have solved the problem. The development of formal methods for representing information in prior distributions is a major enterprise in Bayesian statistics (Bernado & Smith, 2000; Jaynes, 2003). Maximum entropy and transformational invariance methods have been developed to give principled ways of incorporating prior information (e.g., Jaynes, 2003; Tribus & Fitts, 1968). The case for constructing psychologically meaningful priors, and examples of how this can be done with psychological models, is made by Vanpaemel (2009b, 2010). Demonstrations of capturing theory in informative priors in the context of categorization models can be found in Lee and Vanpaemel (2008); Vanpaemel (2011); Vanpaemel and Lee (2007). These authors focus on the Varying Abstraction Model of category learning, which assumes that people learn categories by clustering exemplars (VAM, Vanpaemel & Storms, 2008). They show how the prior in the VAM can be used the express the theoretical assumption that exemplar clustering is likely to be driven by similarity, a theoretical position consistent with the SUSTAIN model (Love et al., 2004) and the Rational Model of Categorization (Anderson, 1991; Griffiths, Canini, Sanborn, & Navarro, 2007).

Limitations of Wills and Pothos' (2012) proposals. Wills and Pothos (2012, p. 112, p. 113) propose two remedies for the challenge of reducing the dependence of models on free parameters. One is to use universal or global parameters, whose specification is general to the whole domain of phenomena that the model is intended to address. The second is to remove the free parameters by fixing each to a single value before seeing the data.

Wills and Pothos (2012, p. 113) themselves seem to concede the first proposal is limited, as they are quick to emphasize that parameter values are meant to change across experimental conditions. Indeed, a key property of useful psychological models is selective influence, which requires that independent experimental manipulations correspond to changes in the estimated values of single model parameters. Selective influence has been evaluated in relation to a number of psychological models (e.g., Rouder et al., 2008; Voss, Rothermund, & Voss, 2004), and discussed in the context of category learning models (e.g., Vanpaemel, 2009a).

The proposal to evaluate the fit of a model at a single parameter value that is dictated by theory is a laudable goal, but seems unrealistic in practice, at least with the current state of psychological theorizing. Often, there is information about parameters available from existing data collected in related experiments or from psychological theory, but this information is unlikely to be complete enough to fix parameters to single values. Evaluating a model using a single, unrealistically precise, value often fails to incorporate important levels of uncertainty.

Wills and Pothos (2012, p. 113) discuss Nosofsky's (1984) use of theorizing about how learners allocate their attention to stimulus dimensions to fix the free attention parameters from the Generalized Context Model (GCM). Recently, Vanpaemel and Lee (submitted) demonstrated how Bayesian methods can be used to capture the attention allocation assumption in an informative prior *distribution* over parameter values rather than in a single point value. In this way, the attention parameter is neither entirely free, as in common practice, nor precisely constrained, as Wills and Pothos (2012) advocate. Instead, by using informative prior distributions in a Bayesian setting, the GCM is evaluated in a way that takes into account existing theorizing about attention, but at the same time acknowledges that this theorizing is incomplete. The Bayesian approach thus allows dependence on free parameters to be reduced, without necessitating the over-confident reduction to singe values.

Testing Empirical Breadth

Wills and Pothos (2012) argue that an important aim in model evaluation is to capture data from a broad set experiments, rather than from one or two experiments, as seems typical in practice (see also Smith & Minda, 2000). We agree that a hallmark of good theorizing and modeling throughout the empirical science is that fundamental variables and processes play roles in a wide range of experiments and phenomena. In classical physics, the underlying mass of an object affects its measured weight, acceleration, momentum, kinetic energy, and so on. In psychology, the underlying acuity of a person's memory affects their measured ability in recall tasks, recognition tasks, recollection tasks, and so on.

Bayesian methods can solve the problem. An important feature of Bayesian inference is that it seamlessly applies to hierarchical, or multilevel, models. The key idea of the hierarchical Bayesian approach is to introduce additional structure to the simple parameter-to-data relationship that characterizes most psychological models. Among the many attractive features of this approach is that it is natural to test the ability of the same model parameters and processes to account for multiple data sets from multiple experimental tasks simultaneously. This enables testing for empirical breadth. Again, the Bayesian approach, hierarchical or otherwise, does not automatically test for empirical breadth, but it does afford the possibility.

