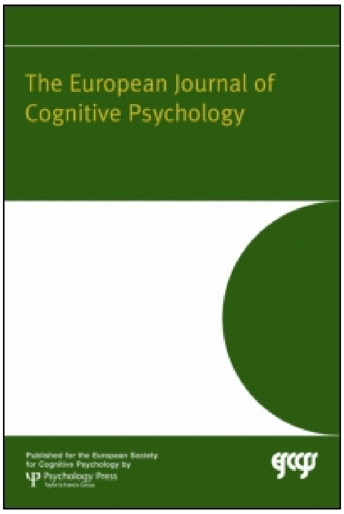

European Journal of Cognitive Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713734596

Does masked and unmasked priming reflect Bayesian inference as implemented in the Bayesian Reader?

Jeffrey S. Bowers [a]
[a] Department of Experimental Psychology, University of Bristol, Bristol, UK

First published on: 18 November 2009

To cite this Article Bowers, Jeffrey S.(2009) 'Does masked and unmasked priming reflect Bayesian inference as implemented in the Bayesian Reader?', European Journal of Cognitive Psychology,, First published on: 18 November 2009 (iFirst)

To link to this Article: DOI: 10.1080/09541440902955900
URL: http://dx.doi.org/10.1080/09541440902955900

EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY
0000, 00 (00), 1–19

Psychology Press
Taylor & Francis Group

Does masked and unmasked priming reflect Bayesian inference as implemented in the Bayesian Reader?

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

Norris and Kinoshita (2008) describe a Bayesian theory of masked and unmasked priming designed to explain a complex pattern of word and nonword priming across a range of tasks. Their theory is implemented within the Bayesian Reader model, and the model makes some predictions that are confirmed in a set of experiments. The authors consider alternative accounts of priming and conclude that only their theory can account for the results obtained. However, contrary to the authors' claims, the Bayesian Reader makes a number of incorrect predictions regarding masked and unmasked priming phenomena, whereas alternative theories can accommodate current findings.

Keywords: Bayesian reader; Word identification; Priming.

Masked and unmasked priming are commonly used in the study of visual word identification. In the case of masked priming, primes are flashed briefly (e.g., 50 ms) and followed with little or no delay by the target. Under these conditions participants are often unaware of the primes, but nevertheless, primes can impact on the processing of the target. Masked priming lasts on the order of seconds, and it is typically assumed to be an automatic byproduct of the online processing of the prime. In the case of unmasked priming, primes are typically presented for a second or more (they are visible), participants often make a response to the primes (e.g., they might be named or categorised), and targets are presented many seconds, minutes, or even days later. This long-term priming is often argued to be the automatic byproduct of learning processes (for a detailed comparison of masked and unmasked priming, see Bowers, 2003).

Although a great deal of theoretical work has been devoted to explaining masked and unmasked priming, Norris and Kinoshita (2008) describe a set of results that they consider problematic for all previous theories. At the

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: 10.1080/09541440902955900

same time, they argue that these findings are readily accommodated within the Bayesian Reader model of visual word identification. That is, they claim that masked priming can be explained as optimal decision making to a target when the prime and target are (mistakenly) treated as a single perceptual object, and unmasked priming reflects optimal learning in which the estimated prior probability of the prime (its prior) is altered as a consequence of encoding the prime in a previous study episode. Norris and Kinoshita argue that this model provides the best account of masked and unmasked priming to date.

In this critique I challenge many of these claims. On the one hand, I show how previous theories of masked and unmasked priming provide a plausible account of the critical data. On the other hand, I describe how the Bayesian Reader fails to account for some key phenomena that Norris and Kinoshita themselves highlight, as well as other relevant masked priming data. I conclude by considering whether the Bayesian Reader can be adapted to accommodate these data.

DATA THAT THEORIES OF PRIMING MUST EXPLAIN

Norris and Kinoshita (2008) outline a few key findings that, in combination, are thought to challenge previous accounts of masked priming. For present purposes, I highlight four of these findings. First, when masked priming is assessed with the lexical decision task, robust repetition priming is obtained for word targets (e.g., prime = *read*, target = *READ*), but little or no priming is obtained for nonword targets (e.g., prime = *blap*, target = *BLAP*; cf. Forster, 1998). Second, and in contrast with the lexical decision task, masked priming extends to nonwords in the same–different matching task (Norris & Kinoshita, 2008, Exp. 2). In this task, a reference item (a word or a nonword) is presented for 1 s in lower case letters, followed by a prime in lower case for ~ 50 ms (in a different spatial location), followed by the target in upper case. The participants simply judge whether the reference and target are the same, and priming reflects the impact of the prime on responses to the target. For example, given the nonword reference *cloor* and the target *CLOOR*, the SAME response is facilitated when the prime and target are the same compared to different (e.g., RTs are faster on the sequence: reference = *cloor*, prime = *cloor*, and target = *CLOOR* compared to the sequence: *cloor*-*deash*-*CLOOR*). Third, masked priming for words and nonwords is eliminated in the same–different matching task when the reference and target are different. For example, if the reference is *slump* and the target *ALARM*, there is no advantage in making a DIFFERENT response when the prime and target are the same (prime = *alarm*, target = *ALARM*) compared to different (prime = *thumb*, target = *ALARM*). Fourth, according to Norris and Kinoshita (2008; although see the discussion in the next section), *unmasked* (long-term) repetition priming extends to both word and nonword targets in the lexical decision task.

