Perception as evidence accumulation and Bayesian inference: Insights from masked priming.

Dennis Norris

MRC Cognition and Brain Sciences Unit

Cambridge, UK

and

Sachiko Kinoshita

Macquarie Centre for Cognitive Science

Macquarie University, Australia

Running head: Bayesian inference

Contact Information:

Dennis Norris
MRC Cognition and Brain Sciences Unit,
15 Chaucer Road, Cambridge
CB2 7EF U.K.
Tel: +44 1223 355 294
Fax: (+44) 11123 359 062
Email: dennis.norris@mrc-cbu.cam.ac.uk

Abstract

In this paper we argue that perception is Bayesian inference based on accumulation of noisy evidence and that, in masked priming, the perceptual system is tricked into treating the prime and target as a single object. Two algorithms are considered for formalizing how the evidence sampled from a prime and target is combined, and only one of the algorithms is shown to be consistent with the existing data from the visual word recognition literature. We then incorporate this algorithm into the Bayesian Reader model (Norris, 2006), and confirm its predictions in three experiments. The experiments show that the pattern of masked priming is not a fixed function of the relations between the prime and target, but can be changed radically by changing the task from lexical decision to a same-different judgment. We conclude with discussion of the implications of the Bayesian framework of masked priming for unconscious cognition and visual masking.

Perception as evidence accumulation and Bayesian inference: Insights from masked priming.

The perceptual system normally operates efficiently and reliably, but sometimes it fails. The perceptual system can be tricked into making mistakes. We can present it with unusual stimuli, such as those that cause visual illusions, or we can push it to its limits. The way perception breaks down can generate important insights into how perceptual processes normally function (c.f. Gregory, 2007; Weiss, Simoncelli, & Adelson, 2002). Here we suggest that one way to make a normally efficient process break down is to use the masked priming task. Masked priming tricks the perceptual system into treating the prime and the target as a single perceptual object, and the way the information from the prime and target is combined tells us something very important about perception: perception is Bayesian inference based on accumulation of noisy evidence. Of course, the idea that perception is Bayesian inference is not new (e.g. Knill, Kersten, & Yuille, 1996). But here we show the importance of backing up this very general claim with a specific proposal for the algorithm used to accumulate evidence. We incorporate this algorithm into the Bayesian Reader model (Norris, 2006) and show that this can both explain existing data and generate new predictions. Because the behavior of the Bayesian Reader is a function of the specific hypothesis that is being tested in performing the task, the model predicts that when the hypothesis is changed by altering task, this will change the pattern of priming. That is, priming is not a fixed property of the relationship between prime and target. For example, in the lexical decision task masked priming is usually limited to word targets. We will show that by switching to a same-different judgment task we can eliminate the priming for words and induce priming for nonwords. Furthermore, masked priming effects are usually insensitive to the perceptual overlap between the letters in the prime and target. However, by changing the decisions required in a same-different task we can make priming depend on the perceptual similarity of prime and target.

The structure of this paper is as follows: we begin with a brief review of the masked priming literature. Initially we will concentrate specifically on studies with word and nonword stimuli using the Forster and Davis (1984) paradigm; although in the discussion we will consider the implications of the theory for studies using stimuli other than words. We will then present a revised version of the Bayesian Reader model and describe two different algorithms that might be used to combine evidence from a prime and a target. Only one of these algorithms allows us to simulate the main phenomena in masked priming using the lexical decision task. Our account of masked priming involves only two additional assumptions beyond those already in the Bayesian Reader: masked primes and targets are treated as a single perceptual event, and task-dependent perceptual hypotheses are evaluated by integrating perceptual evidence. We report three experiments that test novel predictions of the model and conclude with a discussion of implications of the Bayesian interpretation of masked priming for other domains such as unconscious cognition and visual masking.

Masked priming.

When stimuli are presented very briefly, and are followed by a backward mask, they can influence the processing of subsequently presented stimuli, even though participants may have no phenomenological awareness of the masked stimulus (e.g., Dehaene, Naccache, Le Clec, Koechlin, Mueller, Dehaene-Lambertz, et al., 1998; Draine & Greenwald, 1998; Forster & Davis, 1984; Marcel, 1983). This phenomenon of masked priming has long been seen as a fascinating topic of research because it has the potential to shed light on the relationship between conscious and unconscious processing (e.g., Marcel, 1983). However, in recent years, work using the masked priming paradigm has diverged into quite distinct literatures that have very little contact. One line of research continues the tradition of pursuing the question of how masked primes are processed, and the implications this has for views of consciousness (e.g., Dehaene. Changeux, Naccahe, Sackur, & Sergent, 2006; Kouider & Dehaene, 2007; Kunde, Kiesel & Hoffmann, 2003). Another uses backward masking as a means of reducing the visibility of an object and focuses on the mechanism of visual masking for studying visual perception (e.g., Enns & Di Lollo, 2000; Enns & Oriet, 2007). In the third line of research, masked priming has come to be seen as a valuable tool for studying basic processes in visual word recognition (see, e.g., Kinoshita & Lupker, 2003). The appeal of masked priming as a method of studying word recognition is that it affords the possibility of investigating reading in a way that is much less likely to be influenced by strategic factors than tasks using supraliminal presentation: if participants are unaware of the identity of the prime, then they are unlikely to be able to consciously alter they way they process the prime (e.g., Forster, 1998; Forster, Mohan & Hector, 2003). The assumption is that by manipulating the relationship between prime and target, the resulting pattern of priming can provide insights into the nature of the representations underlying word recognition. For example, Forster (2004) suggests that masked priming is an “index of lexical access” (p.277).

In what has come to be the standard procedure used in visual word recognition studies, the Forster and Davis (1984) masked priming paradigm, the target stimulus itself acts as the backward mask. A trial consists of the presentation of a forward mask, usually a sequence of # symbols, followed by the prime, which is presented for about 50ms, followed immediately by the target. The participant's task is almost always to perform lexical decision on the target. The prime is generally presented in lower case and the target in upper case. At prime durations above about 60ms participants can usually discern that something occurred between the prime and the target, but below this most participants are unaware of the presence of any intervening event (Forster & Davis, 1984; Forster, Davis, Schoknecht & Carter, 1987).

When the prime and target are identical (i.e. same letters, but different case) this task produces reliable priming for word targets, but relatively little priming for nonword targets. Forster (1998) reviewed the results of 40 masked priming experiments and found that only three reported a nonword priming effect significant at the 5% level. The mean nonword priming effect over all 40 studies was 8.7ms. So, it does appear that there might be a genuine nonword priming effect, but it is very small, and always smaller than word priming. There is no indication that the pattern of results reported in the literature has changed since Forster's review. Priming is also obtained when the prime and target differ by a single letter (e.g. *antitude-ATTITUDE*) (e.g., Fosrter, 1987; Forster et al., 1987). This is usually referred to as form priming. Once again, the priming effect appears specific to word targets. Even in a study by Sereno (1991) that found significant identity priming for nonwords found no form priming. Studies that have reported form priming for nonwords have generally used non-standard procedures. For example, Masson and

Isaak (1999) used a naming task (in which it is possible to generate pronunciation by applying spelling-to-sound mapping rules), and Bodner and Masson (1997) presented targets in mIxEd case (hence making half of the letters of the target physically identical to the lowercase prime).

A natural interpretation of these results is that the masked priming effect is lexical (e.g., Forster, 1998; Forster, et al., 2003): Why else would it be restricted to items with lexical representations? However, here we will show that masked priming is not restricted to items with lexical representations at all. Masked priming is not 'lexical'.

Given that the target follows the prime immediately at the same location on the screen, one might expect that there might be some low-level visual priming at work here. However, in this particular variant of masked priming, there seems to be no evidence that this is the case. First, if priming operated at some very low level, both words and nonwords should be primed equally. Second, the prime and target are presented in different cases, which should minimize visual overlap. It seems that priming is mediated by abstract letter identities. This interpretation is supported by the results of a study by Bowers , Vigliocco and Haan (1998). Over a number of experiments they demonstrated that cross-case letter similarity (e.g., similar letters – c/C, x/X; dissimilar letters – a/A, b/B) does not modulate the size of identity priming effect: Priming effects are equally large for words consisting of cross-case similar letters (e.g., *kiss/KISS*) and words consisting of cross-case dissimilar letters (e.g., *edge/EDGE*). However, later we will show that, under suitable conditions, priming can depend on cross-case similarity.

Differences between masked and unmasked priming

The most notable difference between masked and unmasked priming is in the effect of identity priming of nonwords. As already noted, masked primes have very little effect on nonword targets. However, lexical decisions to nonword targets are facilitated by visible primes (e.g., Logan, 1990; Scarborough, Cortese, & Scarborough, 1977; Wagenmakers, Zeelenberg, Steyvers, Shiffrin, & Raaijmakers, 2004). Form primes also have a different effect depending on whether they are masked or visible. When the stimulus is a word and the prime is masked, form priming is facilitatory, however, when the prime is visible, there is no benefit from form priming (Colombo, 1986; Martin & Jensen, 1988). Most other differences between visible and masked primes are simply in terms of the magnitude of the effects. For example, semantic priming effects, which are highly robust with unmasked primes in lexical decision and naming tasks (see e.g., Neely, 1991, for a review) are small or unreliable with masked primes (e.g., Frost, Forster, & Deutsch, 1997; Perea & Gotor, 1997; Rastle, Davis, Marslen-Wilson & Tyler, 2000). In contrast, morphological effects are reliable and may even be larger with masked, than unmasked primes (e.g., Frost et al., 1997; Rastle et al., 2000).

In the light of these differences between the effects of masked and visible primes, there is clearly a need for an explanation of what it is that is special about masked priming.

Existing accounts of masked priming

Here, we briefly review how masked priming effects have been explained in the visual word recognition literature, namely, in terms of: (1) persisting activation (e.g., Davis, 2003; Grainger & Jacobs, 1996), (2) entry-opening (Forster, 1987; Forster et al., 2003), and (3) memory recruitment (Masson & Bodner, 2003).

Priming as activation

The most commonly held view of priming is that the prime, in some sense, activates the representation of the target, and that this activation then provides a "head-start" to the activation of the target. This idea is usually implemented in variations on McClelland and Rumelhart's (1981) Interactive Activation Model, such as the Multiple Read-Out Model (MROM, Grainger & Jacobs, 1996) and the Dual-route cascaded model (DRC, Coltheart, Rastle, Perry, Langdon & Ziegler, 2001). Simulations of masked priming in these models have been presented by Davis (2003), Davis and Lupker (2006), Jacobs and Grainger (1992) and van Heuven, Djikstra, Grainger, & Schreifers (2001).

The first problem that has to be faced by these models is how to account for the difference between masked and unmasked primes. Jacobs and Grainger (1992) propose that there is normally a 'reset' between successive stimuli (prime and target), but that the reset does not occur when two stimuli are presented in within certain time limits, as in masked priming. Davis and Lupker (2006), on the other hand, suggest that the mask/target causes a reset of the letter-level activations. There seems to be no principled reason to choose one procedure for implementing masked priming over another.

Another difficulty faced by the activation theories follows from the general problem these models have in simulating responses to nonwords. In the standard interactive activation model there is no mechanism for responding "nonword" in the lexical decision task. As a solution to this problem, the MROM and DRC models use a deadline procedure where the deadline can be modulated by overall activity in the lexicon. However, to date there are no published simulations masked priming of nonwords using interactive activation models. Furthermore, the viability of a deadline procedure for lexical decision has recently been questioned (see e.g., Wagenmakers, Ratcliff, Gomez, & McKoon, in press).

Priming as entry-opening

Forster and Davis' (1984) entry-opening account views masked priming as a "postaccess effect", and suggests that information can be extracted from the entry more rapidly when it has been primed. This fits well with the fact that in the lexical decision task nonword targets are not primed: Nonwords do not have a lexical entry that can be primed. The entry-opening account however has difficulties explaining the pattern of priming in other tasks. Forster (1985, Experiment 3) used a masked priming procedure in an episodic recognition task. Participants were asked to learn a set of 20 nonwords, and were then presented with a set of old and new nonwords that were preceded by masked primes in an old-new recognition task. Forster found significant identity priming for 'old' nonwords, but not for 'new'. This led him to argue that "priming occurs whenever the classification response is dependent on the contacting of any pre-existing mental representation" (p.97). In a further experiment using word stimuli he showed that there was priming for 'old' words but not 'new' words. These results are difficult to reconcile with the view that masked priming is a savings effect due to opening of lexical representation.

<u>Priming as memory recruitment</u>

The main tenet of the memory recruitment account (for a summary, see Masson & Bodner, 2003) is that both masked and unmasked "forms of priming have a common basis, namely, the recruitment of an episodic representation of the prime to assist with the identification of a target item" (p. 67). To make a case for the memory recruitment view, Masson and Bodner challenged the idea that there really are empirical dissociations between the pattern of priming observed with masked and unmasked primes. In particular, to support their claim that masked priming effects are not "lexical", Masson and Bodner cited two experiments by Bodner and Masson (1997) that did show masked priming for nonwords, one using MiXeD case targets and the other using pseudohomophone nonwords. We will discuss these findings later in more detail in the context of the Bayesian Reader. Here, we simply point out that the manipulations used could have changed the way participants performed the lexical decision task (consistent with this, decision latencies were much slower).

More recently, Bodner, Masson and Richard (2006) described the memory recruitment as being consistent with Anderson's rational analysis of memory (Anderson & Milson, 1989; Anderson & Schooler, 1991). In this analysis, memory retrieval is tuned to memories that are most likely to be relevant in the current situation, with relevance determined by the historical structure of the environment. Bodner et al. (2006) suggested that masked priming, similar to Anderson and Milson's (1989) account of semantic priming, is sensitive to the "need probability", i.e., the probability of needing the target word (e.g., cat) in the context of the prime word (e.g., dog). In view of the differences between masked and unmasked priming noted above, this theory faces some major obstacles: In particular, it does not explain why semantic priming in lexical decision and naming tasks, which is highly robust with supraliminal primes, is weak and typically absent with masked primes.

