

**Parafoveal preview effects from word N+1 and word N+2 during reading: A critical review
and Bayesian meta-analysis**

Martin R. Vasilev¹

Bernhard Angele

Bournemouth University

Corresponding author¹ at:

Department of Psychology

Bournemouth University

P104, Poole House

Talbot Campus, Fern Barrow

Poole, Dorset,

BH12 5BB

United Kingdom

Phone: +447497432487

Email: mvasilev@bournemouth.ac.uk

Acknowledgments

M.R.V was supported by a PhD studentship from Bournemouth University and by the Erasmus Mundus II programme of the European Union (grant # 2013-1985/001-001-EMII-EMMC). The funding bodies had no role in the design of this study, the data analysis or the writing of this paper. Portions of this work were completed while the first author was at the University of Groningen and the University of Potsdam. We thank Tim Slattery for valuable discussions on a previous version of this manuscript. All data analysis scripts are available at:

<https://github.com/martin-vasilev/Bmeta>

Abstract

The use of gaze-contingent display techniques to study reading has shown that readers attend not only to the currently fixated word, but also to the word to the right of the current fixation.

However, a critical look at the literature shows that there are a number of questions that cannot be readily answered from the available literature reviews on the topic. First, there is no consensus on whether readers also attend to the second word to the right of fixation. Second, it is not clear whether parafoveal processing is more efficient in languages such as Chinese. Third, it is not well understood whether the measured effects are confounded by the properties of the parafoveal mask. The present study addressed these issues by performing a Bayesian meta-analysis of 93 experiments that used the boundary paradigm (Rayner, 1975). There were three main findings: 1) the advantage of previewing the second word to the right is modest in size and likely not centred on zero; 2) Chinese readers seem to make a more efficient use of parafoveal processing, but this is mostly evident in gaze duration; 3) there are interference effects associated with using different parafoveal masks that roughly increase when the mask is less word-like.

Keywords: parafoveal processing, reading, preview benefit, perceptual span, eye-movements

The advance of eye-tracking technology has allowed an unprecedented opportunity to understand how the reading process unfolds in space and time. One of the advantages of this method is the possibility to precisely manipulate what participants see in real time while they are reading sentences on the screen. The use of such techniques has shown that readers not only process the currently fixated word, but that they also benefit from previewing the upcoming word in parafoveal vision (Rayner, 1998). This so-called *preview benefit* effect is measured as shorter fixation durations on a word for which correct preview information was available during the preceding fixation.

The preview benefit effect is arguably one of the most robust and least controversial findings in the literature, and it has inspired many subsequent experiments as well as models of reading (Schotter, Angele, & Rayner, 2012). A particular issue highlighted by the preview benefit effect is the distinction between gaze location and the attentional focus during reading. While gaze location and attentional focus are usually identical in single-word recognition tasks, readers appear to routinely attend to the upcoming word as well. Many experiments have been dedicated to determining exactly which properties of an upcoming word readers can process while they are still fixating their gaze on the preceding word. It has been shown that readers can pre-process orthographic, phonological, morphological, and possibly semantic properties of the upcoming word (for a review, see Schotter et al., 2012).

However, despite four decades of research on the topic, the size of this effect has not been systematically quantified for the different experimental conditions and dependent measures. Even though many estimates of the size of this effect have been volunteered in the literature, no systematic statistical analysis has been undertaken. Therefore, this does not make it possible to evaluate in a precise manner the predictions of theories and the simulations of computational

models of eye-movement control during reading (Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reilly & Radach, 2006; see also Trukenbrod & Engbert, 2014). Also, while the existence of the preview benefit effect is virtually accepted, there is currently no consensus on whether readers also benefit from previewing the second word to the right of the current fixation. Moreover, the question of whether the preview benefit differs in languages with a different writing system, such as Chinese, has not been answered satisfactorily. Finally, it is currently not clear whether and to what extent the size of the preview benefit is confounded by interference effects associated with seeing a masked word in parafoveal vision.

The purpose of the present paper is two-fold. First, it undertakes a critical evaluation of the literature that highlights unresolved questions related to parafoveal processing during reading. In this aspect, it does not attempt to do a comprehensive review of the literature. Second and more importantly, the paper presents a meta-analytic investigation of parafoveal preview effects during reading in order to answer the questions raised in the introduction.

The Perceptual Span During Reading

When reading a sentence, the eyes do not move smoothly, but alternate between quick jump-like movements, also known as *saccades*, and short periods in which they are relatively stable. These periods of stability, also known as *fixations*, are crucial for the word recognition process. When a given word is fixated, readers view it with their *fovea*, or the central 2° of visual angle, where visual acuity is at its highest. However, it is now well known that readers also obtain information from the *parafovea*, which extends up to 5° of visual angle (Rayner, 1998). An illustration of this is presented in Figure 1A.