THREE STANDARD THEORIES OF MASKED PRIMING

According to Norris and Kinoshita (2008), these findings, in combination, are problematic for previous theories of priming. They organise these theories into three broad camps. On the first view, primes activate letter and word representations, and this provides a "head start" to targets in the repeated condition. Typically, it is assumed that the activated representations that support priming are lexical (e.g., Davis, 2003), although sublexical codes may contribute to masked priming as well (e.g., Humphreys, Besner, & Quinlan, 1988). On the second view, masked primes open (as opposed to activate) lexical representations, and this allows information from the target lexical entry to be extracted more quickly in the repeated condition (e.g., Forster & Davis, 1984).

These two types of theories can provide a straightforward account of why masked priming is restricted to words in the lexical decision task. For example, if priming reflects the activation or opening of preexisting lexical representations, then only preexisting words can be primed. The fact that long-term priming extends to nonwords is not generally considered problematic for these views, as it is typically assumed that long-term priming reflects episodic memory processes separate from the processes involved in identifying words (e.g., Forster & Davis, 1984; but see Bowers & Kouider, 2003). However, these theories appear to have trouble accounting for the robust nonword priming in the same–different matching task. That is, if priming reflects some process acting on a preexisting lexical representations (either activating or opening them), then priming should be restricted to words in this task as well. Furthermore, it is not clear why priming should be lost for words in the same–different matching task when the reference and target are different. That is, a prime should facilitate the identification of the word target in the repeated condition, and this in turn should speed up the decision that the target is different from the reference.

On the third view, masked priming reflects episodic memory (e.g., Masson & Bodner, 2003). The key assumption of this approach is that masked primes set down new episodic traces, and these traces facilitate the identification of targets. Because episodic traces extend to novel information, this approach has no difficulty accounting for the finding that masked priming extends to nonwords in the same–different matching task. Similarly, the long-term word and nonword priming in the lexical decision task can be explained given that long-term episodic traces are assumed to accommodate these effects.

However, Norris and Kinoshita (2008) argue that the episodic account has difficulty accounting for the null masked priming results obtained for words and nonwords in the same–different matching task when the reference and target stimuli are different. Similarly, it might be expected that masked priming would extend to nonwords in the lexical decision task. In both cases, episodic memory traces are laid down by the primes, and this should facilitate the processing of the target in the repeated condition.

More generally, Norris and Kinoshita (2008) assume that these three classes of theories have trouble accounting for the variable priming results observed across tasks and conditions. For example, the different results obtained for words and nonwords in the various conditions of the lexical decision and same–different matching tasks are taken to indicate that "Priming cannot be mediated by some form of automatic spreading activation between lexical representations" (pp. 448–449). Instead, the mixed results are taken to support the conclusion that "priming depends on the hypotheses that support the decision required to make a response" (p. 449). That is, the varied results are thought to support a Bayesian account, as different decisions require different hypotheses to be entertained.

A BAYESIAN ACCOUNT OF MASKED AND UNMASKED PRIMING

According to Norris and Kinoshita (2008), masked priming is the result of a few key computations in the Bayesian Reader. A core assumption of this model is that visual word recognition works according to Bayesian principles of decision making. That is, readers make optimal decisions about words and nonwords based on the available perceptual evidence and background knowledge (priors). For example, if a word is presented briefly such that its perceptual encoding is noisy, then the system can make an optimal guess about the identity of the word based on combining perceptual evidence with the words prior probability (e.g., its frequency). This is computed by Bayes' rule as given by Equation 1:

$$P(Word \mid Input) = P(W) \times P(I|W) \Big/ \sum_{i=0}^{i=n} (P(W_i) \times P(I|W_i)) \quad (1)$$

That is, the probability that a *specific* word has been presented to the model given the current input to the model, or P(Word | Input), equals the prior probability of that word given no evidence, P(W), operationalised as word frequency, multiplied by the likelihood of obtaining this input given that the specific word was in fact presented to the model, P(I|W), normalised by the sum of this product over the entire vocabulary (as calculated by the denominator of Equation 1). As a consequence, if the word *bright* is flashed quickly in a perceptual identification task, and the noisy perceptual input is equally consistent with *bright* and *blight* (that is, P(input | word) is the same for *bright* and *blight*), a person is likely to guess (correctly) *bright* given that the prior probability of *bright* (frequency count of 77/1,000,000 in CELEX) is higher than for *blight* (3/1,000,000). In essence, Bayes' rule provides an optimal bias in decision making under conditions of uncertainty.

Critically, Bayesian decision making in the model can be applied to all variety of hypotheses (not only the identity of specific words). So for instance, lexical decisions reflect optimal decisions about whether a given input was generated by a word (any word) compared to a nonword. That is, the model is computing the following:

$$P(AWord|Input) = P(AWord) \times P(I|AWord)/((P(AWord) \times P(I|AWord)) + (P(ANon-Word) \times P(I|ANon-Word))) \quad (2)$$

Unlike in Equation 1, the model does not need to identify a specific word above some probability, just whether the input is more likely to be a word (any word). The computations in this case are more complex, and the intuition of what the model is doing is more difficult to grasp (at least for me). The denominator sums over two hypotheses (that the input is a word vs. nonword) rather than over all the words in the vocabulary. In order to make computations regarding nonwords the authors assume "virtual nonwords" that are orthographically similar to real words, and in effect, the model is deciding whether the input is more similar to real words or virtual nonwords. See Norris (2006) for details on this. Although different computations are employed for the sake of word identification and lexical decisions, they are both based on lexical representations.