A Bayesian interpretation of priming

The Bayesian approach to perception leads to a very different explanation of masked priming according to which priming is driven primarily by the nature of the representations and decisions required by the task. An important consequence of this assumption is that priming is not directly determined by either the nature of the target (e.g. word vs. nonword), or even by the relationship between prime and target. The same pairing of prime and target that produces substantial priming in one task may produce none at all in another. 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.

In Bayesian terms, the most straightforward way to explain priming is to assume that a prime will change the priors of the target. That is, priming makes the target more or less predictable. After the sentence fragment "The cat sat on the ...", the word "mat" is far more probable than it is when it appears in isolation. This is the standard way of dealing with contextual information in, for example, automatic speech recognition systems. However, this view of priming implies that the prime and the target are discrete events. The prior of the target is changed as a result of identifying the prime. If masked priming did operate simply by changing priors in this way, one would expect masked priming to behave in exactly the same way as supraliminal priming, which is clearly not the case.

However, at least at a phenomenological level, the masked prime and the target are not seen as two discrete objects. Within the Bayesian framework, if the prime and target are processed as a single perceptual object, masked priming will behave very differently from supraliminal priming.

There are two algorithms that could be used to combine the information from the prime and target into a single percept. The first can be thought of as integrating at the level of the input data. The second can be thought of as integrating the evidence for particular hypotheses. To make this distinction concrete, and to show that only the second algorithm provides a viable account of masked priming effect, we will illustrate how both might be implemented in the Bayesian Reader. But first we give a brief account of the Bayesian Reader model itself.

The Bayesian Reader

As the name implies, the Bayesian Reader is a model of visual word recognition that assumes that readers operate as Bayesian decision makers. Readers make optimal decisions based on both the available perceptual evidence and on knowledge of word frequency (i.e. word probabilities). During normal reading the decision is about which word is in the input. But Bayes' theorem can also be used to make other decisions, such as whether the input is a word or not. An important consequence of this assumption is that behavior can change quite dramatically depending on the nature of the decision. For example, the Bayesian Reader correctly predicts that neighborhood effects should be inhibitory in perceptual identification tasks, but facilitatory in lexical decision (Andrews, 1997; Perea & Rosa, 2000.) As we will show later, it also predicts that the pattern of masked priming will vary as a function of task too. Priming effects are not some automatic consequence of lexical activation, but can come and go depending on the task.

Bayes' theorem is given in equation 1.

$$P(H_x \mid E) = P(H_x) \times P(E \mid H_x) \Big/ \sum_{i=0}^{i=n} (P(H_i) \times P(E \mid H_i)) \qquad (1)$$

Given knowledge of the prior probabilities with which events or hypotheses (H) occur, Bayes' theorem indicates how those probabilities should be revised in the light of new evidence (E). Given the prior probabilities of the possible hypotheses $P(H_i)$, and the likelihood that the evidence is consistent with each of those hypotheses $P(E|H_i)$, Bayes' theorem can be used to calculate $P(H_i|E)$, the revised, or posterior probability of each hypothesis, given the evidence. In the case of word recognition, each hypothesis corresponds to a word, and the prior probability of each hypothesis is given by the frequency of occurrence of the word. $P(H_i|E)$ is therefore the likelihood of observing the current perceptual evidence if the input really was $word_{i..}$ An important part of what follows will be the claim that the hypotheses that people evaluate are determined by the demands of the task. That is, they are not a fixed property of the perceptual system. So, for example, responses to word stimuli are not always driven by $P(Word_i|E)$,

In the Bayesian Reader, words are represented as vectors. Each word vector is constructed from a set of vectors representing the individual letters of the word, and each letter vector contains 26 elements. A 5-letter word is therefore represented by a 130 element vector. Perceptual evidence is accumulated by a noisy sampling process where each sample is constructed by adding zero-mean Gaussian noise to each element of the input vector. The process is easiest to appreciate by considering what would happen if the perceptual representation of words was unidimensional, with words differing just in terms of their location on this dimension. Successive samples would each consist of a single number. After each sample the model computes the mean and standard error of the mean of the samples. Clearly, as more samples are accumulated the sample mean will tend to move closer to the true mean (i.e. the input) and the standard error of the mean of the samples will decrease. The principles of the model are illustrated graphically in Figure 1 which represents the case where there are only two words in the lexicon. The lower two curves represent the likelihood functions, $f(X| W_i)$, of the two words at some early point in processing where the standard error or the mean is large. The likelihood functions indicate the likelihood of observing the input data X, given that the input was word W_i. As more samples are accumulated the standard error or the mean will decrease and the likelihood functions will become narrower, as indicated by the upper two curves. Over time the model will home in on the correct word (its likelihood will increase) while the likelihood of neighboring words will decrease.

Insert Figure 1 about here

Although Norris (2006) applied this procedure only to word recognition, the method is totally general and should apply equally well to recognition of any other objects or to any task requiring categorization where the necessary priors and likelihoods are available.

Lexical decision in the Bayesian Reader

The default model of the model is to calculate the probability of each word given the input, and given the assumption that the input is a word. This is clearly not the case in lexical decision, where half of the stimuli are nonwords. In lexical decision the task is not to decide which word is present, but to decide whether any word is present. To do this the Bayesian Reader performs lexical decision by comparing the overall evidence that the input was produced by a word, $P(X|\textit{input is a word})$, with the evidence that it was produced by a nonword ($P(X|\textit{input is a nonword})$). In order to make a 'Yes' response in lexical decision the model does not need to identify which particular word it is presented with. It just needs to know that the input is more likely to be a word that a nonword. Similarly, in order to make a 'No' response, the model does not need to have identified the letters in the nonword with any great confidence; it just needs to know that the input is unlikely to have been generated by a word. Full details of how the Bayesian Reader performs lexical decision are given in Appendix A. Having described the basic operation of the Bayesian Reader we now discuss two different algorithms that might be used to integrate the information in a masked-prime and target.

Two algorithms

1. *Integrating data*

In the existing formulation of the Bayesian Reader, the posterior probabilities of words are calculated after each input sample on the basis of the priors and the computed likelihoods. According to the algorithm implementing the current instantiation of the Bayesian Reader, this involves calculating the mean of the input samples. When processing a prime and target in a masked priming experiment the estimated mean of the input will be the average of the samples from both prime and target. If the representation of the prime and target are different, the resulting mean would correspond to a point somewhere between the two. Any priming effect would therefore be determined largely by the featural overlap between the letters in the prime and target and, contrary to the data, this would apply equally to words and nonwords. Perhaps it might be better to call this 'smearing' the data, rather than integrating the data. Given that primes and targets are generally presented in different cases, smearing would always occur in masked priming experiments.

However, there are far more general reasons to believe that the perceptual system is unlikely to operate simply by integrating data. Imagine that one is standing in the road and a red car is driving towards you at high speed. If one were to rely on integrating data, one’s visual percept would simply be of an ever lengthening red blur. The consequence is likely to be immediate removal from the gene pool. What one needs to do is to treat the changing data as support for the hypothesis that there is a car driving towards you and that this, in turn, supports the hypotheses that this situation demands rapid evasive action.

That is, what should be integrated over time is not the raw perceptual data, but evidence for alternative hypotheses.

2. *Integrating evidence*

Rather than integrating raw perceptual data, an alternative conceptualization of masked priming is that integration takes place by accumulating evidence for higher level perceptual hypotheses such a letters or words. A simple illustration of the importance of accumulating evidence for hypotheses comes from the integration of auditory and visual cues to speech recognition. Speech recognition is generally facilitated by being able to see the lip movements of the speaker (MacLeod & Summerfield, 1987; Middelweerd & Plomp, 1987). Additionally, as the well known McGurk effect (McGurk & MacDonald, 1976) shows, when there is a conflict between the visual information and the speech, this can lead to a percept that corresponds to neither. The critical point here is that integration must take place at a level above the raw data. It simply makes no sense to think of integrating auditory and visual information at the level of perceptual data. The only way to combine the two sources of evidence is in terms of their independent contribution to the likelihoods of alternative hypotheses about the identity of the incoming features, phonemes or words. The need to combine different sources of evidence in support of alternative perceptual hypotheses plays in indispensable role in perception in general. As the example of the red car shows, even within a single modality, evidence needs to be integrated at the level of perceptual hypotheses rather than data. The problem of integrating across successive saccades presents a similar problem (Irwin, 1996). However, rather interestingly, a recent review (Simons, Mitroff, & Franconeri, 2003) suggests that the problem is akin to overlaying two overhead transparencies. That is, they suggest that the problem is solved by integrating data.

The idea that perception requires the construction of perceptual hypotheses has parallels with the notion of ‘object-files’ (Kahneman, Treisman & Gibbs, 1992) in the visual perception literature. Object-files are able to maintain temporary episodic representations of objects in the visual field. In effect, objects are hypotheses. Object-based representations play a central role in the “object updating” theory developed by Enns and colleagues (Enns & Di Lollo, 1997; Di Lollo, Enns & Resnik, 2000; Enns, 2004; Lleras & Moore, 2003; Enns, Lleras & Moore, in press). They argue that a number of phenomena in visual masking can best be explained by the assumption that visual processing involves continuous updating of object representations. They make a distinction between “image updating” and “object updating” which parallels the contrast between integrating data and integrating evidence. We will say more about the object updating theory in the discussion.

It is important to note that both forms of integration are conceptually very different from the idea that the prime alters the predictability of the target. 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. At one level, the mathematical consequences of priming are the same regardless of whether the prime is supraliminal or subliminal: the prime alters priors. However, the revision of priors during processing of masked primes is quite independent of whether the prime alters the predictability of the target. Integration is a result of a failure to treat prime and target as discrete events. With unmasked primes, the fully identified prime changes the prior probability of the target. With masked primes, the priors are altered as a result of partial processing of the perceptual evidence for the prime. Under the circumstances of masked priming, the evidence from the prime thus leads to a revision of priors that has nothing to do with the probability of encountering the target, and does not provide an accurate basis for calculating posterior probabilities. In the terms used by Huber, Shiffrin, Lyle and Ruys (2001) and Weidemann, Huber and Shiffrin (2005), participants fail to discount the evidence from the prime: they fail to appreciate that the prime and target are different objects. Indeed, Huber et al. suggested that subliminal priming might be explained by their ROUSE (Responding Optimally to Unknown Sources of Evidence) model as a failure to discount the prime. The fact the ROUSE model also assumes optimal responding is a further source of similarity between their model and ours. Our claim is therefore that while word recognition is normally an efficient and near optimal process for performing decisions based on accumulating evidence from noisy input, the highly unusual circumstances of masked priming cause the system to break down and perform sub-optimally. The perceptual system is tricked into processing the prime and target as a single object.

Integrating evidence in the Bayesian Reader

The original formulation of the Bayesian Reader was designed to deal with the case where decisions (perceptual identification, or lexical decision) were based on a sequence of samples from a static input. The specific algorithm used to perform the necessary computations was selected because it had the most transparent relation to the standard formulation of Bayes' theorem. The algorithm works by computing the mean of the data samples accumulated so far, and then using this to compute probabilities. Because it involves computing the mean of all of the samples received so far, this algorithm integrates at the level of data rather than hypotheses.

However, there are alternative algorithms that can compute exactly the same function by continuously updating probabilities or likelihood ratios, without the need to compute the sample mean. A standard way of combining multiple sources of evidence is by multiplying likelihood ratios (see Appendix B). This is the procedure one would use to combine the evidence from a number of independent diagnostic tests for a particular disease. The results of the tests are converted to likelihoods ratios, and these are then multiplied together and used to compute the probability that the patient has the disease. Each noisy sample can be considered to be analogous to the outcome of a diagnostic test. This procedure formed the basis of early models of reaction time developed by Stone (1960) and Laming (1968), and also Ratcliff's (1978) diffusion model. At each time step evidence is accumulated by multiplying likelihood ratios from successive samples. The critical factor here is that formally it does not matter where the evidence comes from. It could come from samples from a single input, or could equally well come from quite difference sources. So, for example, during the processing of a prime, evidence might accumulate for the presence of the lower case letter 'a'. While processing the target, evidence might accumulate for the upper-case letter 'A'. However, both could be combined to provide support for the hypothesis that the input contains the case-independent abstract letter representation a/A. Alternatively, both might provide evidence for words containing the letter a/A in the appropriate position. Multiplying likelihoods is equivalent to using each sample to update the probability of each hypothesis, and using these new probabilities as the priors for the next update. This is the framework we will adopt here to describe the process of evidence accumulation.

In masked priming, the evidence from the prime can be used to revise the priors for words or letters, and evidence from the target will simply continue that process. Because the raw perceptual input samples are discarded after they are used to revise priors, the fact that the mean location of the input samples might change between prime and target need be of no consequence. Evidence from the prime and target letters is simply making independent contributions to the calculation of posterior probabilities of words or letters. What is being carried over from moment to moment are probabilities, and not a literal record of the perceptual input. Note that because priors are undergoing continuous revision, there is no need for any discontinuity in processing at the transition between prime and target. Although, it is quite likely that there might be some low-level visual disruption at the transition between prime and target, the system does not need to know that there are separate primes and targets, or take any special action at the end of the prime or onset of the target.

The main question that this analysis poses is: Which priors are being revised? The simple answer is that it depends on the decision required to perform the task. One could imagine priors being revised at the level of letters, words, or even at the level of the final decision. However, in lexical decision the response should be driven entirely by the pooled evidence for individual words, so this is the only level at which evidence should be integrated. Given the evidence for individual words, combined with an estimate of the likelihood that the input is a nonword, the probability that the input is a word follows automatically. That is, after each sample the probability that the input is a word must be computed afresh directly from word probabilities. The response probability computed after the previous sample should not be allowed to influence the response probability at the current sample.