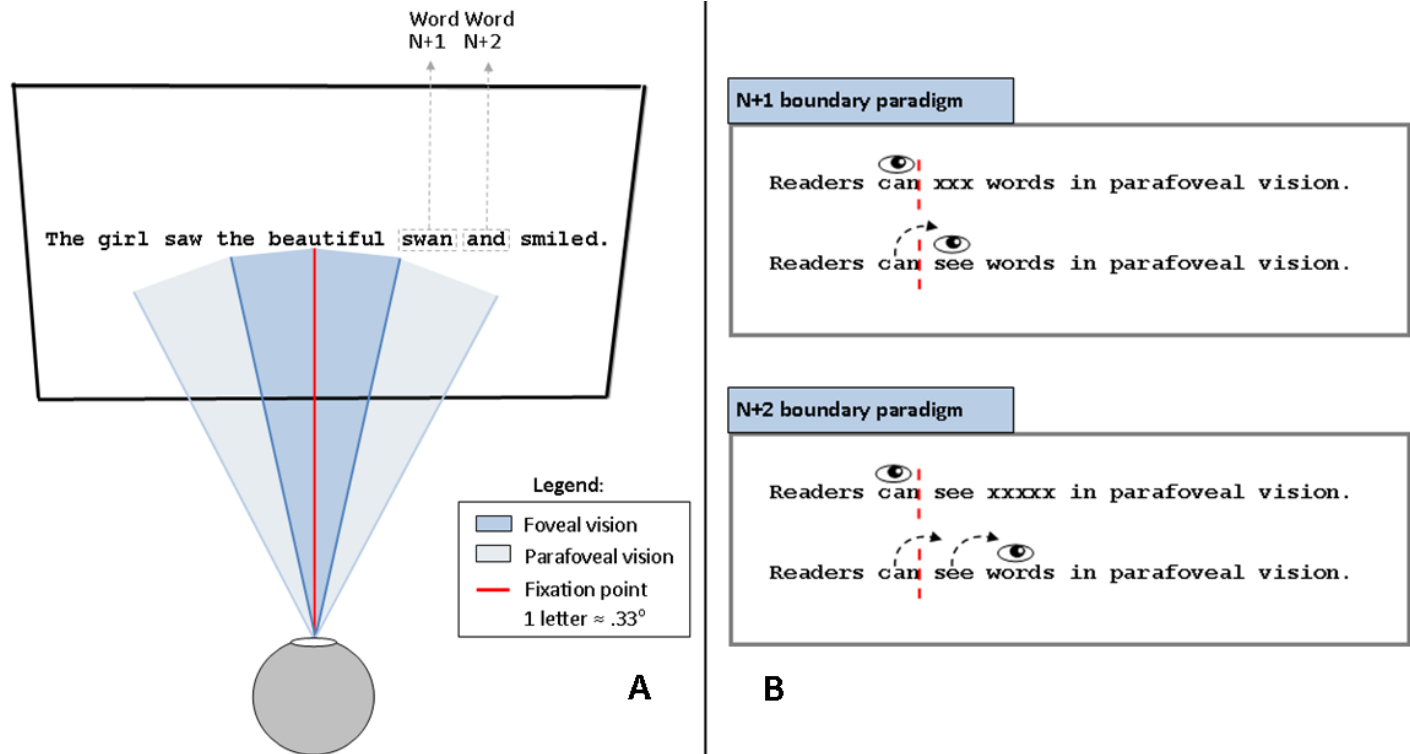


Figure 1. An illustration of the perceptual span during reading (1A). A demonstration of the gaze-contingent boundary paradigm (Rayner, 1975) for studying N+1 and N+2 preview effects (1B). The invisible boundary is represented by a red vertical line. Panel B illustrates only the invalid preview condition; during valid preview, no change occurs on the screen.

One paradigm that has been widely used to study parafoveal processing during reading is the gaze-contingent boundary technique (Rayner, 1975). In this technique, an invisible boundary is placed before the target word in order to manipulate what participants see before the target word is fixated in foveal vision (see Figure 1B for an illustration). In the control condition, participants view the actual target word in parafoveal vision (this will be referred to as *valid preview* from now on). In the experimental condition, the target word is masked (e.g. with a string of Xs) and participants do not acquire any useful information from it prior to crossing the boundary (this will be referred to as *invalid preview*). Once the eyes cross the invisible boundary,

the parafoveal mask is permanently replaced with the target word (during valid preview, no physical change occurs on the monitor, i.e. the target word is replaced with itself). By using this manipulation, it is possible to calculate how much parafoveal information participants obtain from the word to the right of the current fixation by subtracting fixation durations during valid preview from fixation durations during invalid preview. In this paper, the standard terminology is adopted, where the word before the boundary is referred to as word N, the first word following the boundary is referred to as word N+1, and the second word following the boundary is referred to as word N+2. As shown in the lower panel of Figure 1B, the gaze-contingent display change manipulation can also be applied to word N+2.