However, decision rules do not have to be directed at the lexical level at all. For instance, if participants are asked to make judgements regarding the visual forms of letters (e.g., judging that a/A have a different form, whereas a/a have the same form), then Bayesian decisions can be directed at a perceptual level in which the visual form of letters are computed as opposed to an orthographic level in which letters and words are coded in an abstract format (e.g., Bowers, Vigliocco, & Haan, 1998). Indeed, perception in general is thought to be characterised by Bayesian decision processes. For example, Norris and Kinoshita (2008) argue that this general analysis applies equally well to perceptual decisions about faces, shapes, and numbers.

The core claim regarding masked priming in the Bayesian Reader is that the prime and the target are (mistakenly) processed as one perceptual object. As a result, the encoding of the target is simply a continuation of the encoding of the prime. That is, the prime gives the target a "head start" when the prime and the target are related. The nature of the "head start" that contributes to priming depends on the hypothesis under consideration. If the task is to identify a specific word in a perceptual identification task, the evidence for the identity of the prime is integrated with the evidence of the identity of the target. This facilitates the identification of the target in the repeated condition effectively by altering (dynamically) the prior of the *target*. For instance, the prior probability of encountering the word *PUDDLE* by chance is quite low (e.g., 2/1,000,000 in CELEX norms), but its prior is revised upwards in the context of the prime *puddle* (in the same way, the prior of *PUDDLE* would be dynamically altered if it was embedded within a sentence about rain). The revised prior in turn reduces the processing required to reach a decision to the target *PUDDLE*.

If instead of attempting to identify a target participants are making lexical decisions, then evidence from the prime is again integrated with evidence of the target, but the evidence is now with regards to the lexical status of the target. That is, a decision is reached by computing the overall evidence in support of Equation 2. When the prime and target are the same word, the evidence from prime and target is consistent, and together they increase the probability of Equation 2, more so than when the prime and target are different words. This explains why masked repetition priming is obtained for words. By contrast, when the prime and target are the same nonword, the prime contributes to decreasing the probability of Equation 2, but no more so than when the prime and target are different nonwords. As a consequence, no priming is predicted for nonwords. That is, on this analysis, an optimal decision algorithm predicts priming for words but not for nonwords in the context of making lexical decisions (see Norris & Kinoshita, 2008, for details).

Norris and Kinoshita (2008) use a similar logic to explain the different pattern of priming observed in the same–different matching task. However, in this situation, the authors assume that the lexicon of the model is now one item—that is, the model's lexicon is simply composed of the reference stimulus (be it a word or nonword). The task of the model is to determine whether the input matches the reference P(I|produced by the reference), or whether it was more likely produced by a another input (a virtual nonword) one letter different from the reference P(I|produced by a neighbour of reference). When the reference and the target are different (that is, a NO response is required), the relation between the prime and target is irrelevant (as long as the prime and target are equally similar to the reference). This follows because the primes in the repeated and baseline conditions are both different from the reference, and thus both primes will equally contribute to a NO response (e.g., if the reference is *thice* and the target is *CLOOR*, the prime–target pairs *cloor*-*CLOOR* and *deash*-*CLOOR* equally contribute to the NO response). By contrast, when the reference and target are the same, then the relation between the prime and target does matter, with facilitation expected in the repeated condition. This follows because when the prime and the target match (*cloor*-*CLOOR*) they both contribute to the YES response, whereas when they mismatch (*deash*-*CLOOR*) the prime does not contributed to the YES response. This applies equally well to word and nonword targets, as the model encodes the reference as a familiar item (the only item in the lexicon), regardless of whether it is a word or nonword (that is, to the model, target nonwords are words).

As noted earlier, masked priming in the lexical decision and same–different matching tasks reflects the integration of evidence from the prime and target (due to the prime and target mistakenly being treated as a single perceptual object), and this dynamically alters the prior of whatever hypothesis is under consideration. By contrast, when the prime is unmasked, the prime and target are not mistaken as one perceptual object. Accordingly, the encoding of the prime does not dynamically alter the prior of the target. Rather the visual prime alters the prior of the *prime* in long-term memory, reflecting the fact that the item is more common (and thus more likely to occur in the future). This is effectively a change in its estimated frequency, which facilitates its later identification or categorisation in a lexical decision task.

Given that unmasked priming reflects a change of the prime's prior in long-term memory, the model can account for the longevity of this phenomenon. Furthermore, given that the prime acts to increase the priors of both words and nonwords, the model can explain why long-term priming extends to both words and nonwords in the lexical decision task.