The principle that masked priming in lexical decision should be driven solely by integrating evidence at the word level is also supported by the data. If evidence were integrated at the level of letters, this would benefit words and nonwords equally. This is clearly not the case. It is of interest to note that as an exception to this, as mentioned earlier, Bodner and Masson (1997) did report finding priming for nonwords in a lexical decision task. One of these experiments used mixed-case targets, and the other used pseudohomophone nonwords. A natural consequence of both of these manipulations is that responses become much slower. This raises the interesting possibility that the pattern of priming might change if the task were changed from that of performing lexical decision in the optimal way specified by the Bayesian Reader, to one of first identifying the letters, and then deciding whether the target is a word or not. If this were the case, priming should operate at the letter level, and this would benefit words and nonwords equally.

The implication of this analysis is that in a task where word probabilities were not relevant to the decision, priming would not be determined by lexical status. The level at which priors are revised is not some fixed property of the perceptual system, but should be determined by the hypotheses that are being compared in performing the decision required by the task. We will soon see that this leads to an important novel prediction.

Modeling masked priming in lexical decision

Masked priming is simulated by presenting the prime for a number of samples corresponding to the duration of the prime (10 samples in all of the simulations reported here). At this point the priors of each word are revised in the light of the evidence accumulated so far. Processing of the target word then proceeds exactly as it would had there not been a prime, with the exception that the words now have new priors. In the simulations all probabilities are calculated as though there are discrete phases corresponding to the prime and target. However, as described earlier, this is for computational convenience only, and we assume that the process of updating probabilities is continuous. That is, the computer program running the model uses an algorithm which, under the specific circumstances of these simulations, computes the same function as an algorithm that continuously updates probabilities.

First consider what happens in the case of identity priming with word targets. The prime will cause the prior of the target to be revised upwards. The target will therefore be responded to faster. To a slightly lesser extent, the same will be true of form primes because they are similar to the target. They will also increase the probability of the corresponding word target, and this will lead to a faster decision.

However, in the case of an identity primed nonword, there is no representation of the nonword that can have its prior revised. Any influence of a nonword prime must be mediated by its effect on words. In fact the only factor that influences processing of nonwords is how similar they are to words. The easiest way to understand what happens in the model is to consider a simplified case of a lexicon with only one word, and where the experiment uses two nonwords located equally far from the word in perceptual space.

One nonword corresponds to the target nonword (and identity prime) the other is an unrelated control prime. For clarity, this is illustrated in Figure 2. When the identity prime is presented, the probability of the word will go down. The likelihood of the word, given the input from the nonword is shown in the solid bar at the mean of the target nonword distribution. At the end of the prime there is no record of the individual samples. All that has happened is that the prime has changed the probability of the word. But, presenting the control nonword prime would have had exactly the same effect because the likelihood of the word at the end of the prime would have been identical. There is no record of the perceptual form of the prime, only of how far away it was from the word. Consequently, it does not matter whether the nonwords presented as the prime or the target are the same or different. All that counts is how different they are from the word. At this point it might seem as though this explanation is similar to the idea that priming depends on lexical representations. However, the representations need not be lexical. What counts are the representations required to perform the task. It just happens that lexical decision depends on lexical representations.

Insert Figures 2 & 3 about here

A simulation of masked priming for words and nonwords is shown in Figure 3. This simulation uses the same stimuli that will be used in Experiments 1 and 2. In both the simulations and in Experiment 1 there are two kinds of unrelated control prime.

Congruent primes have the same lexical status as the target, whereas incongruent primes have the opposite lexical status. To the extent that there is any response priming, either in the data or the simulations, incongruent trials would be expected to produce slower responses than congruent trials. As would be expected from the description of the model given above, there is a facilitatory priming effect for words for identity primes, but no facilitation for nonwords. There is no specific representation that accumulates evidence that the input is a nonword, therefore there is no facilitatory priming for nonwords. There is a general tendency for identity priming to have an inhibitory effect on nonwords. A nonword prime will initially increase the probability of words in its neighborhood and these words will therefore compete with the nonword more strongly that if they had not been primed. The simulations here assume that participants perform the lexical decision task optimally and that priming changes only word priors. As noted earlier, to the extent that participants adopt a suboptimal strategy, such as a letter-by-letter spelling check, there might also be some facilitation of letter identification which should be modeled as a change in letter priors. This would benefit words and nonwords equally. Note that, in common with the simulations reported in Norris (2006) responses to nonwords are faster than to words, the opposite of what is normally observed. However, whereas the pattern of priming necessarily follows from the structure of the model, the relative speed of word and nonword responses can be altered by changes in parameters. When the model is supplemented with parameters to represent additional sources of noise in processing, it can give precise quantitative simulations of both word and nonword latencies and response-time distributions (Norris, submitted).

The fact that the model correctly predicts that masked priming should be greater for words than nonwords helps illustrate the difference between priming via evidence integration and priming by changing priors. In unmasked priming, where we assume that the prime alters the priors of the target, there should be priming for both words and nonwords (e.g. Norris, 1984, Zeelenberg, Wagenmakers & Shiffrin, 2004), as even nonwords should become represented in memory, and can therefore be assigned priors. When participants see a nonword it is more likely that they will encounter that specific nonword again than some random nonword (i.e., its “need probability” will be increased, cf. Anderson & Milson, 1989; Anderson & Schooler, 1991). In fact, this will be true of any encounter with a novel letter string in everyday life. A novel letter string is likely to be a new word or acronym. Therefore, a single encounter with a novel letter string will increase the probability of encountering that letter string in the future.

Response congruence

Theories of masked priming proposed in the unconscious cognition domain will be reviewed later in the paper. A view common to many of these theories is that, in some sense, participants are unconsciously applying the task instructions to the prime as well as the target. For example, Dehaene, Naccache, Le Clec, Koechlin, Mueller, Dehaene-Lambertz, et al. (1998) proposed that “Subjects would unconsciously apply task instructions to the prime”. Similar accounts have been proposed by Klinger, Burton and Pitts (2000), Kunde, Kiesel and Hoffmann (2003), and Forster (2004). Although this idea is broadly consistent with the Bayesian view, in the absence of any account of how the task is performed, is it impossible to make clear predictions about the outcome of masked priming experiments. For example, if participants apply task instructions to the prime in a lexical decision task, one might predict that priming would depend mainly on whether the prime and target were to be classified in the same way. That is, nonword primes should facilitate only responses to nonword targets, and word primes should facilitate only word targets. Priming should depend on whether responses to primes and targets are congruent. By “response”, we mean the category required by the task (i.e., “categorize the letter string as a word or a nonword”) and not the motor response (e.g., “press the left/right key”). Indeed, Dehaene et al.’s (1998) proposal is based on the finding of such congruence effects in a categorization (“Is the number bigger/smaller than 5?”) task, in which masked primes that are response-congruent produced faster responses to the target relative to response-incongruent primes (e.g., 3-1 is faster than 7-1). However, as shown by the simulations in Figure 2, the Bayesian Reader predicts that there should be no response congruence effects in lexical decision. The reason for this is quite simple: the prime increases the evidence for words in some quite broad area of perceptual space, but does not focus in on an area sufficiently small to alter the probability of a word versus a nonword response. Words of a particular form are primed, but there is no effect of response congruence. But are the predictions of the Bayesian Reader correct? We are aware of only one published study that directly tested (and did not find) response congruence effect in lexical decision: That result (Perea, Fernandez & Rosa, 1998) is only available in Spanish. In the first experiment we test this prediction in a lexical decision experiment that manipulated the lexical status of both targets and primes. This experiment also serves to establish whether stimuli to be used in Experiment 2 show the standard effects reported in the literature. That is, do they show priming for words, but not nonwords?

Experiment 1

The purpose of Experiment 1 was to test if the lexical decision task is sensitive to response congruence between the prime and target. It also served to establish that the stimuli show the benchmark effects observed with the lexical decision task, namely, the frequency effect, and identity priming effect with words, but not with nonwords.

Method

Participants. Twenty-four volunteer Macquarie University students participated in Experiment 1 for course credit. All participants were native Australian-English speakers.

Design. Experiment 1 used the lexical decision task. The experiment constituted a 3 (Target type: high-frequency words vs. low-frequency words vs. nonwords) x 3 (Prime type: identity vs. response-congruent vs. response-incongruent) factorial design, with both factors manipulated within subjects. The dependent variables were lexical decision latency and error rate.

Materials. The critical stimuli were 80 high-frequency words, 80 low-frequency words and 80 nonwords, which were also used in Experiment 2. All items were 5-letters long. The high-frequency words ranged between 81 to 1599 occurrences per million (Kucera & Francis, 1967), with a mean of 316.5. The low-frequency words ranged between 1 to 20 occurrences per million, with a mean of 9.2. The high-frequency words and low-frequency words ranged in the number of orthographic neighbors (N, as defined by Coltheart, Davelaar, Jonasson, & Besner, 1977) between 0 and 17, and were matched on mean N (4.11 and 4.36, respectively). Nonwords were generated by changing one or two letters in a real word and were approximately matched with the words on N. For the nonwords, N ranged between 0 and 10 with a mean of 3.41. The stimuli are listed in Appendix C.

Sixty words from each stimulus class were used as targets in the lexical decision task, and the remaining 20 were used as (response-congruent and response-incongruent) control primes. To equate the number of word and nonword trials there were 60 nonword filler targets and 20 control primes.

The 60 high-frequency words, 60 low-frequency words, and 60 nonwords from the critical set as well as the 60 nonwords from the filler target set were each divided into three sets containing 20 items, Sets A, B, and C. Three list versions were constructed such that each lists contained 60 high-frequency words and 60 low-frequency words and 120 nonwords as targets. Within each list, one third of stimuli of each type were paired with an identity prime, one third with a response-congruent prime, and the remaining one third, a response-incongruent prime. Lexical status and frequency of the response-congruent prime was matched to that of the target (e.g. *range-CHILD*, *abbey-THIGH*, *deash-CLOOR*); the response-incongruent prime had the opposite lexical status to the target (e.g., *deash-CHILD, range-CLOOR*). A control prime stimulus was used once as a response-congruent prime and once as a response-incongruent prime, but it was never used as a target. The assignment of the three sets to the prime condition was counterbalanced so that within a list a target occurred only once, and across three lists, each target appeared in each of the three prime conditions just once. Eight participants were assigned to each of three lists.

Participants were given 16 practice items and there were 4 warm-up items at the start of each test block. These items were selected according to the same criteria as the test stimuli and were not included in the analysis.

Apparatus and Procedure. Participants were tested individually, seated approximately 40 cm in front of a Dell 19 inch Flat Trinitron monitor, upon which stimuli were presented. Each participant completed 240 test trials, presented in two half blocks, each block consisting of 30 high-frequency word, 30 low-frequency word and 60 nonword targets, with a self-paced break between the blocks. A different random order of trials was generated for each participant.

Participants were instructed at the outset of the experiment that they would be presented with a series of letter strings in uppercase, and their task was to decide for each letter string whether it was a word or a nonword, as fast and accurately as possible. Responses were collected via an external response pad. Participants were instructed to press a key marked “+” for words and a key marked “-“ for nonwords. No mention was made of the presence of primes.

Stimulus presentation and data collection were achieved through the use of the DMDX presentation software (Forster & Forster, 2003) running on a Dell Dimension 8400 computer running on an Intel Pentium 4 processor. Stimulus display was synchronized to the screen refresh rate (13.3 ms).

Each trial began with the presentation of a forward mask containing a row of five hash signs (#####) for 500 ms, which was replaced by a prime in lowercase letters presented for 53 ms. The prime was then replaced by a target displayed in uppercase letters, which remained on the screen either until the participant made a response or for 2,000 ms. Both primes and targets were presented in the centre of the screen in white Courier 10 point font against a black background. Participants were given no feedback on either response times or error rates during the experiment.

Results

In summary, the lexical decision latency data showed the expected effects. For words, there was a large effect of frequency and identity priming, which was greater for low-frequency words than high-frequency words (contrary to results of earlier studies, e.g., Forster & Davis, 1984; but consistent with studies that used low-frequency words that are known to participants, see Kinoshita, 2006). There was no identity priming for nonwords. As predicted by the Bayesian Reader, there was no response congruence effect, either for words or nonwords.

Data analysis

In this and all subsequent experiments, the preliminary treatment of trials was as follows. Any trial on which a subject made an error was excluded from the RT analysis. To reduce the effects of extremely long and short latencies, the cutoff was set for each participant at 3 S.D. units from each participant’s mean latency and those shorter or longer than the cutoff was replaced with the cutoff value. In Experiment 1, this affected 1.6% of trials. Lexical decision latencies and error rates of the critical word and nonword target set (60 high-frequency words, 60 low-frequency words and 60 nonwords) are reported, with words and nonwords analyzed separately. Words were analyzed using a two-way analysis of variance (ANOVA) with Frequency (high-frequency words vs. low-frequency words) and Prime type (identity vs. response-congruent vs. response-incongruent) as factors. Nonwords were analyzed with just the Prime type as a factor. In the by-subjects analysis, both were within-subject factors, and in the by-items analysis, Frequency was a between-item factor. In line with the recommendations of Raaijmakers, Schrijnemakers, and Gremmen (1999) and Raaijmakers, (2003), we only report analyses treating subjects as a random factor. Effects were considered to be significant when the by-subjects analysis was significant at the .05 level. In every case where the by-subjects analysis was significant, the items analysis was also significant. Mean lexical decision latencies and error rates are presented in Table 1.

Insert Table 1 about here

Words. For latency, there were significant effects of both Frequency, $F(1,23) = 78.19$, $MSe = 1571.15$, and Prime type $F(2,46) = 18.88$, $MSe = 1905.91$. Orthogonal contrasts showed that there was no difference between response-congruent and response-incongruent prime conditions, $F < 1.0$, but the average of these prime condition was significantly slower than the identity prime condition, $F(1,23) = 36.40$. There was an interaction between Frequency and Prime type, $F(2, 46) = 5.62$, $MSe = 540.32$. Orthogonal interaction contrasts showed that the response-congruence effect did not interact with Frequency, $F(1,23) = 1.74$, $p = .20$, but the identity priming effect was greater for low-frequency words than for high-frequency words, $F(1,23) =11.33$.

