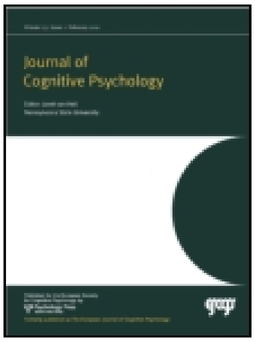

European Journal of Cognitive Psychology

ISSN: 0954-1446 (Print) 1464-0635 (Online) Journal homepage: http://www.tandfonline.com/loi/pecp20

What are the Bayesian constraints in the Bayesian reader? Reply to Norris and Kinoshita (2010)

Jeffrey S. Bowers

EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY
2010, 22 (8), 1270–1273

Psychology Press
Taylor & Francis Group

What are the Bayesian constraints in the Bayesian reader? Reply to Norris and Kinoshita (2010)

Jeffrey S. Bowers
Department of Psychology, University of Bristol, Bristol, UK

In this brief reply, I argue that the Bayesian reader can account for any pattern of data (including those not actually observed) because the predictions of the model are largely independent of any Bayesian principles. It is a good thing that the model is flexible, as the implemented model has been falsified by existing data.

Norris and Kinoshita (2008) claim that their Bayesian theory of priming provides a principled account for a wide range of results. On their view, the hypothesis that word identification is near optimal provides important constraints to the model, and this in turn allows it to explain *why* priming takes the form that it does. By contrast, previous "mechanistic" theories are said to succeed (to the extent that they do) by relying on a collection of ad hoc assumptions.

I challenged these claims in a recent critique, and also pointed out that their model makes a number of incorrect predictions (Bowers, 2010). In their response, Norris and Kinoshita (this issue 2010) focus on the importance of Bayesian principles in constraining theories, and did not fully respond to my claim that their model is falsified by existing data. In this brief response, I revisit both points.

HOW CONSTRAINING ARE THE BAYESIAN PRINCIPLES IN ACCOUNTING FOR MASKED PRIMING?

Norris and Kinoshita (2008) write "the Bayesian approach predicted all of these results from very simple basic principles" (p. 449). My claim, by contrast, is that most of the model's predictions are due to implementational decisions unrelated to Bayesian principles. In order to evaluate these

Correspondence should be addressed to Jeffrey S. Bowers, Department of Experimental Psychology, University of Bristol, 12a Priory Road Clifton Bristol BS8 1TU UK. E-mail: j.bowers@bris.ac.uk

http://www.psypress.com/ecp DOI: 10.1080/09541446.2010.532120

conflicting claims, consider the first principle noted by Norris and Kinoshita (this issue 2010), namely, that "words can be represented in terms of a multidimensional perceptual space" (p. 4). This amounts to the claim that the input to the model constitutes a concatenation of letter representations. However, the nature of the letter representation has varied across iterations of the Bayesian reader. Initially, Norris and Kinoshita (2008) employed the letter scheme from the Interactive Activation model, but because this prevented the model from accommodating key priming results, Norris, Kinoshita, and van Casteren (2010) adopted the coding scheme of the overlap model Gomez, Ratcliff, and Perea (2008). But this letter coding scheme also prevents the model from accounting for key data, and Norris et al. conclude, "a more complete model would focus on deriving a translation and scale-invariant representation of order" (p. 268), in line with the spatial coding scheme developed by Davis (2010). That is, the data determine which letter coding scheme is selected, not a "rational analysis" carried out on a multidimensional space.

This flexibility is widespread. For example, the original version of the Bayesian reader included "virtual nonwords", and the model assessed the relative probability that the input string was word or a virtual nonword. Norris added "background" nonwords to the calculations in order to eliminate a bias to respond YES in the lexical decision task. A problem with this implementation was that it predicted faster RTs to nonwords compared to words (contrary to fact), and in order to rectify this, Norris and Kinoshita (2008) eliminated the virtual and background nonwords and computed lexical decision in another way. Note, the new model has another problem regarding RTs in the lexical decision task (cf. Norris, 2009), and the authors will need to develop another hypothesis for making lexical decisions or alter some other aspect of the model to address this issue.