TWO PROBLEMS WITH THE CLAIM THAT THE BAYESIAN READER PROVIDES A BETTER ACCOUNT OF MASKED AND UNMASKED PRIMING COMPARED TO PREVIOUS THEORIES

Problem 1: The authors have described a caricature of previous theories

A key claim of Norris and Kinoshita (2008) is that only the Bayesian model can account for the complex pattern of results obtained with masked and unmasked primes in the lexical decision and same–different matching tasks. However, this claim is based on a simplification of previous theories. Consider the most common theory of masked priming; namely, that the prime preactivates the target. There are a number of ways that this general approach could account for the complex pattern of results reported previously. For example, masked priming might reflect the preactivation of both sublexical (e.g., letters, onsets, bodies, etc.) and lexical representations (e.g., Bowers, Arguin, & Bub, 1996; Humphreys et al., 1988), with the sublexical codes supporting priming for nonwords in the same–different matching task. Furthermore, in order to account for the absence of nonword priming in the lexical decision task, a familiarity bias could be added to the theory (perhaps at a decision stage) that acts to slow down processing of nonword targets in the repeated condition. On this view, a masked nonword prime facilitates the identification of a repeated nonword target by preactivating the relevant sublexical representations. This should facilitate the (correct) NO response. On the other hand, a repeated nonword target might be perceived as more familiar (due to its improved perception), and this may serve as evidence that the target is a word (given that familiar letter strings tend to be words). This familiarity biases participants to respond (incorrectly) YES. It is the bias to respond YES to familiar targets that eliminates priming for nonwords in the lexical decision task. The hypothesis that familiarity acts to reduce or eliminate masked (as well as long-term) nonword priming in the lexical decision task has a long history (e.g., Bodner & Masson, 1997; Feustel, Shiffrin, & Salasoo, 1983; Humphreys, Evett, & Quinlan, 1990). I'll label this the *conflict hypothesis*.

Is this a post hoc account of the data? A key form of evidence taken to support the conflict hypothesis is that masked priming for nonwords is robust in all tasks other than the lexical decision task. For example, masked nonword priming is obtained in the naming task, and the effects are not simply the product of an articulatory onset effect (e.g., Bowers et al., 1996; Masson & Isaak, 1999). Similarly, masked nonword priming is obtained in perceptual identification tasks in which the prime and target are both presented briefly and masked (e.g., Evett & Humphreys, 1981). In these tasks, the improved perceptual encoding and increased familiarity of the nonword targets both contribute to the same (correct) response, and accordingly, nonword priming is expected. In the same way, the improved perception and increased familiarity of nonword targets in the same–different matching task in the SAME condition both contribute to the correct response, so again, the finding is predicted by the conflict hypothesis.

Furthermore, robust masked nonword priming is obtained in the lexical decision task under conditions designed to emphasise the role of improved perception over that of a familiarity bias, and inhibitory nonword priming effects are obtained under conditions designed to enhance the role of bias. For example, masked repetition priming is obtained for nonwords under conditions in which a high proportion of the trials include repeated primes and targets (Bodner & Masson, 2001). In this condition, most items are perceived as familiar, which may discourage a strategy of responding based on familiarity. For another manipulation that has the same effect, see Bodner and Masson (1997).

What about the fact that priming is eliminated for words and nonwords in the same–different matching task when the probe and target are different? Again the conflict hypothesis can be extended to explain these results (indeed, the hypothesis predicts the effect). Consider the situation in which the reference is *slump* and the target is *ALARM* (a DIFFERENT trial). If the prime is *alarm* (a repetition condition), participants should be faster to identify *ALARM* due to the prime improving the perception of the target, but the increased fluency of perceiving the target can be taken as evidence that the target has been repeated (which should inhibit a DIFFERENT response). That is, once again, a conflict between better perception and a familiarity bias may undermine priming. Indeed, similar accounts have been used in related contexts. For instance, Jacoby and Whitehouse (1989) found that participants were more likely to falsely categorise unstudied words as *old* in a recognition memory task when targets were preceded by a masked repetition prime. According to the authors, the improved fluency in processing the target (due to the prime) biases the participants to respond OLD (or in the current context, SAME). The role of familiarity (fluency) has been extensively studied in the memory literature, and again, it is possible to enhance or reduce the role of bias in order to increase or decrease memory errors (e.g., Goldinger & Hansen, 2005; Jacoby, Woloshyn, & Kelley, 1989; Whittlesea, Jacoby, & Girard, 1990)—just as bias can be systematically manipulated in order to increase or decrease masked priming in the lexical the decision task (e.g., Bodner & Masson, 1997).

Additional evidence for the conflict hypothesis comes from unmasked (long-term) priming studies. It is often assumed that unmasked primes strengthen preexisting orthographic representations or establish new representations, and these long-term modifications support priming (cf. Bowers & Kouider, 2003). That is, long-term priming is thought to reflect one-trial learning that effectively increases the frequency of the prime (much as assumed by Norris & Kinoshita, 2008). A large literature has demonstrated that long-term nonword priming is robust in most tasks, including naming (e.g., Brown & Carr, 1993) and perceptual identification (e.g., Bowers, 1994; Rueckl, 1990; Stark & McClelland, 2000) tasks. The one task that typically fails to support nonword priming (contrary to the claim of Norris & Kinoshita, 2008) is the lexical decision task (e.g., Bentin & Moscovitch, 1988; Bowers, 1994; Brown & Carr, 1993; Forbach, Stanners, & Hochhaus, 1974; Fowler, Napps, & Feldman, 1985; McKoon & Ratcliff, 1979), and the explanation is the same. That is, repetition facilitates the identification of the target as a nonword, but the increased familiarity of the repeated nonwords is taken as evidence that the target is a word (cf. Feustel et al., 1983).