For error rate, there was a significant effect of Frequency, $F(1,23) = 27.02$, $MSe = 86.47$. The main effect of Prime type was non-significant, $F < 1.0$. There was no interaction between the two factors, $F(2, 46) = 1.24$, $MSe = 43.83$.

Nonwords. There were no significant effects in either the latencies or errors.

Discussion

The results of Experiment 1 provide a straightforward confirmation of the predictions of the Bayesian Reader: there are no response congruence effects in masked priming using the lexical decision task. However, it is worth noting at this point that this does not mean that the Bayesian approach would never predict what appears to be a response congruence effect. If the task is simply to determine which of two symmetrical hypotheses is correct (i.e. there is a specific representation for each hypothesis. e.g. a left arrow or a right arrow: Vorberg, Mattler, Heinecke, Schmidt, & Schwarzbach, 2003) then the response will be driven directly by the evidence from the identity of the target and prime. If the prime generates evidence for one response (hypothesis), and the target for the opposite response, then this will slow responding relative to either a neutral prime or an identity prime.

How to prime nonwords

Selective priming of words rather than nonwords in lexical decision comes about because nonwords have no pre-existing representation whose prior can be revised. But, what would happen if the task did require a specific representation of a nonword? If the precise form of the nonword were relevant to the decision, might it be possible to show masked priming of nonwords? More interesting still, if the form of the word did not matter, might it be possible to eliminate word priming?

A task which seems to satisfy the necessary requirements is the sequential same-different matching task (e.g., Bruder, 1978; Marmurek, 1989; for a review, see, e.g., Proctor, 1981). If participants are required to compare two sequentially presented nonwords (a reference, and a target) to determine whether they are the same or different, this decision must be made on the basis of the letters in the nonwords. However, if the two nonwords are different, then the form of the reference nonword matters, but the exact form of the target does not. The only important thing is to establish that the target is different from the reference. The situation is directly analogous to the case of nonwords in lexical decision, but now it is only when the two nonwords are different that the form of the second nonword does not matter. When they are the same, the form does matter. The underlying assumption here is the same as with lexical decision. Participants have a specific hypothesis about what to expect if the target is the same as the reference, but no specific hypothesis about what form a different target would take. The same-different decision is made by comparing the likelihood that the target has the same form as the reference with the likelihood that it is different. Consequently, we should be able to see masked priming of nonword targets when reference and target are the same, but not when they are different.

If this reasoning is correct, it should apply to words as well as nonwords. That is, there should be masked priming for 'same' decisions, but no word priming for 'different’ decisions. The predicted pattern of priming is independent of the lexical status and frequency of the stimuli.

Insert Figure 3 about here

The simulation of masked priming in the same-different task is shown in Figure 3. 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, and this is compared to the nearest letter-string. Words other than the reference play no role in these simulations. The reference stimulus is always set to have a prior probability of 0.5 ('same' and 'different' trials are equally likely), and there is always a representation of the form of the reference, regardless of whether or not it is a word. The task is to determine whether the input corresponds to the reference, or a letter-string 1 letter different from the reference. As in the case of identification and lexical decision, the probability of the reference string is determined by the product of the probabilities of the constituent letters. At the end of the prime the only prior to be revised is P(same).

So, moving from one different nonword to another, assuming that they are both roughly the same perceptual distance from the reference, will have no influence on P(same). When the prime and targets are identical on a 'same' trial, the prime will contribute to P(same), i.e. the probability that the input is the same as the reference string.

The corresponding control prime will reduce the probability of the reference string, and cause inhibition. This reasoning applies equally to words and nonwords. There will be priming for 'same' trials, and no priming for 'different' trials. Because only the reference string is being considered there should be no effect of word frequency. We now present an experimental test of the model's predictions, using the same stimuli used in the lexical decision task in Experiment 1. The main prediction is that whereas in lexical decision there was priming for words but not nonwords, in the same-different task there will be priming for 'same' responses, but not for 'different' responses; further, in the same-different task priming effects will not be influenced by lexical status.

Experiment 2

Method

Participants. An additional twenty-four participants from the same subject population as Experiment 1 participated in Experiment 2.

Design. Experiment 2 used the same-different matching task. The experiment constituted a 3 (Target type: high-frequency words vs. low-frequency words vs. nonwords) x 2 (Prime type: identity vs. control) factorial design, with both factors manipulated within subjects. These factors were crossed with the Response (*Same* vs. *Different*) factor. The dependent variables were decision latency and error rate.

Materials. The critical stimuli were the same 80 high-frequency words, 80 low-frequency words and 80 nonwords used in Experiment 1; these were used as targets in the same-different task. An additional 40 high-frequency words, 40 low-frequency words and 40 nonwords were selected to be used as control primes, as well, 40 high-frequency words, 40 low-frequency words and 40 nonwords were used as the reference item requiring a *Different* response. They were all 5-letter long. Mean frequency for the high-frequency reference words was 200.5 per million, for the low-frequency reference words, 10.0. The reference stimuli are listed in the Appendix C.

The 80 high-frequency words, 80 low-frequency words and 80 nonwords were each divided into four sets containing 20 items each. Four list versions were constructed such that each list contained the same 240 items as targets, and across the four lists, each set appeared in each of the four experimental conditions resulting from a factorial combination of Prime type (identity vs. control) and Response type (Same vs. Different). Lexical status and frequency of the control prime and *Different* reference stimuli were matched to those of the target (e.g., *block-range-CHILD, crust-abbey-THIGH, thice-deash-CLOOR*). Orthographic overlap between the reference, prime and the target in the control and *Different* conditions was avoided as much as possible. Six participants were assigned to each of the four lists. Participants were given 16 practice items and there were 4 warm-up items at the start of each test block. These items were selected according to the same criteria as the test stimuli and were not included in the analysis.

Apparatus and Procedure. The apparatus and software used for stimulus presentation and response collection were identical to Experiment 1. Participants were instructed at the outset of the experiment that they would be presented with a pair of letter strings, one after another, and their task was to decide for each pair whether the second string (target) was the same or different from the first string (reference), ignoring difference in case, as fast and accurately as possible. Participants were instructed to press a key marked “+” for *Same* and a key marked “-“ for *Different* responses. As in Experiment 1, no mention was made of the presence of primes. Each participant completed 240 test trials, presented as two blocks of 120 trials each, with a self-paced break between blocks. In each block, half required a *Same* response and half required a *Different* response, and the three types of targets (high-frequency word, low-frequency word, nonword) were represented equally. A different random order of trials was generated for each participant.

Each trial began with the presentation of a reference stimulus in lowercase letters above a forward mask consisting of five hash signs for 1 second. Both the reference and the forward mask then disappeared, and the forward mask was replaced by a prime in lowercase letters presented for 53 ms, which was then replaced by a target presented in uppercase letters. The target remained on the screen either until the participant made a response or for 2,000 ms.

Results and Discussion

The results of Experiment 2 are as predicted. There is priming for ‘same’ targets, but not for ‘different’ targets. This is true for both words and nonwords. In contrast to lexical decision, the same-different task produces priming for nonwords (in the ‘same’ condition) combined with an absence of priming for words (in the ‘different’ condition). Furthermore, lexical effects in Experiment 2 were minimal. Words were responded to only 12ms faster than nonwords, which contrast with a lexicality effect of 96ms in Experiment 1. There was no significant effect of word frequency.

The preliminary treatment of data was identical to Experiment 1. In Experiment 2, the data-trimming procedure affected 1.5% of trials. *Same* and *Different* decision latencies and error rates were analyzed separately, using a two-way ANOVA with Target type (high-frequency words vs. low-frequency words vs. nonwords) and Prime type (identity vs. control) as factors. Mean decision latencies and error rates are presented in Table 2.

Insert Table 2 about here

Same responses. In the analysis of latency, the main effect of Target type was significant, $F(2, 46) = 5.23$, $MSe = 647.36$. Orthogonal contrasts showed that Words were significantly (by 12 ms) faster than Nonwords, $F(1, 23) = 6.58$, $MSe = 1155.96$. The frequency effect did not reach significance, $F(1, 23) = 3.24$, $MSe = 1048.16$, $p = .09$. The main effect of Prime type was highly significant, $F(1, 23) = 235.27$, $MSe = 1236.75$. On average, identity-primed targets were 90 ms faster than control-primed targets. There was no interaction between Target type and Prime type, $F < 1.0$.

In the analysis of error rate, only the main effect of Prime type was significant, $F(1, 23) = 12.13$, $MSe = 143.01$. Identity-primed targets were on average 7% more accurate than control-primed targets. All other effects were non-significant, $F < 1.0$ in all cases.

Different responses. In the analysis of latency, none of the main effects or interactions reached significance, $F < 2.70$, $p > .08$.

In the analysis of error rate also, none of the effects reached significance, $F_1 < 2.15$, $p > .16$.

The results of Experiment 2 have a parallel in the results observed in an episodic recognition task by Forster (1985) described earlier. Whether the studied items were nonwords (Experiment 3) or words (Experiment 4), only the “old” items showed priming; “new” items showed no priming. Forster’s results are directly analogous to the same-different results from present Experiment 2, with the ‘old’ items being effectively ‘same’ items that were presented some minutes earlier. Forster’s results are clearly exactly what would be expected from the Bayesian analysis of masked priming[1].

Experiments 1 and 2 show that masked priming is critically dependent on the nature of the task, or hypothesis. When the hypothesis being tested is “the target is a word”, then words are primed but not nonwords. When the hypothesis is “the target is the same as the reference”, ‘same’ responses are primed regardless of lexical status, but ‘different’ responses are not. Might changes in task change other aspects of masked priming? One issue raised in the introduction was whether masked priming was based on abstract letter identities or on more basic visual properties of the stimuli. All of the evidence suggests that priming depends on abstract letter identities. For example, the study by Bowers, Vigliocco and Haan (1998) showed that priming was not modulated by the visual similarity of the primes and targets. However, might this also be dependent on the nature of the task? In lexical decision, or in our same-different task, the physical form of the targets is irrelevant to the decision. All that matters are abstract letter identities. But what if the physical form of the stimulus did matter to the decision being made? In the next experiment we investigate whether, by changing the task, we can induce a shift between reliance on abstract letter identities and reliance on visual form.

Experiment 3 (cross-case and same-case letter match)

Experiment 3 used a letter match task. The task is essentially the same as the same-different task used in Experiment 2, except here the stimuli are all single letters. Participants are instructed to decide as quickly as possible if a target letter is the same letter as the reference letter presented in advance. As in the same-different task used in Experiment 2, the target is preceded by a briefly presented prime which is forward-masked by a # sign, and backward-masked by the target. Experiment 3 used two versions of this task, cross-case match and same-case match. In the cross-case match task, the reference and target are always in opposite case (e.g., a/A, a/B) and participants are required to decide if they are the same letters, ignoring difference in case. In the same-case match task, participants are instructed to respond *Same* only if they are the same letters in the same case (e.g., a/a, A/A), and the reference and targets in the *Different* condition included the same letters in different case (e.g., a/A) as well as different letters in same case (e.g., c/x). The prime-target pairs (which were always in opposite case) were unchanged in the two tasks.

Kinoshita and Kaplan (2008) used the cross-case letter match task and found robust identity priming effects. As in the present Experiment 2, priming was limited to *Same* responses; *Different* responses did not show sensitivity to the identity between the prime and target (e.g., a-b-B = a-g-B). In all of their three experiments, the size of priming did not differ between cross-case similar letters (e.g., c/C, x/X) and cross-case dissimilar letters (e.g., a/A, b/B), and this was so whether the prime was in different case (e.g., a-A), as in the standard masked priming procedure, or in the same case (but in different size, e.g., A-A) as the target. These results were taken as evidence that priming in the cross-case letter match task was based on abstract letter identities. We expected the present cross-case letter match task to replicate these results. The novel prediction concerned the same-case match task. Based on the argument presented above, when the decision is about case-specific identity, priming should be modulated by cross-letter similarity, and cross-case similar letters (e.g., c/C), but not dissimilar letters (e.g., a/A) are expected to show priming.

Method

Participants. Thirty-two volunteer Macquarie University students participated in the experiment for course credit; Twelve were assigned to the cross-case match task, and twenty to the same-case match task, in the order of arrival.

Design. Each task constituted a 2 (Letter type: cross-case dissimilar vs. similar) x 2 (Prime type: identity vs. control) factorial design, with both factors manipulated within subjects. The dependent variables were decision latency and error rate.

Materials. The critical stimuli were 6 cross-case dissimilar letters (A/a, B/b, D/d, E/e, G/g, R/r) and 6 similar letters (C/c, O/o, P/p, S/s, V/v, X/x), selected on the basis of Boles and Clifford’s (1989) letter similarity ratings. We avoided the letters i/I, l/L and q/Q, because of their confusability to other letters (e.g., l and I, q and g in the Courier font, the font used to present the letters). Each trial involved a reference, prime and target (e.g., a-a-A), and participants were required to decide if the target was the same or different from the reference. The prime and target were always in opposite case in both the cross-case and same-case match task. In the cross-case match task, the reference and target were always in opposite case, and participants were asked to decide if the reference and target were the same, ignoring the difference in case. For the *Different* reference-target pairs and the control prime-target pairs, the letters were chosen from the same letter similarity class (Dissimilar letters or Similar letters). In the same-case match task, the reference and target were always in the same case, and participants were required to respond *Same* only if the reference and target were the same letter in both identity and case. For the *Different* condition, the reference-target pairs for the Dissimilar letters were changed so that they involved the same letter identity differing only in case (e.g., reference-A, target-a). This was done to encourage participants to focus on the case-specific letter identities. (This was done only for Dissimilar letters because Similar letters in different case (e.g., c/C, x/X) differ only in relative size, and it is was considered unlikely that size information survives masking). Examples of reference-prime-target triplets in each of the 8 experimental conditions in each task are shown in Table 3.