Bayesian methods have solved the problem. Hierarchical Bayesian methods were proposed long ago in statistics (e.g., Lindley & Smith, 1972), and are now standard textbook material (e.g., Gelman et al., 2004; Gelman & Hill, 2007). The general idea behind the hierarchical Bayesian approach has been advocated repeatedly in the psychological literature (e.g., Kruschke, 2010; Lee, 2006; Lee & Vanpaemel, 2008; Rouder & Lu, 2005; Shiffrin et al., 2008), and the case for hierarchical Bayesian models allowing the evaluation of empirical breadth in psychology is made by Lee (2011). Recent psychological applications of this approach to seek breadth can, for example, be found in Pooley, Lee, and Shankle (2011) who deal with the combination of recognition and recall memory, and in

Lee and Sarnecka (2011), in a developmental area closely related to categorization. These authors show how two different tasks in which children must demonstrate conceptual understanding can be modeled simultaneously in terms of common latent developmental stage variables, and different process models developed for the two different empirical tasks.

Limitations of Wills and Pothos' (2012) proposals. Wills and Pothos (2012, p. 119) propose the evaluation of models against a large set of key phenomena in a way that parameters are held constant across the whole set of phenomena, or are fixed to single values without recourse to data. As discussed above, in the context of reducing dependence on free parameters, both of these proposals are limited. For models with free parameters, it is often desirable that they meaningfully vary across conditions, and it seems generally the case that theory is insufficiently well-developed to fix parameters to single point values.

Conclusion

The Bayesian framework provides a complete and coherent solution to the basic scientific challenge of drawing inferences over structured models from sparse and noisy data. That is exactly what is needed to evaluate psychological models, including models of categorization. In this comment, we have discussed how Bayesian methods can address the four main evaluative issues raised by Wills and Pothos (2012). We have pointed to research that has already adopted Bayesian methods successfully, and indicated how the evaluation of psychological models will benefit even further from an expanded use of the Bayesian framework.

For one of the issues raised by Wills and Pothos (2012), the basic principles of Bayesian methods provide a solution. The correct application of Bayesian inference automatically controls for model complexity, and prevents over-fitting. But, for the other issues, the Bayesian approach does not in and of itself provide a complete solution. Instead, it can provide an appropriate framework for workable and principled solutions. Bayesian inference can operate naturally with decision theoretic utilities that can be used to emphasize the scientifically important qualitative properties of data. Bayesian inference requires the specification of a prior distribution over model parameters, and so can allow theoretical information about their content and constraints to be incorporated into evaluation. And Bayesian inference can work effectively with hierarchical models that can be applied to capture the way in which the same psychological variables and processes are manifest in multiple psychological phenomena, and so allows for direct evaluations of empirical breadth.

Of course, it will take some effort to realize all of the potential of the Bayesian approach, and there are mundane practical considerations. Some categorization models are computationally easier to implement within a fully Bayesian analysis than others. Foundational models like the GCM are relatively straightforward, but learning models like ALCOVE (Kruschke, 1992) or SUSTAIN are more challenging. These are computational rather than conceptual issues, and the successful Bayesian evaluation of closely related learning models, like the Expectancy Valence model (Wetzels, Vandekerckhove, Tuerlinckx, &

Wagenmakers, 2010), strongly suggests that they are surmountable.

Thus, we believe that the work that needs to be done to understand categorization and other cognitive activities is to develop good theories and models, and to collect the strong empirical evidence needed to evaluate those theories and models. The challenge is not to find a framework for evaluation. The Bayesian framework already exists, and is ready to do that part of the work.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer.

Bernado, J. M., & Smith, A. F. M. (2000). *Bayesian Theory*. Chichester, UK: Wiley.

Casella, G., & Berger, R. (2002). *Statistical Inference* (Second ed.). Pacific Grove: Duxbury.

Cox, R. T. (1961). *The Algebra of Probable Inference*. Baltimore, MD: Johns Hopkins University Press.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, *70*, 193–242.

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, *116*, 439–453.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis* (Second ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Griffiths, T. L., Canini, K. R., Sanborn, A. N., & Navarro, D. J. (2007). Unifying rational models of categorization via the hierarchical Dirichlet process. In D. J. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 323–328). Austin, TX: Cognitive Science Society.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science.* Cambridge, UK: Cambridge University Press.

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. In *Mathematical Proceedings of the Cambridge Philosophical Society* (Vol. 31, pp. 203–222). Cambidge University Press.

Jeffreys, H. (1961). *Theory of Probability.* Oxford, UK: Oxford University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 377–395.

Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, *91*, 1343–1370.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105*, 10687–10692.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 658–676.

Kruschke, J. K. (2011). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS.* Burlington, MA: Academic Press/Elsevier.

Lee, M. D. (2004). A Bayesian analysis of retention functions. *Journal of Mathematical Psychology*, *48*, 310–321.

Lee, M. D. (2006). A hierarchical Bayesian model of human decision-making on an optimal stopping problem. *Cognitive Science*, *30*, 555–580.

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, *15*, 1–15.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*, 1–7.