Norris and Kinoshita (2008) only consider cases in which long-term nonword priming is obtained in the lexical decision task. However, it is important to note that all the studies they cite shared one key feature;

namely, participants performed lexical decisions at both study and test (Logan, 1990; Norris, 1984; Scarborough, Cortese, & Scarborough, 1977; Wagenmakers, Zeelenberg, Steyvers, Shiffrin, & Raaijmakers, 2004; Zeelenberg, Wagenmakers & Shiffrin, 2004). Under these conditions, priming may not only reflect the improved perception of the target, but also the strengthening of the mapping between the stimulus and response, which together facilitate the correct response more than a familiarly bias delays responding. When different tasks are performed at study and test, the nonword priming results are lost. Indeed, despite the fact that Norris and Kinoshita (2008) cite Wagenmakers et al. (2004) and Zeelenberg et al. (2004) as providing evidence that priming extends to nonwords in the lexical decision task, the key finding of these studies was that long-term nonword priming could be manipulated to be positive, null, or negative as a function of task conditions designed to modify the relative weighting of improved perception versus a familiarity bias in performing the task. Wagenmakers et al. and Zeelenberg et al. took their findings to support the conflict hypothesis, a conclusion that is inconsistent with the Norris and Kinoshita interpretation of their findings.

To summarise, a similar pattern of word and nonword priming is obtained in masked and unmasked priming tasks. The parallel results can be explained by assuming that a familiarity bias impacts on performance in both masked and long-term priming tasks, as described by the conflict hypothesis.

It should be emphasised that these effects can be explained in other ways as well. For instance, even in the absence of the conflict hypothesis, an activation account might explain the contrasting priming results obtained with nonwords in the lexical decision and same–different matching tasks. That is, masked priming may be lexically mediated, which would explain the null nonword priming typically obtained in the lexical decision task. By contrast, the robust nonword priming in the same-different matching task might reflect *new* lexical representations established during the encoding of the reference. This is not an unreasonable assumption given that each reference item is presented in the clear prior to the presentation of the prime. Indeed, the 1 s presentation of the reference would be expected to support long-term priming, and long-term priming may reflect one-trial learning (Bowers & Kouider, 2003). One-trial learning of the reference is exactly what Norris and Kinoshita (2008) assume in order to explain the masked priming results in the Bayesian Reader.

Alternatively, episodic theories may also account for this pattern of results. These theories assume that an episodic record of a masked prime improves the perception of a target as well as its familiarity (e.g., Masson & Bodner, 2003). So again, episodic theories appeal to the conflict hypothesis in order to account for the selective loss of masked (and unmasked) priming in the lexical decision task. For present purposes, the main point is not to endorse activation or episodic theories of priming. Rather, it is to show that Norris and Kinoshita (2008) are unjustified to claim that these data falsify previous theories.

Another criticism that Norris and Kinoshita (2008) raise is that previous theories cannot account for the fact that priming is sensitive to task demands. For example, when summarising their results, they write:

> First, and perhaps most surprisingly, priming does not depend on some fixed relationship between prime and target ... Priming cannot be mediated by some form of automatic spreading activation between lexical representations. (p. 448)

They take the varied results across tasks and conditions as support for their Bayesian account, as different decisions require different hypotheses to be entertained.

But it is unclear why they find this pattern of results problematic for previous theories. The large literature on masked and unmasked priming is replete with examples of different priming results in different tasks, and all previous theories assume that the nature of the task matters. For example, according to Forster and Davis (1991), masked priming in the naming task (but not lexical decision task) is influenced by the onset effect—that is priming is impacted by the partial naming of the prime. In a similar way, masked priming between homophones is more robust in the naming than lexical decision task, and this is attributed to the greater weight of phonology in naming compared to lexical decisions (Bowers et al., 1998). The nature of the nonword foils impacts on the nature of masked word priming in the lexical decision task, consistent with the claim that participants focus on different types of information in different conditions (e.g., Davis & Lupker, 2006; de Moor, Verguts, & Brysbaert, 2005; Grainger & Ferrand, 1996). Even on the assumption that masked primes automatically activate letters and words in a fixed manner (regardless of task), the decision processes must vary as a function of task, and it seems reasonable to assume that the benefits (or costs) in processing the target will be a function of the extent to which the preactivated representations overlap with the processes involved in responding to the target. This is consistent with the more general theoretical framework for model construction in the domain of visual word identification, as detailed by Grainger and Jacobs (1996). The authors highlight the importance of drawing a distinction between task-independent processes on the one hand, and task-specific, decision-related processes on the other. In their elaboration of the Interactive Activation (IA) model (the MROM model), different patterns of results across different word processing tasks are expected given that task specific decision processes contribute to performance under various conditions.

The same logic applies to long-term priming, where the nature of the priming depends on the nature of the task (cf. Bowers & Kouider, 2003). Indeed, this is the essence of transfer-appropriate processing long advocated in the memory literature, and there has been no corresponding commitment to the claim that long-term memory is optimal (cf. Schacter, 2001). Sensitivity to task instructions and Bayesian inference are orthogonal issues.