Insert Table 3 about here

In each of the cross-case and same-case match task, presenting a letter in uppercase once and lowercase once as a target in each of the 8 experimental conditions resulted in 96 test trials (12 x 8 experimental conditions) as a complete set. Participants were given 12 practice trials and there were 4 initial buffer trials at the start of each block. These trials used letters not used as the critical stimuli and were not included in the analysis.

Apparatus and Procedure. The apparatus and software used for stimulus presentation and response collection were identical to Experiment 1. Each participant completed 192 test trials, comprised of two blocks with each block containing 96 trials. A different random order of trials was generated for each participant.

Participants were instructed at the outset of the experiment that they would be presented with a pair of letters, one above the other, and their task was to decide whether they were the same or different letters, as fast and accurately as possible. In the cross-case match task, they were told to ignore the difference in case; in the same-case match task, they were told to respond *Same* only if the letters were in the same letter in the same case. Responses were collected via an external response pad. Participants were instructed to press a key marked “+” for *Same* and a key marked “-“ for *Different*. No mention was made of the presence of primes.

Each trial began with the presentation of a reference letter, presented above a forward mask, a hash sign (#), for 750 ms. It then disappeared, and at the same time, the forward mask was replaced by a prime letter presented for 53 ms. The prime was then replaced by a target letter, which remained on the screen either until the participant made a response or for 2,000 ms. The reference, prime and target were presented in the centre of the screen in white Courier 14 point font against a black background. Participants were given no feedback on either response times or error rates during the experiment.

Results

As predicted, there were significant priming effects *Same* responses for both *cross-case* and *same-case* matches, and the effect of letter similarity was significant only for *same-case* matches.

The preliminary treatment of trials was as identical to previous experiments. In the cross-case match task, the 3 S.D. cutoff replacement procedure affected 1.4% of trials; in the same-case match task, it affected 1.2% of trials. For each task, we report a two-way ANOVA with Letter type (dissimilar vs. similar) and Prime type (identity vs. control) as factors, for *Same* and *Different* responses.

Cross-case match task. Mean decision latencies and error rates for the cross-case match task are presented in Table 4.

Insert Table 4 about here

Same responses. For latency, the main effect of Prime type was highly significant, $F(1, 11) = 112.09$, $MSe = 493.81$. Targets preceded by identity-primes were responded to 68 ms faster than the control-primed targets. The main effect of Letter type was non-significant, $F(1, 11) = 3.38$, $MSe = 948.52$. There was no interaction between Prime type and Letter type, $F < 1.0$.

For error rate, the main effect of Prime type was significant, $F(1, 11) = 4.65$, $MSe = 89.75$. Targets preceded by identity-primes were 5.9% more accurate than the control-primed targets. The was no significant effect of Letter type, $F(1,11) = 1.01$, $MSe = 52.08$, and no interaction between Prime type and Letter type, $F < 1.0$.

Different responses. For latency, the main effect of Prime type was non-significant, $F(1, 11) = 1.81$, $MSe = 916.43$. The main effect of Letter type was significant, $F(1, 11) = 19.18$, $MSe = 734.07$, with Dissimilar letters being responded to 34 ms more slowly than Similar letters. This effect of Letter type for Different responses was not expected and suggests that the reference-target pairs were more similar for the Dissimilar letters than the Similar letters. To check this possibility, we analyzed Boles and Clifford's (1989) letter similarity rating data. Mean similarity rating for dissimilar letter pairs used in the present experiment was 267.33 and for similar letter pairs 195.08 (on a scale of 100-500), a statistically significant difference, $F(1, 23) = 17.89$, $MSe = 1751.07$. Thus, this finding seems to be an artifact of the Different reference-target letter pair combinations being more similar (and therefore more difficult) for the dissimilar letters than the similar letters. We should note that this confound is difficult to avoid with the current set of letters used, because the average similarity to all other letters in the set is higher for the Dissimilar letters than the Similar letters (270.66 vs. 238.00). There was a marginal interaction between Prime type and Letter type, $F(1, 11) = 3.29$, $MSe = 1149.52$, $p = .09$. Identity primes tended to speed up *Different* response to Dissimilar letters, but had little effect for Similar letters.

For error rate, none of the effects was significant, all $F < 1.64$, $p > .21$.

Same-case match task. Mean decision latencies and error rates are presented in Table 5.

Insert Table 5 about here

Same responses. For latency, the main effect of Prime type was significant, $F(1, 19) = 13.62$, $MSe = 958.34$. The main effect of Letter type was significant, $F(1, 19) = 8.49$, $MSe = 806.24$. Of critical interest, there was a significant interaction between Prime type and Letter type, $F(1, 19) = 6.99$, $MSe = 891.37$. Simple effects analysis showed that the priming effect for Dissimilar letters (8 ms) was non-significant, $F < 1.0$, but the effect for Similar letters (43 ms) was significant, $F(1, 19) = 27.30$, $MSe = 683.72$.

For error rate, none of the main effects or interactions was significant, all $F < 3.15$, $p > .08$.

Different responses. For latency, the main effect of Letter type was significant, $F(1, 19) = 42.63$, $MSe = 1105.87$. Dissimilar letters were responded to 48 ms more slowly than Similar letters. None of the other main or interaction effect was significant, all $F < 1.0$.

For error rate, the main effect of Prime type was significant, $F(1, 19) = 7.08$, $MSe = 42.55$. This effect reflected greater error rates to identity-primed letters (9.2%) than control-primed letters (5.6%). The main effect of Letter type was also significant, $F(1, 19) = 14.51$, $MSe = 103.69$. Consistent with the latency data, Dissimilar letters (11.4%) were more error-prone than Similar letters (3.4%). These two factors did not interact, $F < 1.0$.

Discussion

Experiment 3 demonstrated that the nature of masked identity priming between letters in different case can be modulated in a letter match task. Using prime-target letter pairs that are cross-case similar (e.g., c/C, x/X) and dissimilar (e.g., a/A, b/B), when decision was required at the level of abstract letter identity (responding *Same* to a-A or c-C), the size of identity priming effect was unaffected by cross-case letter similarity, replicating the result observed by Kinoshita and Kaplan (2008). In contrast, when decision was at the level of case-specific letter identity (responding *SAME* to a-a or C-C but *Different* to a-A), the priming effect was eliminated for Dissimilar letters but remained for Similar letters. So, not only is the pattern of priming with words and nonwords task dependent but, as predicted, so is the pattern of letter priming (abstract or case-specific).

General Discussion

The data reported here support the Bayesian analysis of masked priming presented in the introduction. We proposed that masked priming is driven by the failure of the perceptual system to treat the prime and target as perceptually distinct events. Evidence from the prime and the target is continuously integrated to alter the probabilities or likelihoods of the representations required to perform the task. This leads to the prediction that the pattern of priming will necessarily change as a function of the nature of the task. This was confirmed by showing that the standard pattern of priming seen in the lexical decision task: priming for words but not nonwords, is not observed with 'same' trials in a same-different task. Furthermore, as also predicted, neither words nor nonwords showed priming for 'different' trials.

We then showed that analogous task-dependent changes in priming can be seen at the levels of physical versus abstract letter identity. In a cross-case letter match task priming was unaffected by cross-case letter similarity, but a similarity effect did modulate priming when the task required participants to match letters on the basis of their physical identity.

These data make a number of important points. First, and perhaps most surprisingly, priming does not depend on some fixed relationship between prime and target. The relationships between primes and targets were the same in the lexical decision and the same-different experiments, and in the cross-case and same-case letter match tasks: they used the same stimuli paired in the same way. However, the pattern of priming was completely different. Priming cannot be mediated by some form of automatic spreading activation between lexical representations. Second, priming is not a direct consequence of the nature of the primes and targets. In particular, priming does not depend on the targets being words. There is no priming for 'different' words in the same-different task, but there is priming for 'same' nonwords. Third, although priming in lexical decision and letter-name same-different tasks seems to depend only on abstract letter identities, priming can be made dependent on the physical form of letters when the task requires a judgment about physical identity. Once again, priming is not dependent on any fixed relation between prime and target. Fourth, representations do not need to be in long-term memory in order to be primed. Nonwords do not have representations in long-term memory, yet they can be primed. In the case of nonwords in the same-different task, the relevant representations are constructed dynamically from the reference stimulus. Finally, and most importantly, priming depends on the nature of the representations used to perform the task. More specifically, priming depends on the hypotheses that support the decision required to make a response.

Implications for theories of masked priming of words

The data reported here present major problems for all three of the accounts of masked priming described in the introduction. In particular, none can offer any explanation why there should be no word priming for the “different” trials in the same-different task. According to both the activation based theories and the entry-opening theory, priming should be an automatic consequence of lexical processing; it should not disappear when the task is changed. In the memory recruitment model, episodic representations should assist with the identification of the target, so the theory gives no reason to predict that the episodic representations that produce priming for words in lexical decision will not produce priming in the same different task.

Memory recruitment might possibly explain why there should be priming for “same” responses for both words and nonwords in the same-different task, as nonwords would then have a clear episodic representation. However, it is unclear from this view why “different” responses showed no priming for words or nonwords.

The data from Experiment 3 showing task effects on masked priming of letters is beyond the scope of these models; they simply have nothing to say about how the same-different task might be performed on letters. Of course, it might well be possible to extend the scope of these 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.

We noted in the introduction that research on masked priming has separated into three separate streams that have little contact. Our own work evolved in the context of a theory of visual word recognition. However, the Bayesian account of decision making based on noisy accumulation of evidence is completely general. There is nothing in the Bayesian approach that should restrict it to words, or to the lexical decision and same-different tasks. Indeed, we have already shown that the same principles hold for letter perception. It seems that those principles also hold for data collected in studies of unconscious cognition and visual masking. Recent research in both areas has shown that masked priming can be modulated by task instructions. Furthermore, some of the explanations of these data fit comfortably within the Bayesian framework.

‘Object updating’: Priming of faces and shapes

The closest parallel to our own work comes from very recent research on face perception by Enns and Oriet (2007). In a number of papers, Enns and colleagues (Di Lollo, Enns & Resnik, 2000; Enns, 2004; Enns & Di Lollo, 1997; Enns, Lleras & Moore, in press; Lleras & Moore, 2003) have argued that perception involves ‘object updating’ rather than ‘image updating’. This is the same as our claim that perception involves integrating evidence rather than integrating data. They also assume that information is sampled from the stimulus, and that this can support perceptual hypotheses, although the notion of “hypothesis” here is informal, rather than specifically Bayesian. Additionally, they argue that performance can be influenced by the participant’s intentions:

"performance should be strongly influenced by the degree to which the participant has been able to form a well-defined task template or filter for the anticipated display prior to its onset" (Enns & Oriet, 2007, p213). They supported this claim with data from two masked priming experiments. The first experiment used faces as stimuli. Participants were asked to classify faces according to sex, emotion (happy vs. sad) or race (Asian vs. Caucasian). The face targets could be primed by other faces that shared task-relevant features (e.g. both female in the sex judgment task), task irrelevant features (e.g. same race in the sex judgment task), or no features. Priming was found only when the prime and target shared task-relevant features.

A second experiment used geometric stimuli varying in color or shape, where the task was to categorize the stimuli according to either color or shape. Once again, priming was found only when the prime and target shared task-relevant features. The results of both of these experiments are consistent with the idea that participants set up task-specific perceptual hypotheses, and that this leads them to ignore irrelevant information.

Unconscious cognition: Number priming

One of the most significant features of the Forster and Davis (1984) paradigm is that it is insensitive to the format of the prime; lower-case primes will prime upper-case targets. Similar format-independent effects have been found with numbers, where priming can be obtained between Arabic numerals and numeral words, and vice-versa (Dehaene, Naccache, Le Clec, Koechlin, Mueller, Dehaene-Lambertz, et al. 1998). Priming also appears to be driven by the semantic properties of numerals. In a task where participants are required to judge whether a number is greater or less than five, responses are influenced by the magnitude of number primes, even if those numbers have never been encountered previously as targets (Kunde, Kiesel & Hoffmann, 2003; Naccache & Dehaene, 2001). This indicates that the effect is not simply a consequence of learned associations between primes and responses. In this paradigm we would argue that participants are sampling the input and collecting evidence in favor of the hypotheses “greater than five” and “less than five”. Neither the format of the numeral, nor its specific identity is relevant to these hypotheses.

In an interesting variant of the number priming task, Kunde et al. (2003) showed that this form of masked priming is also determined by the exact form of the instructions. In Experiments 2 and 3 of that paper the target stimuli were the Arabic numerals 1, 4, 6 and 9, and the numerals 1-9 were used as primes. In the first of these experiments the instructions were to judge whether a number was greater than or less than five. With these instructions they obtained the standard finding of priming from the non-target numerals 2, 3, 7, and 8. In this task, participants are effectively being instructed to test a hypothesis about the magnitude of the stimuli, and the effect of primes is determined by their magnitude. However, in Experiment 3 the instructions were to press the left hand button in response to the numerals 1 and 4, and the right hand button in response to the numerals 6 and 9. With these instructions the non-target numeral no longer produced priming. These instructions specify a hypothesis that is specific to the identity of the target numerals. The semantic relationship between target and non-target numerals is of no consequence. In a parallel with our own results, prime-target pairs that produce priming under one set of instructions produce no priming with another set of instructions.

Kunde and colleagues explained their data in terms of their “action-trigger” hypothesis: “If a stimulus matches to the release conditions of an action trigger, the related action is automatically activated causing congruency effects if the stimulus was a prime)." (Kiesel, Kunde & Hoffmann, 2007, p 4). Although the action-trigger hypothesis places an emphasis on the importance of task instructions, it differs significantly from the view advanced here. A clear implication of the action-trigger hypothesis is that the prime and target are in some sense treated as separate events. This also seems to be the implication of Dehaene et al’s suggestion that “Subjects would unconsciously apply task instructions to the prime”.