Lee, M. D., & Sarnecka, B. W. (2011). Number knower-levels in young children: Insights from a Bayesian model. *Cognition*, *120*, 391–402.

Lee, M. D., & Vanpaemel, W. (2008). Exemplars, prototypes, similarities and rules in category representation: An example of hierarchical Bayesian analysis. *Cognitive Science*, *32*, 1403–1424.

Lee, M. D., & Wagenmakers, E.-J. (2005). Bayesian statistical inference in psychology: Comment on Trafimow (2003). *Psychological Review*, *112*, 662–668.

Lee, M. D., & Wetzels, R. (2010). Individual differences in attention during category learning. In R. Catrambone & S. Ohlsson (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 387–392). Austin, TX: Cognitive Science Society.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–41.

Love, B. C., Medin, D. L., & Gureckis, T. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309–332.

Myung, I. J., Balasubramanian, V., & Pitt, M. A. (2000). Counting probability distributions: Differential geometry and model selection. *Proceedings of the National Academy of Sciences*, *97*, 11170–11175.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, *4*, 79–95.

Myung, J., & Pitt, M. (2009). Optimal experimental design for model discrimination. *Psychological Review*, *116*, 499–518.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104–114.

Nosofsky, R. M. (1998). Selective attention and the formation of linear decision boundaries: Reply to Maddox and Ashby (1998). *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 322–339.

Nosofsky, R. M., & Palmeri, T. J. (1997a). Aan exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.

Nosofsky, R. M., & Palmeri, T. J. (1997b). Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. *Perception, & Psychophysics*, *59*, 1027–1048.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 924-940.

Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, *113*, 57–83.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*, 421–425.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*, 472–491.

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.

Pooley, J. P., Lee, M. D., & Shankle, W. R. (2011). Understanding Alzheimer's using memory models and hierarchical Bayesian analysis. *Journal of Mathematical Psychology*, *55*, 47–56.

Popper, K. R. (1959). *The Logic of Scientific Discovery.* London: Hutchinson.

Robert, C. P. (1994). *The Bayesian Choice.* New York, NY: Springer-Verlag.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*, 573-604.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte,

M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, *105*, 5975–5975.

Shiffrin, R. M., Lee, M. D., Kim, W.-J., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*, 1248–1284.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 3–27.

Steyvers, M., Lee, M. D., & Wagenmakers, E.-J. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*, *53*, 168–179.

Tribus, M., & Fitts, G. (1968). The widget problem revisited. *IEEE Transactions on Systems Science and Cybernetics*, *4*, 241–248.

Vanpaemel, W. (2009a). BayesGCM: Software for Bayesian inference with the generalized context model. *Behavior Research Methods*, *41*, 1111–1120.

Vanpaemel, W. (2009b). Measuring model complexity with the prior predictive. *Advances in Neural Information Processing Systems*, *22*, 1919–1927.

Vanpaemel, W. (2010). Prior sensitivity in theory testing: An apologia for the Bayes factor. *Journal of Mathematical Psychology*, *54*, 491–498.

Vanpaemel, W. (2011). Constructing informative model priors using hierarchical methods. *Journal of Mathematical Psychology*, *55*, 106–117.

Vanpaemel, W., & Lee, M. D. (2007). A model of building representations for category learning. In D. J. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 1605–1610). Austin, TX: Cognitive Science Society.

Vanpaemel, W., & Lee, M. D. (submitted). Using priors to formalize theory: Optimal attention and the Generalized Context Model. *Manuscript submitted for publication.*

Vanpaemel, W., & Storms, G. (2008). In search of abstraction: The varying abstraction model of categorization. *Psychonomic Bulletin & Review*, *15*, 732–749.

Vanpaemel, W., & Storms, G. (2010). Abstraction and model selection in category learning. *Behavior Research Methods*, *42*, 421–437.

Voorspoels, W., Storms, G., & Vanpaemel, W. (2011). Representation at different levels in a conceptual hierarchy. *Acta Psychologica*, *138*, 11–18.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*, 1206–1220.

Wetzels, R., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E. (2010). Bayesian parameter estimation in the Expectancy Valence model of the Iowa gambling task. *Journal of Mathematical Psychology*, *54*, 14–27.

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

Zhang, S., & Lee, M. D. (2010). Optimal experimental design for a class of bandit problems. *Journal of Mathematical Psychology*, *54*, 499–508.

Footnotes

[1]We are not claiming that free parameters are always undesirable in psychological modeling. When psychological models are used as measurement models, for example, as in psychometric applications, estimating free parameters is useful and sensible. In these applications, psychological models allow latent psychological variables—like cognitive abilities in individuals—to be related to people's observed task behavior, and the scientific goal is to infer the free parameters, not eliminate them.