Problem 2: The Bayesian approach does not explain a number of relevant findings

Although previous theories can be extended to account for the previous findings, the Bayesian Reader, as currently implemented, has trouble with some of these and related findings. Let me highlight a few of these problems, some of which appear to pose a fundamental challenge to the model.

First, the Bayesian Reader model in its current form cannot account for the parallel set of masked and long-term nonword priming results. According to Norris and Kinoshita (2008), long-term priming is the byproduct of altering a prime's prior at study, which leads to the prediction of faster responses to repeated words and nonwords in the lexical decision task. By contrast, masked priming reflects a mistake in processing the prime and the target as a single object, and as detailed earlier, this leads to the prediction of null nonword priming in the lexical decision task. However, as detailed earlier, the striking finding is how similar the nonword priming effects are, with long-term and masked nonword priming typically lost in the lexical decision task, and robust in all other tasks. The parallel results may reflect the role of familiarity in modulating performance in various masked and long-term priming tasks, as described by the conflict hypothesis (e.g., Bodner & Masson, 2001; Wagenmakers et al., 2004).

Perhaps more importantly, the theory has difficulty accounting for masked form priming effects obtained with word targets. Norris and Kinoshita (2008) claim that masked form priming effects are facilitatory, but in fact, inhibitory effects are quite common. For example, Davis and Lupker (2006) report that whereas masked form priming is facilitatory when primes are nonwords and targets are words, inhibitory effects are obtained when the prime and target are both words, and the prime is higher in frequency. Furthermore, inhibitory from priming is increased when the prime and target share a neighbour (e.g., prime = *short*, target = *SNORT*, which share the neighbour *sport*, compared to a condition in which they do not share a neighbour, e.g., prime = *heard*, target = *BEARD*). Both of these finding are predicted by an activation based account of priming when implemented in the IA model (Davis & Lupker, 2006). These inhibitory effects follow from the competition that occurs in the model; for example, identifying *SNORT* is relatively difficult when primed with *short* because *short* proves to be a strong competitor to *snort* (and *short* is activated by both the prime and the target). Many other examples of inhibitory masked priming effects can be found (e.g., Bijeljac-Babic, Biardeau, & Grainger, 1997; Brysbaert, Lange, & van Wijnendaele, 2000; de Moor & Brysbaert, 2000; Drews & Zwitserlood, 1995; Grainger, Colé, & Segui, 1991; Grainger & Ferrand, 1994; Segui & Grainger, 1990).

The existence of inhibitory form priming effects is problematic for the Bayesian Reader, as the model can only predict facilitatory effects. To see this, recall that masked priming in the Bayesian Reader is the product of the prime and the target being treated as one object (the target). When the prime is similar to the target, the target's prior is increased (compared to a condition in which the target is preceded by an unrelated prime), which in turn facilitates responding to the target. In principle, a form-related word prime can reduce a target's prior compared to a condition in which no prime is presented. That is, if the prime is sufficiently well identified before the onset of the target, then the prime's prior would be increased at the expense of the target's. Nevertheless, a form-related prime cannot not reduce a target's prior more than an unrelated word prime (e.g., the prior for the target *SNORT* cannot be reduced more by the form related prime *short* compared to the unrelated prime *tramp*), and accordingly, there is no scope for the inhibitory effects reported in the literature. These findings are generally taken to reflect competition between coactivated and form similar lexical representations, a process that that is not captured in the computations of the Bayesian Reader. Similarly, the Bayesian Reader has no mechanism to explain the inhibitory long-term priming effects sometimes obtained with repeated nonwords in the lexical decision task (e.g., Bowers, 1994; Wagenmakers et al., 2004). These effects may reflect processes associated with the conflict hypothesis.

Another empirical difficulty for the model is that masked priming extends to nonwords in the perceptual identification task (e.g., Humphreys et al., 1988). According to Norris and Kinoshita (2008), nonword priming relies on preexisting lexical representations of nonwords. That is, nonword priming in the same–different matching task relies on one-trial learning of the reference nonword. Although one-trial learning of the nonwords is plausible in this context given that the nonword is presented in the clear for 1 s, there is no opportunity to learn the nonword primes in the masked perceptual identification task, and accordingly, some other explanation is required. One possibility is that the priming occurs sublexically (e.g., Bowers et al., 1996), or alternatively, the prime and target fuse under these specific test conditions, making nonword targets more legible in the repeated condition (cf. Davis & Forster, 1994). Whatever the explanation, nonword priming in the perceptual identification is problematic for the implemented Bayesian Reader.

CAN THE BAYESIAN READER BE MODIFIED IN ORDER TO ACCOUNT FOR MASKED AND LONG-TERM PRIMING?

As has been made clear, there are no empirical reasons to prefer a Bayesian Reader theory of priming. Still, it should be acknowledge that the Bayesian Reader is an implemented model, whereas some of the alternative accounts of masked and long-term priming (e.g., episodic theories, and the serial search model) are not. Furthermore, the implemented IA model that has been used to account for various masked priming data (e.g., Davis, 2003) also fails to explain some of the critical findings. For example, the implemented model cannot account for the parallel pattern of long-term and masked priming, as it does not include the learning mechanisms required to support long-term priming (although see Davis, 1999), nor a familiarity processes required to support the conflict hypothesis. Clearly, alternative models need to be implemented and modified in order to provide a more complete account of the data. A fair question, then, is whether the Bayesian Reader can be modified to account for the data.