So, somewhat surprisingly, our Bayesian account of priming developed in the context of word recognition has more in common with accounts developed to explain visual masking and unconscious processing than it does with other models of word recognition. All of the accounts we have just described emphasize the importance of task instructions, or hypotheses. The object updating theory comes closest to our own view with its suggestion that there is a sampling process and that the prime and target are treated as a single object. What all of these theories lack though is an account of how the tasks are performed, or the hypotheses are tested. Without a model of task performance, it is impossible to derive specific predictions from these theories that would explain the data reported here. For example, simply specifying that participants unconsciously apply the task instructions to the prime (Dehaene at al., 1998) does not make a clear prediction as to whether there will be response congruence effects in lexical decision. Furthermore, none of the theories explains why there should be priming for words but not nonwords in lexical decision, or for ‘same’ response but not for ‘different’ responses in the same-different task. The Bayesian account replaces these informal notions with a formal account of the decision processes. This enables us to give a precise specification of how these tasks are performed, and to generate specific predictions. These predictions were confirmed in the experiments reported here.

Conclusion

Three important assumptions underlie our explanation of masked priming. The first is the basic principle that people approximate optimal Bayesian decision makers. Although this principle is instantiated in a model of visual word recognition (but see Norris & McQueen, in press, for a Bayesian model of spoken word recognition) it applies to perception in general. This claim is supported by data from face perception and number identification. The second is that the masked priming paradigm tricks the perceptual system into processing primes and targets as a single perceptual object. Although people approximate optimal Bayesian decision makers, there are still limitations on their performance. Because people fail to perceive the prime as a separate object they fail to discount the evidence from the prime, and so integrate the prime with the target. Finally, the way the prime and target are integrated is in terms of integrating evidence, or multiplying likelihood ratios. This is the crucial insight. Words are normally static objects. Models that integrate at the level of raw data can cope with static objects, but they break down when faced with dynamically changing input produced by masked primes. Masked priming has tricked the perceptual system into revealing an important secret: All perception involves Bayesian inference based on accumulation of noisy evidence.

References

Andrews, A. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 802-814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 234-254.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4*(439-461).

Anderson, J.R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*, 702-719.

Anderson, J.R., & Schooler, L.J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396-408.

Balota, D.A., Cortese, M., M.J., Sergent-Marshall, S.D., Spieler, D., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283-316.

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*, 268-293.

Bodner, G.E., Masson, M.E.J., & Richard, N.T. (2006). Repetition proportion biases masked priming of lexical decisions. *Memory & Cognition, 34,* 1298-1311.

Bogacz, R. (2007). Optimal decision-making theories: Linking neurobiology with behaviour. *Trends in Cognitive Sciences, 11,* available online, 2 February 2007.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J.D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review, 113*, 700-765.

Boles, D.B., & Clifford, J.E. (1989). An upper- and lowercase alphabetic similarity matrix, with derived generation similarity values. *Behavior Research Methods, Instruments, & Computers, 21*, 579-586.

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*, 1705-1719.

Bruder, G. (1978). Role of visual similarity in the word superiority effects obtained with the simultaneous matching task. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 88-100.

Colombo, L. (1986). Activation and inhibition with orthographically similar words. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 226-234.

Coltheart, M., Davelaar, E., Jonasson, J.T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed,), *Attention and Performance, VI*. Hillsdale: Erlbaum. pp.535-555.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review,* 108, 204-256.

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*. New York: NY: Psychology Press, pp. 121-170.

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

Dehaene, S., Changeux, J-P., Naccahe, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences, 10*, 204-211.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van der Moortele, P.F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature, 395(6702)*, 597-600.

Di Lollo, V., Enns, J.T., & Rensink, R.A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129,* 481-507.

Draine, S.C., & Greenwald, A.G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General, 127*, 286-303.

Enns, J.T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research, 44,* 1321-1331.

Enns, J.T., & Di Lollo, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological Science, 8*, 135-139.

Enns, J.T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4*, 345-352.

Enns, J.T., Lleras, A., & Moore, C.M. (in press). Object updating: A force for perceptual continuity and scene stability in human vision. In R. Nijhawan (Ed.), *Problems of space and time in perception and action*. Cambridge University Press.

Enns, J.T., & Oriet, C. (2007). Visual similarity in masking and priming: The critical role of task relevance. *Advances in Cognitive Psychology, 3*, 211-216. Available online at http://www.ac-psych.org.

Forster, K.I. (1985). Lexical acquisition and the modular lexicon. *Language and Cognitive Processes, 1*, 87-108.

Forster, K.I. (1987). Form-priming with masked priming: The best match hypothesis. In M. Coltheart (Ed.), *Attention & Performance XII: The psychology of reading*. Hove, UK: Lawrence Erlbaum Associates, pp. 127-146.

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

Forster, K.I. (2004). Category size effects revisited: Frequency and masked priming effects in semantic categorization. *Brain and Language, 90*, 276-286.

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

Forster, K.I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *Quarterly Journal of Experimental Psychology*, *39*, 211-251.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*, 116-124.

Forster, K.I., Mohan, K., & Hector, J. (2003). The mechanics of masked priming. In S. Kinoshita & S.J. Lupker (Eds.), *Masked priming: The state of the art*. New York: NY: Psychology Press, pp. 3-38.

Frost, R., Forster, K.I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 829-856.

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

Gregory, R.L. (2007) *Illusion: The Phenomenal Brain*, Oxford University Press.

Huber, D. E., Shiffrin, R. M., Lyle, K. B., & Ruys, K. I. (2001). Perception and preference in short-term word priming. *Psychological Review, 108*(1), 149-182.

Irwin, D. (1996) Integrating information across saccadic eye movements. *Current Directions in Psychological Science*, 5(3), 94–100.

Jacobs, A.M., & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 1174-1188.

Kahneman, D., Treisman, A., & Gibbs, B.J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*, 175-219.

Kiesel, A., Kunde,W., & Hoffmann, J. (2007). Mechanisms of subliminal response priming. *Advances in Cognitive Psychology, 3*, 307-315.

Kinoshita, S. (1997). Masked target priming effects on feeling-of-knowing and feeling-of-familiarity judgments. *Acta Psychologica*, *97*, 183-199.

Kinoshita, S. (2006). Additive and interactive effects of word frequency and masked repetition in the lexical decision task. *Psychonomic Bulletin & Review*, *13*, 668-673.

Kinoshita, S., & Kaplan, L. (2008). Priming of abstract letter identities in the letter match task. *Quarterly Journal of Experimental Psychology*, in press.

Kinoshita, S., & Lupker, S.J. (Eds.) (2003). *Masked priming: The state of the art*. New York: NY: Psychology Press.

Klinger, M.R., Burton, P.C., & Pitts, G.S. (2000). Mechanisms of unconscious priming: I. Response competition, not spreading activation. *Journal of Experimental psychology: Learning, Memory and Cognition, 26*, 441-455.

Knill, D. C., Kersten, D., & Yuille, A. (1996). A Bayesian formulation of visual perception. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian Inference*. Cambridge: Cambridge University Press. pp. 1-21.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical transactions of the Royal Society*, *B*, *362*, 857-875.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Kunde, W., Kiesel, A., & Hoffmann, J. (2003). Conscious control over the content of unconscious cognition. *Cognition, 88*, 223-242.

Laming, D.R.J. (1968). *Information theory and choice-reaction times*. New York: Academic Press.

Lleras, A., & Moore, C.M. (2003). When the target becomes the mask: Using apparent motion to isolate the object-level component of object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 106-120.

Logan, G.D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology, 21*, 1-35.

Marcel, A.J. (1983). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology, 15*, 238-300.

Marmurek, H.H. (1989). Familiarity effects and word unitization in visual comparison tasks. *Memory & Cognition, 17,* 483-489.

Martin, R.C., & Jensen, C.R. (1988). Phonological priming in the lexical decision task: A failure to replicate. *Memory & Cognition, 16*, 505-521.

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*. New York: NY: Psychology Press, pp. 57-96.

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

McClelland, J.J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review, 88,* 375-407.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*(2), 131-141.

McGurk, H., & MacDonald, T. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Middleweerd, M. J., & Plomp, R. (1987). The effect of speechreading on the speech-reception threshold of sentences in noise. *Journal of the Acoustical Society of America*, *82*(6), 2145-2147.

Murray, W., & Forster, K.I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review, 111*, 721-756.

Naccache, L., & Dehaene, S. (2001). Unconscious semantic priming extends to novel unseen stimuli. *Cognition, 80*, 223-237.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G.W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, pp.264-336.

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. (1986). Word recognition: Context effects without priming. *Cognition, 22*, 93-136.

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

Norris, D. (submitted). Why the lexical decision task really does tell us a lot about word recognition: Modeling reaction-time distributions with the Bayesian Reader

Norris, D., & McQueen, J. M. (in press). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*..

Pelli, D.G., Farell, B., & Moore, D.C. (2003, June 12). The remarkable inefficiency of word recognition. *Nature, 423,* 752-756.

Perea, M., Fernandez, L., & Rosa, E. (1998). El papel del status lexico y de la frecuencia del estimulo-senal en la condition no relacionada con la tenica de presentacion enmascarada del estimulo-senal. (The role of the lexical status and the frequency of the unrelated prime in the masked priming technique). *Psicologica, 19,* 311-319.

Perea, M., & Gotor, A. (1997). Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming. *Cognition, 62*, 223-240.

Perea, M., & Rosa, E. (2000). The effects of orthographic neighborhood in reading and laboratory identification tasks: A review. *Psicologica, 21*, 237-340.

Proctor, R.W. (1981). A unified theory for matching-task phenomena. *Psychological Review, 88*, 291-326.

Raaijmakers, J. G. (2003). A Further Look at the “Language-as-Fixed-Effect Fallacy”. *Canadian Journal of Experimental Psychology, 57*(3), 141-151.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to Deal with “The Language-as-Fixed-Effect Fallacy”: Common Misconceptions and Alternative Solutions. *Journal of Memory and Language, 41*(3), 416-426.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*, 89-102.

Rastle, K., Davis, M.H., Marslen-Wilson, W.D., & Tyler, L.K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes, 15*, 507-537.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59-108.

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.

Sereno, J.A. (1991). Graphemic, associative, and syntactic priming effects at brief stimulus onset asynchrony in lexical decision and naming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*, 459-477.

Simons, D. J., Mitroff, S. R., & Franconeri, S. L. (2003). Scene perception: what we can learn from visual integration and change detection. In M. Peterson & G. Rhodes (Eds.), Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes (335-355). Oxford University Press.

Snodgrass, J.G., & Mintzer, M. (1993). Neighborhood effects in visual word recognition: Facilitatory or inhibitory? *Memory & Cognition, 21*, 247-266.

Stone, M. (1960). Models of choice reaction time. *Psychometrika, 25*, 251-260.

Van Heuven, W.J., Dijkstra, T., Grainger, J., & Schreifers, H. (2001). Shared neighborhood effects in masked orthographic priming. *Psychonomic Bulletin & Review, 8*, 96-101.

Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences, 100*, 6275-6280.

Wagenmakers, E-J., Ratcliff, R., Gomez, P., & McKoon, G. (2007). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, in press.

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.

Weidemann, C.T., Huber, D.E., & Shiffrin, R.M. (2005). Confusion and compensation in visual perception: Effects of spatiotemporal proximity and selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 40-61.

Weiss, Y., Simoncelli, E.P., & Adelson, E.H. (2002). Motion illusions as optimal percepts. *Nature neuroscience, 5*, 598-604.

Whaley, C.P. (1979). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*, 143-154.

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.

Appendix

Lexical decision in the Bayesian Reader.

Representation of letters

In the representations used in the simulations reported here, each position-specific letter is represented by a 26 element binary vector where one coordinate is set to 1 and the remainder to 0. Each letter vector therefore corresponds to a point in multidimensional perceptual space.

Sampling

Each sample presented to the model is derived by adding zero-mean Gaussian noise with standard deviation σ to each coordinate of the vector corresponding to each input letter. After each new sample x_t is received, the centroid $\bar{x}$ of the sample vectors is updated for each letter position. This is simply calculated by computing the average value of the sample vectors for each coordinate. The root mean square (RMS) radial dispersion s_d is then computed, and the corresponding standard error of the dispersion of the centroid $s_{\bar{x}}$ is derived. These are calculated on the basis of the distributions of distances between the location of each sample vector and the location of the centroid of the sample vectors. The euclidean distance d_t between the sample centroid $\bar{x}$ and a sample vector x_t is given by

$$d_t = \|\bar{x} - x_t\| = \sqrt{\sum_{i=1}^{i=n} (\bar{x}_i - x_{ti})^2} \tag{A1}$$

where n is the number of coordinates of the vector (number of dimensions/features) which, in the present case is 26 per letter.

Then, on the basis of N samples, the RMS radial dispersion s_d is given by

$$s_d = (\sum_{t=1}^{t=N} d_t^{\,2} \Big/ N)^{1/2}\,, \tag{A2}$$

and, by analogy with the unidimensional case, $s_{\bar{x}}$ is derived as:

$$s_{\bar{x}} = s_d / \sqrt{N} \tag{A3}$$

where N is the sample size. The essential role of $s_{\bar{x}}$ is to act as a scaling factor in the log likelihood which reflects both the actual noise in the data and the increased precision that comes with growing amounts of data.