To take another example, consider how Norris and Kinoshita (2008) account for the robust nonword priming in the same-different task. They built in representations for the nonword reference stimuli (effectively making the nonwords words), and changed the vocabulary of the model:

> The same–different task is simulated in the same way as lexical decision. The only difference between the two is that, for the same–different task there is only one item in the "lexicon"—the reference stimulus . . . . (p. 444)

It is hard to see how the flexibility to vary the likelihood function (e.g., different letter coding schemes entail different likelihood functions), the hypothesis (e.g., different ways to make lexical decisions), or the priors (e.g., word frequency or age-of-acquisition) leads to models that are more constrained than alternative models.

HOW SUCCESSFUL IS THE BAYESIAN READER IN ACCOUNTING FOR EXISTING DATA?

Norris and Kinoshita (this issue 2010) did not respond to my claim that the implemented Bayesian reader is falsified by inhibitory masked form priming effects, but they did respond to my claim that the Bayesian reader cannot account for the parallel set of masked and unmasked nonword priming results. They write:

> Bowers believes that we have made claims about long-term unmasked priming in lexical decision, and that those claims are wrong. It's hard to see how claims that we never made could be wrong. The full extent of our comments on unmasked priming was that "A supraliminal word prime might have a genuine effect on the probability of encountering that word again in the experiment, so the prime should be expected to alter the prior of the target." On its own this makes no predictions, and in particular, we made no claims about the similarity or differences between short-term priming where the target follows the prime immediately, and long-term ("episodic") priming where many trials intervene between the "prime" and "target".

This is a surprising claim, given that the Norris and Kinoshita (2008) paper has a section entitled "Differences between Masked and Unmasked Priming" (p. 435), given that they motivate their analysis of masked priming with "[i]n light of these differences between the effects of masked and visual primes, there is clearly a need for an explanation of what it is that is special about masked priming" (p. 435), given their criticisms of alternative theories, for example, "[t]he first problem with these models is how to account for the difference between masked and unmasked primes" (p. 436), and given their explicit prediction that long-term priming should extend to nonwords when the prime is visible, "[i]n unmasked priming, where we assume that the prime alters the priors of the target, there should be priming for both words and nonwords" (p. 441).

To conclude, both Bayesian and non-Bayesian theories of priming have been influenced and changed in response to the data. Of course, data fitting should not be the only constraint for theory building: The ability to account for a diverse range of empirical findings in a parsimonious way, the ability to make novel predictions, adaptive constraints, etc., are all important. But Bayesian and non-Bayesian theories can be informed equally by these considerations, and as a consequence, both approaches can address *why* questions. A critical advantage of the "mechanistic" approach, however, is that algorithmic and implementational constraints come into play. It is the mechanism that addresses *how* questions.

Original manuscript received October 2010
Revised manuscript received October 2010

REFERENCES

Bowers, J. S. (2010). Does masked and unmasked priming reflect Bayesian inference as implemented in the Bayesian Reader? *European Journal of Cognitive Psychology*, *22*, 779–797.

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, *117*, 713–758.

Gomez, P., Ratcliff, R., & Perea, M. (2008). A model of letter position coding: The overlap model. *Psychological Review*, *115*, 577–601.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*, 327–357.

Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review*, *116*, 207–216.

Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General*, *137*(3), 434–455.

Norris, D., & Kinoshita, S. (2010). Explanation versus accommodation: Reply to Bowers (2010). *European Journal of Cognitive Psychology*, *22*, 1261–1269.

Norris, D., Kinoshita, S., & van Casteren, M. (2010). A stimulus sampling theory of letter identity and order. *Journal of Memory and Language*, *62*, 254–271.