It depends on what Norris and Kinoshita (2008) consider core to their theory. If the computations that support lexical decisions, and the computations involved in integrating information from the prime and target are considered fundamental claims, then it unclear whether the Bayesian Reader can be modified in order to account for the data. It is hard to see how inhibitory masked form priming effects can be accommodated without introducing new computations. On the other hand, it seems unlikely that the previous results pose a challenge to the Bayesian approach in general. In order to accommodate these findings, no doubt some optimality argument can be made for familiarity processes, as well as a role for competition between form similar words. So if the core claim is that priming is mediated by Bayesian decision processes, than the Bayesian Reader model can no doubt be adapted.

But the ability to develop different Bayesian theories of masked and long-term priming undermines one of the key arguments that Norris and Kinoshita (2008) put forward in support of the Bayesian Reader. That is, the authors claim that the Bayesian Reader is not only more successful in accounting for existing data, but is also more constrained to make the predictions that it does. For example, when considering the relative merits of the different theories, Norris and Kinoshita write: "Of course, it might well be possible to extend the scope of [previous] models and to modify them to be consistent with the data, but the Bayesian approach predicted all of these results from very simple basic principles" (p. 449). Or more generally, "This new interpretation allows us to situate masked priming within a much broader framework, and based on Bayesian principles, generate predictions for a much wider range of perceptual tasks" (p. 436). Elsewhere, Norris (2006) writes: "One of the most important features of the Bayesian Reader is that its behavior follows entirely from the requirement to make optimal decisions based on the available information. Given the initial assumptions, and a specification of the decision to be made, there is only one way the model can behave. The model has not been influenced or changed in any way by consideration of the data" (p. 351).

However, the last quote includes the critical qualification; namely, that the predictions of the model rely on "... initial assumptions, and a specification of the decision to be made" (Norris, 2006, p. 351). This undermines the claim that the Bayesian Reader is more principled or constrained in its predictions. For example, in order to explain the robust nonword priming results in the same–different matching task, Norris and Kinoshita (2008) assumed that the model learns new representations of the reference nonwords. Regardless of the merits of this hypothesis, the decision to treat word and nonword stimuli equivalently in this task was not based on any fundamental Bayesian principles.

Similarly, Norris and Kinoshita (2008) found that masked priming is insensitive to response congruence effects in a behavioural study (Exp. 1). That is, lexical decision times to target words were the same when preceded by unrelated *word* or unrelated *nonword* primes, and the model showed a similar insensitivity to the lexicality of the prime. However, if congruency effects had been obtained in behavioural experiments, it would not be difficult to accommodate the effects as well. In the Bayesian framework, primes can alter priors at the level of letters, words, or decision units (amongst other representations). The authors assume that masked priming in the lexical decision task reflects a revision of the priors at the lexical-orthographic level. However, if congruency effects had been obtained, it could have been argued that decision units play a role in priming, such that nonword primes increased the prior that the decision is NONWORD, which should delay responding when the target is in fact a word. That is, the success of accounting for the null congruency effect is not a principle of the Bayesian approach, but a decision to model the lexical decision task in a specific way.

Indeed, Bayesian models are not restricted to make optimal or adaptive decisions. For example, visual word identification is sensitive to Age-of-Acquisition (AoA) effects, with early acquired words identified more quickly than late acquired words after controlling for frequency effects (cf. Stadthagen-Gonzalez, Bowers, & Damian, 2004). This is nonoptimal (the probability of reading the word "dragon" in adult text is minimal), but as noted by Norris (2006), the findings can be explained by adding AoA information into the word priors. In short, unless there is some principled way of deciding the nature of the input to the model, what type of information is used to make a decision, what are the priors, etc., a Bayesian approach does not provide a principled account of priming (or anything else).

Of course, the same can be said of all theories. For example, models from the IA framework are free to add new processes in order to account for problematic priming results (e.g., it is possible that the decision process within IA models will need to be sensitive to familiarity in order to implement the conflict hypothesis), and IA type models are free to vary what types of information are used to support priming (e.g., lexical vs. sublexical), etc. But researchers advocating alternative frameworks have never claimed that their predictions were so strongly constrained, or that their theories have been uninfluenced by any consideration of the data.

In the end, the relative success of Bayesian compared to alternative theories will not rest on the fact that one approach is more principled than another. Rather, it will depend on how well the various theories accommodate the relevant data, and how well the theories predict new phenomena. Currently, the implemented Bayesian Reader is less successful than alternative approaches in accounting for existing data. Perhaps a future Bayesian model will provide the better theory, but for now, alternative approaches should not be dismissed so quickly.

REFERENCES

Bentin, S., & Moscovitch, M. (1988). The time course of repetition effects for words and unfamiliar faces. *Journal of Experimental Psychology: General*, *117*(2), 148–160.

Bijeljac-Babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory and Cognition*, *25*, 447–457.

Bodner, G. E., & Masson, M. E. J. (1997). Masked repetition priming of words and nonwords: Evidence for a nonlexical basis for priming. *Journal of Memory and Language*, *37*(2), 268–293.