<u>Calculating the likelihood of each letter:</u>

Let x_j denote the sub-vector of a sample x that corresponds to the letter in the j^{th} position in the letter string stimulus The likelihood of each letter is then calculated on the basis of the distance, $D_{ji} = \|\bar{x}_j - L_i\|$ between each letter in its binary vector form and the mean of the input samples for that letter position. The likelihood, $f(x_j \mid L_i)$ is modelled by

$$f(x_j \mid L\ ) = \left(2\pi\sigma_M^2\right)^{-n/2}\ e^{-D_{ji}^2 / 2\sigma_M^2}\ . \qquad (A4)$$

Where σ_M is the standard error of the mean. Note that as all the calculations are based on likelihood ratios, the first term of equation (A4) can be ignored, as it is common to all likelihoods, giving:

$$f(x\ \mid L\ ) \propto\ e^{-D^2 / 2\sigma_M^2} \qquad (A5)$$

The probability of each letter is given by:

$$P(L_x \mid x_x) = f(x_x \mid L_x) \Big/ \sum_{i=1}^{i=26} f(x_i \mid L_i) \qquad (A6)$$

Having obtained the probability of each letter in each position in the input, the likelihood of observing the sequence of letters LS = (L_1 , … , L_n)corresponding to each word is given by the product of the probabilities of the letters in the word:

$$P(LS\ \mid x) = \prod_{i=1}^{n} P(L_i \mid x\ ) \qquad (A7)$$

Word probabilities can then be calculated from Bayes' theorem,

$$P(W \mid x) = P(W) \times P(LS_W \mid x) \Big/ \sum_{k=1}^{k=m} (P(W_k) \times P(LS_k \mid x)) \qquad (A8)$$

where m is the number of words in the lexicon and LS_W is the letter sequence corresponding to the word *W*.

Note that, ideally, instead of using the multidimensional distance measure, likelihoods would be calculated independently for each coordinate/feature and then combined to calculate letter probabilities.

In the original implementation of the model, the procedure for calculating the likelihood that the input was a nonword involved two terms. The first was the likelihood that the input was produced by a *virtual-nonword* which was located as close as possible to the mean of the input samples, but was as least some minimum distance from the nearest word. In the simulations reported in Norris (2006), that distance always corresponded to the distance between two letter-strings differing by a single letter. An additional *background-nonword* term was incorporated which made allowance for the possibility that nonwords could be located anywhere in lexical space. That is, nonwords should be considered to be selected from some large set of nonwords that could potentially appear in the experiment. The combination of the *virtual-nonword* and the *background-nonwords* is therefore designed to capture the fact that, in the lexical decision task, nonwords can appear anywhere in lexical space, but are generally similar to words.

Because the new implementation is based on letter probabilities there is no straightforward representation of distance in lexical space, and likelihoods can only readily be computed for specific letter-strings. The *virtual-nonword* is now represented by the letter string with the highest $P(LS)$ that is not a word. This is assigned a prior probability corresponding to the mean prior of the words.

Computation of the *background-nonword* term takes advantage of the fact that the sum of letter probabilities at each position, and the sum of the probabilities of all possible letter-strings, must both necessarily be 1.0. If a particular string of letters has been identified, each with P(letter|input) = 1.0, and these letters form a word, say WORK, then the input must be a word. If the final letter in the string is ambiguous between K and D (P(K|input) = 0.5, P(D|input) = 0.5) then the input must be a word because the sum of the letter-string probabilities corresponding to words is 1.0. In contrast, if the input is WORB and final letter is completely inconsistent with any word, (P(B|input) = 1.0), then the input must be a nonword. The sum of the letter-string probabilities corresponding to words could therefore be used to compute the likelihood that the input is a word because the summed probability of the letter-strings corresponding to nonwords is 1-summed probability of letter-strings corresponding to words.

Defined in this way, the *background-nonwords* include the *virtual-nonword.* In effect the virtual nonword is just like any other background nonword, apart from the fact that it is assigned a larger prior (the mean word prior). Therefore the virtual-nonword letter-string probability $P(LS_{vn})$ has to be separated from the summed *background-nonword* string probability in order to multiply each by the appropriate priors.

The prior of the *virtual-nonword*, *P(VN)*, is set to be the same as the mean word prior, (1.0/(2 x the number of words in the lexicon)), and the sum of the word and nonword priors is each set to 0.5, as the target is equally likely to be a word as a nonword. The prior of the *background-nonwords*, *P(BN)* = 0.5 – *P(VN)*. The *likelihood of the background-nonwords (BNL)*, excluding the *virtual-nonword* (VN) is then given by

$$BNL = P(BN) \times \left(1.0 - P(LS_{vn}) - \sum_{i=1}^{i=m} P(LS_i)\right) \qquad (A9)$$

which corresponds to the likelihood that remains after subtracting the virtual nonword and word likelihoods. And the likelihood of the virtual-nonword is

$$VNL = P(LS_{vn}) \times P(VN) \qquad (A10)$$

The likelihood that the input is a word is given by

$$WL = \sum_{i=1}^{i=m} (P(W_i) \times P(LS_i)) \qquad (A11)$$

(where $P(W_i)$ is now the prior adjusted for the fact that the sum of all word priors must be 0.5) and the probability that the input is a word is given by

$$P(a\,word) = WL / (WL + VNL + BNL) \qquad (A12)$$

Appendix B

Convert probability to odds:

$odds = P/(1-P)$

Calculate post-test odds by multiplying prior-odds by the likelihood ratios of the sources of evidence to be combined:

$post\text{-}test\ odds = odds \times LR1 \times LR2\ etc$

Convert odds to probability:

$final\ P = post\text{-}test\ odds/(1+post\text{-}test\ odds)$

Appendix C

List of stimuli used in Experiment 1 and 2

High-frequency target words. LEAVE, WOMEN, WATER, GROUP, MAJOR, DEATH, WHITE, EARLY, HEAVY, DRIVE, HOTEL, EARTH, STORY, HEART, AFTER, MONEY, PEACE, VOICE, TRADE, LARGE, CHILD, SCENE, YOUNG, SENSE, BROWN, TABLE, SOUTH, CLASS, PRESS, BLOOD, HOUSE, MONTH, THINK, BLACK, BROAD, LEAST, CLEAR, SHARE, GREEN, PLANT, PLANE, DAILY, COURT, TODAY, STILL, WOMAN, BELOW, REACH, STAGE, VALUE, ORDER, EVERY, GREAT, HUMAN, ISSUE, FIRST, AMONG, THOSE, READY, UNDER, FIELD, TRIAL, RIVER, PRICE, LEVEL, MIGHT, ROUND, NORTH, NIGHT, COULD, SHALL, STOCK, SMALL, WHILE, WRONG, PARTY, ALONG, STOOD, BASIC, YEARS

Low-frequency target words. WHEAT, LEMON, SCOUT, FLOUR,GRILL, GLOVE, PETTY, BOAST, ARROW, AGONY, SUPER, KNOCK, BLEAK, BRISK, FRAIL, PEARL, SPADE, ATLAS, WRECK, DEMON, THIGH, PLANK, BRICK, BLAST, SQUAT, HOUND, TOWEL, GREED, BLUSH, BOUGH, SONAR, BLINK, ADOPT, CREST, BLAND, CREEP, STARE, STEER, CREED, CHORE, NIECE, ESSAY, ATTIC, THIEF, BARGE, LUNAR, PURSE, SPORT, SWEEP, BLAZE, FEAST, ANKLE, GLOBE, GLAND, VIRUS, CANDY, CARGO, MOUSE, BOOTH, WEIRD, TRUNK, TOAST, LYRIC, COMIC, CANOE, FILLY, SLEET, GREED, MINUS, ROAST, BATON, SEWER, TIMER, SLACK, REPEL, CREPE, SCARE, CHEEK, SLATS, STEAK

Nonword targets. CLOOR, SHEND, SHICE, BREET, THARE, FRONE, BEARK, TRULD, PROUT, GOUND, WOUCH, TRAND, CHERD, STERM, THACE, THEAK, WHERT, THIRP, DREAT, BREAT, BLOUT, SONTH, GRAIM, PIGHT, SELCH, MINCH, FLEEP, STARP, THEST, RILSE, CHESK, HILCH, THOVE, DRASH, SLART, FRIND, SMACE, BLACE, BLING, THEEN, TRONG, THENG, NOULD, CHOUT, GRALD, COULK, SOULD, GRICH, FRISS, ROULT, FLOST, WONTH, SHONG, WHOSK, CLEST, COURM, THEAN, SHAIR, SMAIR, PLART, SHISK, CHAND, BLONG, CHALL, STIRE, BANCH, WHOCE, BRULT, CHIST, LITCH, PRING, DRING, THERP, STERK, SPING, SHINT, WOULT, GLAST, PREAT, BROUL

High-frequency word control primes. third, until, union, chief, spent, floor, march, thing, point, about, bring, sound, given, music, which, built, truth, staff, eight, visit, range, total, where, local, state, force, piece, teeth, final, image, again, serve, power, front, seven, above, often, began, start, write

Low-frequency word control primes. slump, sword, alarm, width, couch, bunch, choir, ivory, shelf, exert, alibi, medal, ghost, path, dough, onion, thumb, merry, salad, quack, abbey, sober, swell, fever, flock, stack, scrap, stamp, arena, lever, juice, torso, nurse, gloom, choke, float, chill, cliff, blunt

Nonword control primes. chade, quile, clost, warld, frint, wreat, torth, glome, pronk, fooph, glain, bedge, crawn, flide, dench, hathe, brith, shird, crame, crulb, deash, bruck, broon, shool, clond, chilk, thind, sheck, sheel, chirk, trape, fluge, slass, chond, wrunk, lorch, sland, lough, floop, shund

High-frequency *Different* reference words. paper, moral, three, radio, space, cable, youth, glass, dress, flood, horse, mouth, thank, block, bread, lease, clean, sharp, greet, short, found, study, doubt, faith, right, bound, forth, light, would, shell, stick, smell, whole, wring, parts, alone, stool, basis, yearn

Low-frequency *Different* reference words. polar, scrub, algae, dodge, logic, mound, vowel, breed, flush, cough, solar, blank, adapt, crust, blond, creek, stark, steep, creep, chord, elbow, alley, shaft, tenth, quill, silly, fleet, freed, sinus, boast, bacon, sever, tiger, slick, rebel, crept, scars, cheer, slate, steam

Nonword *Different* references. trind, bruld, cousk, thear, brean, geark, moung, dreak, starm, chard, thice, brone, shent, cound, thire, carth, drout, mousk, wherp, worch, bleam, boult, shain, briss, moulk, shing, thead, couse, jance, flast, skirm, shace, coure, brich, yould, plard, smain, whesk, wouth, clast

Author notes

Dennis Norris, MRC Cognition and Brain Sciences Unit, U.K., and Sachiko Kinoshita, Macquarie Centre for Cognitive Science (MACCS) and Department of Psychology, Macquarie University, Sydney, Australia.

Research reported in this paper was partly supported by funding from the Australian Research Council Discovery Project Grant (DP0877084) to the authors. Thanks to Maarten van Casteren for programming the Bayesian Reader, and Ian Nimmo-Smith for valuable advice, and to Shaun Greenfield for research assistance.

Mail correspondence concerning this article may be addressed either to Dennis Norris, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, United Kingdom, CB2 2EF, or Sachiko Kinoshita, Macquarie Centre for Cognitive Science (MACCS) and Department of Psychology, Macquarie University, Sydney, NSW, Australia, 2109. Electronic mail may be sent to: dennis.norris@mrc-cbu.cam.ac.uk or sachiko@maccs.mq.edu.au.

Footnotes

1. In apparent contradiction to the results reported by Forster (1985, Experiment 3, 4) showing that masked priming facilitated only episodic recognition decisions for OLD items, Rajaram (1993, Experiment 3) reported that masked priming increased the bias towards responding OLD for both old and new items, only for those decisions that are not accompanied by retrieval of contextual information – the KNOW responses (see Kinoshita, 1997, Experiment 1 for a replication). Unlike Forster, Rajaram did not find masked priming facilitated recognition judgment latency. An important difference between these studies is that Forster used a small set of items as the study set (20 nonwords in Experiment 3, 30 words in Experiment 4) and as a result, recognition performance was high (mean hit rate 93.2%, false alarm rate 6.9%). In contrast, in the studies that showed the bias effect, a larger number of items were studied and recognition performance was lower (in Rajaram, 60 words were studied and the hit rate was 63.5% and false alarm rate 20.5%; in Kinoshita, 60 words were studied and the hit rate was 70.2%, false alarm rate 7.3%). That is, the bias effect reported by Rajaram and Kinoshita is limited to conditions in which recognition performance is low, and likely reflects misattribution of source of "feeling of familiarity" enhanced by masked priming (see Kinoshita, 1997, for discussion).

Table 1.

Mean Lexical Decision Latencies (RT, in ms), Standard Deviations (in parentheses) and Percent Error Rates (%E) in Experiment 1 (lexical decision task)

	Target type					
	High-frequency word		Low-frequency word		Nonword	
Prime type	RT	%E	RT	%E	RT	%E
Identity	526 (68)	3.3	568 (86)	9.0	667 (121)	11.3
Response-congruent	561 (58)	2.7	634 (82)	12.3	682 (124)	10.4
Response-incongruent	561 (53)	4.0	621 (71)	12.9	672 (108)	12.7
Identity priming effect [a]	35	0	60	3.6	10	0.3
Response congruence Effect [b]	0	1.3	-13	0.6	10	2.3

a. Identity priming effect = Average of response-congruent and response-incongruent – identity

b. Response congruence effect = response-incongruent – response-congruent

Table 2.

Mean Decision Latencies (RT, in ms), Standard Deviations (in parentheses) and Percent Error Rates (%E) in Experiment 2 (same-different matching task)

	Target type					
	High-frequency word		Low-frequency word		Nonword	
Response type and prime type	RT	%E	RT	%E	RT	%E
Same						
Identity	415 (84)	5.0	428 (80)	4.6	436 (90)	4.4
Control	511 (78)	10.2	515 (81)	12.3	523 (95)	12.5
Priming effect	96	5.2	87	7.7	87	8.1
Different						
Identity	501 (75)	5.6	497 (82)	5.8	508 (88)	6.3
Control	508 (77)	4.4	491 (71)	4.4	497 (61)	5.2
Priming effect	7	-1.2	-6	-1.4	-11	-1.1

Table 3.

Examples of reference, prime and target triplets used in Experiment 3 (letter match task)

	Prime type	
Task, Response and Letter type	Identity	Control
Cross-case match - Same		
Dissimilar	A-A-a, a-a-A	A-B-a, a-b-A
Similar	C-C-c, c-c-C	C-X-c, c-x-C
Cross-case match - Different		
Dissimilar	A-B-b, a-b-B	A-R-b, a-r-B
Similar	C-X-x, c-x-X	C-P-x, c-p-X
Same-case match - Same		
Dissimilar	a-A-a, A-a-A	a-B-a, A-b-A
Similar	c-C-c, C-c-C	c-X-c, C-x-C
Same-case match - Different		
Dissimilar	B-B-b, A-b-B	B-R-b, A-r-B
Similar	c-X-x, C-x-X	c-P-x, C-p-X

Note. Response type (Same/Different) refers to the response required to the reference and target pair; Prime type (Identity/Control) refers to the prime status with respect to the target; Letter type (Similar/Dissimilar) refers to featural similarity of uppercase and lowercase letter pair

Table 4.

Mean Decision Latencies (RT, in ms), Standard Deviations (in parentheses) and Percent Error Rates (%E) in Cross-case Match Task in Experiment 3

	Letter type (cross-case similarity)			
	Dissimilar		Similar	
Response type and prime type	RT	%E	RT	%E
Same				
Identity	419 (90)	7.3	405 (93)	10.8
Control	489 (106)	14.6	471 (88)	15.3
Priming effect	70	7.3	66	4.5
Different				
Identity	500 (88)	5.9	484 (94)	5.9
Control	530 (115)	6.3	478 (90)	3.5
Priming effect	30	0.4	-6	-2.4

Table 5.

Mean Decision Latencies (RT, in ms), Standard Deviations (in parentheses) and Percent Error Rates (%E) in Same-case Match Task in Experiment 3

	Letter type (cross-case similarity)			
	Dissimilar		Similar	
Response type and prime type	RT	%E	RT	%E
Same				
Identity	452 (79)	8.8	416 (79)	9.6
Control	460 (78)	11.0	459 (81)	13.1
Priming effect	8	2.2	43	3.5
Different				
Identity	489 (73)	13.6	440 (71)	4.8
Control	494 (86)	9.2	446 (74)	2.1
Priming effect	5	-4.4	6	-2.7

List of figures

Figure 1 Likelihood functions of two words differing on a single perceptual dimension. The lower curves show likelihood functions early in time, which the more peaked curves show the likelihood functions later on in processing.

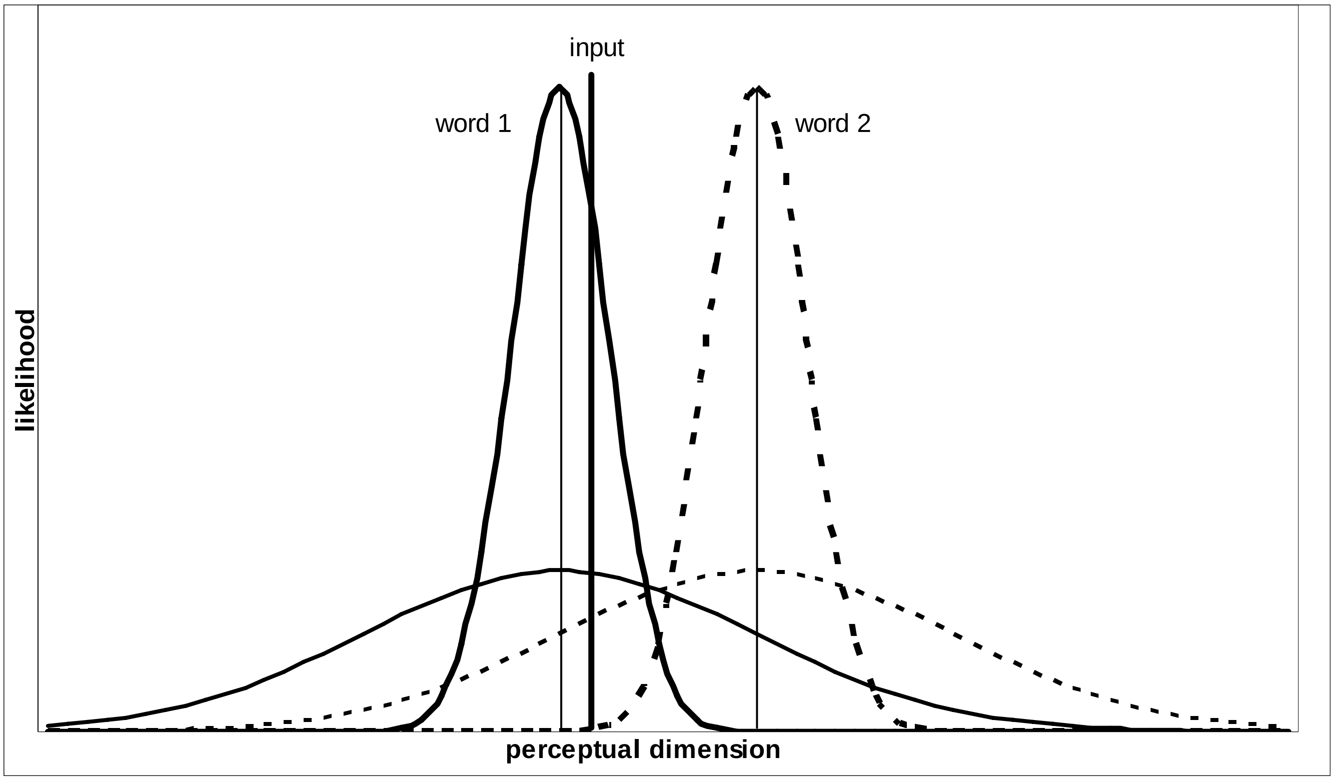
Figure 2. Priming of nonwords

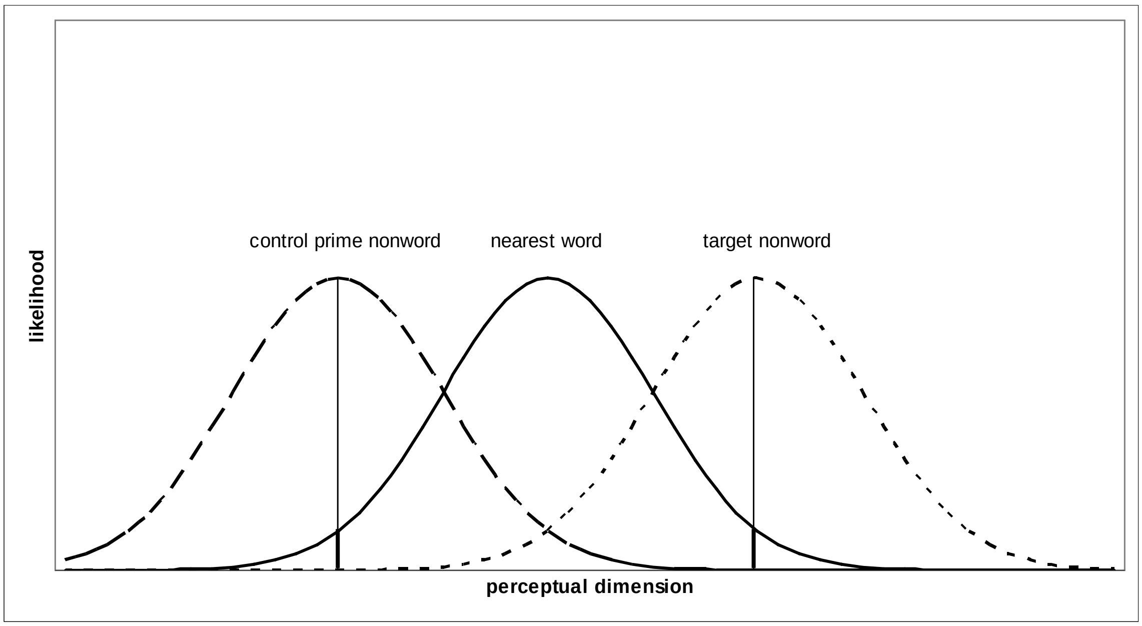
The figure illustrates why 'nonword' responses in the Bayesian Reader should not be influenced by whether the prime is the same as the target. All that matters is the distance between the nonword (whether prime or target) and the representation of the nearest word.

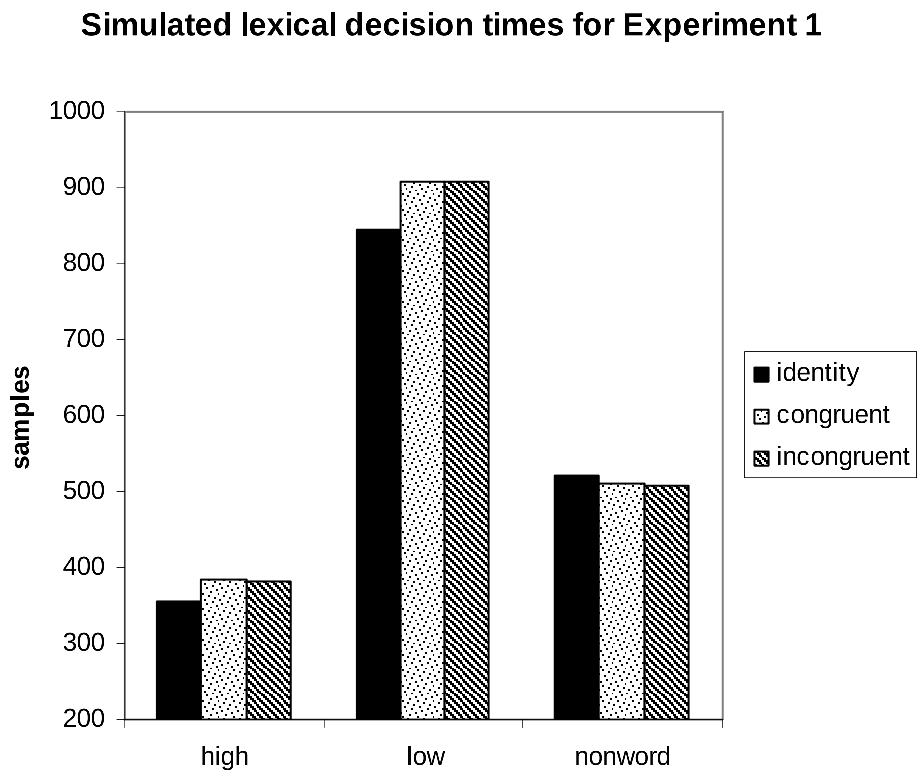
Figure 3. Bayesian Reader simulation of lexical decision times in Experiment 1

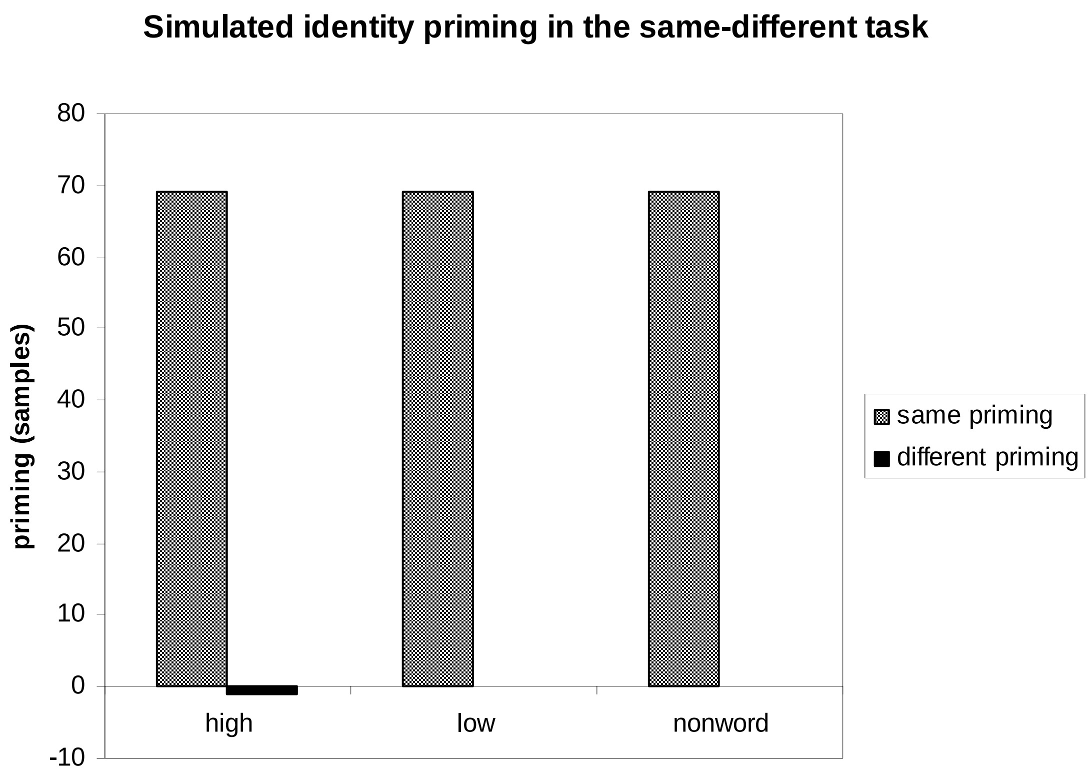
The graph shows the number of samples required for the model to make a lexical decision response. Note that although the number of samples is in a similar range to the range of RTs, the simulations have not been scaled to fit the RTs, and the simulations make no allowance for components of decision time outside the scope of the model.

Figure 4. Simulation of priming effects in same-different task (Experiment 2)


input
word 1
word 2
likelihood
perceptual dimension


control prime nonword
nearest word
target nonword
likelihood
perceptual dimension


Simulated lexical decision times for Experiment 1
samples
1000
900
800
700
600
500
400
300
200
high
low
nonword
identity
congruent
incongruent


Simulated identity priming in the same-different task
priming (samples)
80
70
60
50
40
30
20
10
0
-10
high
low
nonword
same priming
different priming