Bodner, G. E., & Masson, M. E. J. (2001). Prime validity affects masked repetition priming: Evidence for an episodic resource account of priming. *Journal of Memory and Language*, *45*(4), 616–647.

Bowers, J. S. (1994). Does implicit memory extend to legal and illegal nonwords. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*(3), 534–549.

Bowers, J. S. (2003). An abstractionist account of masked and long-term word priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 39–56). Philadelphia: Psychology Press.

Bowers, J. S., Arguin, M., & Bub, D. N. (1996). Fast and specific access to orthographic knowledge in a case of letter-by-letter surface alexia. *Cognitive Neuropsychology*, *13*(4), 525–567.

Bowers, J. S., & Kouider, S. (2003). Developing theories of priming with an eye on function. In J. S. Bowers & C. J. Marsolek (Eds.), *Rethinking implicit memory* (pp. 19–40). London: Oxford University Press.

Bowers, J. S., Vigliocco, G., & Haan, R. (1998). Orthographic, phonological, and articulatory contributions to masked letter and word priming. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1705–1719.

Brown, J. S., & Carr, T. H. (1993). Limits on perceptual abstraction in reading: Asymmetric transfer between surface forms differing in typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1277–1296.

Brysbaert, M., Lange, M., & van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology*, *12*, 65–86.

Davis, C. J. (1999). The Self-Organising Lexical Acquisition and Recognition (SOLAR) model of visual word recognition. Unpublished doctoral dissertation, University of New South Wales, Sydney, Australia.

Davis, C. J. (2003). Factors underlying masked priming effects in competitive network models of visual word recognition. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 121–170). Philadelphia: Psychology Press.

Davis, C., & Forster, K. I. (1994). Masked orthographic priming: The effect of prime target legibility. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *47A*(3), 673–697.

Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 668–687.

De Moor, W., & Brysbeart, M. (2000). Neighborhood-frequency effects when primes and targets are of different lengths. *Psychological Research/Psychologische Forschung*, *63*, 159–162.

De Moor, W., Verguts, T., & Brysbeart, M. (2005). Testing the "multiple" in the multiple read-out model of visual word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 1502–1508.

Drews, E., & Zwitserlood, P. (1995). Morphological and orthographic similarity in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1098–1116.

Evett, L. J., & Humphreys, G. W. (1981). The use of abstract graphemic information in lexical access. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *33A(V)*, 325–350.

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General*, *112*(3), 309–346.

Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory and Cognition*, *2*, 337–339.

Forster, K. I. (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research*, *27*(2), 203–233.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*(4), 680–698.

Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language*, *30*, 1–25.

Fowler, C. A., Napps, S. E., & Feldman, L. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory and Cognition*, *13*(3), 241–255.

Goldinger, S. D., & Hansen, W. A. (2005). Recognition by the seat of your pants. *Psychological Science*, *16*, 525–529.

Grainger, J., Colé, P., & Segui, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory and Language*, *30*, 370–384.

Grainger, J., & Ferrand, L. (1994). Phonology and orthography in visual word recognition: Effects of masked homophone primes. *Journal of Memory and Language*, *33*, 218–233.

Grainger, J., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language*, *35*(5), 623–647.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518–565.

Humphreys, G. W., Besner, D., & Quinlan, P. T. (1988). Event perception and the word repetition effect. *Journal of Experimental Psychology: General*, *117*(1), 51–67.

Humphreys, G. W., Evett, L. J., & Quinlan, P. T. (1990). Orthographic processing in visual word identification. *Cognitive Psychology*, *22*(4), 517–560.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, *118*(2), 126–135.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*, *118*(2), 115–125.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, *22*, 1–35.

Masson, M. E. J., & Bodner, G. E. (2003). A retrospective view of masked priming: Toward a unified account of masked and long-term repetition priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 57–96). Philadelphia: Psychology Press.

Masson, M. E. J., & Isaak, M. I. (1999). Masked priming of words and nonwords in a naming task: Further evidence for a nonlexical basis for priming. *Memory and Cognition*, *27*(3), 399–412.

McKoon, G., & Ratcliff, R. (1979). Priming in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*, 463–480.

Norris, D. (1984). The effects of frequency, repetition and stimulus quality in visual word recognition. *Quarterly Journal of Experimental Psychology*, *36A*, 507–518.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*(2), 327–357.

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.

Rueckl, J. G. (1990). Similarity effects in word and pseudoword repetition priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*(3), 374–391.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 1–17.

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers*. Boston: Houghton Mifflin.

Segui, J., & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime–target frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 65–76.

Stadthagen-Gonzalez, H., Bowers, J. S., & Damian, M. F. (2004). Age of acquisition effects in visual word recognition: Evidence from expert vocabularies. *Cognition*, *93*, B11–B26.

Stark, C. E. L., & McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*(4), 945–972.

Wagenmakers, E.-J., Zeelenberg, R., Steyvers, M., Shiffrin, R., & Raaijmakers, J. G. W. (2004). Nonword repetition in lexical decision: Support for two opposing processes. *Quarterly Journal of Experimental Psychology*, *57A*, 1191–1210.

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, *29*(6), 716–732.

Zeelenberg, R., Wagenmakers, E. J., & Shiffrin, R. M. (2004). Nonword repetition priming in lexical decision reverses as a function of study task and speed stress. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*(1), 270–277.