The Preview Benefit Effect: A Re-examination

The advantage of previewing word N+1 in parafoveal vision is known as the preview benefit effect (Rayner, 1998) and is often thought to reflect the benefit in terms of word recognition times: larger preview benefits are interpreted to mean faster recognition times once word N+1 is fixated in foveal vision. However, the exact size of this effect is not known- in part because it varies from study to study, and in part because it hasn't been systematically investigated. Previous literature reviews have estimated that the preview benefit effect is between 20-50 ms (Rayner, 2009; Rayner, White, Kambe, Miller, & Liversedge, 2003). However, there are a number of issues with such estimates. First, they are not derived in ways that are sufficiently well-documented. Second, they are not based on all the available evidence and are thus susceptible to researchers' intuition and their own experience with this type of research. Third, due to the fact that a number of different types of fixation duration measures are used, the effect size for each individual measure is also not known with certainty.

There are few standard fixation duration measures that are used for calculating the

preview benefit effect. First fixation duration (FFD) measures the very first fixation on the target word. Single fixation duration (SFD) is similar to FFD, but it reflects cases when the target word was fixated only once. Another measure is gaze duration (GD), which measures all fixations on the target word before the eyes move to another word. These three measures are often referred to as first-pass reading. In addition to them, second-pass reading measures are also sometimes used. The most common one is total viewing time (TVT), which measures all fixations on the target word (including when the word is re-fixated during a regression).

The use of multiple fixation duration measures is something that has not been seriously considered until very recently (von der Malsburg, & Angele, 2015). One issue with this practice is that there is a high degree of co-linearity between these measures. This means that the preview benefit of one measure already contains all the information that is computed with other measures. For example, as all the fixations that contribute to FFD also contribute to GD and TVT, this has the consequence that, in practice, the preview benefit effect is additive and that both the effect size and the general magnitude of fixation time measures are virtually always related in the following way: $FFD \leq SFD \leq GD \leq TVT$. Due to the fact that the effect size for each measure is not known with certainty, not much can be said about how the effect is distributed across the four measures (e.g. is it constrained mostly to the first fixation? Do GD and TVT equally add to it?). A precise estimate of each measure is useful not only for planning experiments (e.g. sample size calculations), but also for formulating more precise hypotheses and for interpreting the results of boundary studies. For example, given that it is often not clear how to interpret an effect found in one measure but not the other measures, a more precise estimate of the effect will make it possible to make more precise hypotheses about the expected effects of boundary experiments.

Another aspect of the preview benefit effect that has not been addressed in a satisfactory way is whether it is similar in size across languages. Research in this area has a long-standing bias in the sense that most studies are conducted in a few western languages, such as English, German and Finnish. Fortunately, this has changed a bit in recent years, and studies are now being published in languages such as Chinese, Thai, and Korean, just to name a few (see Radach & Kennedy, 2013 for a recent discussion of this). Some languages, such as Chinese, use different writing systems. In this sense, the preview benefit effect may be different in non-alphabetical languages.

In fact, one common assumption is that the preview benefit effect is bigger in Chinese because information is more densely packed and there are no inter-word spaces (e.g. see Yang, Wang, Tong, & Rayner, 2012b). This, however, has not been addressed empirically, and it is not immediately clear if and to what extent the preview benefit effect differs between alphabetical languages and Chinese. This question holds great theoretical implications, because parafoveal processing plays a key role in computational models such as the E-Z Reader (Reichle et al., 1998) and SWIFT (Engbert et al., 2005) that were originally developed on alphabetical languages. Finally, there is no good estimate of the size of the preview benefit effect that is based on all languages that have been studied with the boundary paradigm.

Preview Benefit vs. N+1 Preview Effects

As mentioned previously, the preview benefit has traditionally been viewed as the advantage of previewing the target word in terms of word recognition times. However, there are two recent lines of evidence that suggest this may be an oversimplification. The first one is the finding that parafoveal masks that are used during invalid preview conditions may introduce processing costs (Gagl, Hawelka, Richlan, Schuster, & Hutzler, 2014; Hutzler et al., 2013;

Kliegl, Hohenstein, Yan, & McDonald, 2013; Marx, Hawelka, Schuster, & Hutzler, 2015; Yan, Risse, Zhou, & Kliegl, 2012). This evidence suggests that the preview benefit effect is a combination of both preview benefits and preview costs. Because the effect is calculated by subtracting fixation durations after valid preview from fixation durations after invalid preview, any processing costs that inflate the latter will also result in an inflated preview benefit effect.

In alphabetical languages at least, there are a number of ways in which the target word can be masked in parafoveal vision. For example it can be replaced with Xs (*cottage* → *xxxxxxx*), random letters (*cottage* → *fdiuekl*), an unrelated word (*cottage* → *kitchen*) or a pseudo-word (*cottage* → *oxypane*). Therefore, there are a number of baselines that could be used for calculating the preview benefit effect. The problem of what constitutes a proper baseline is not unique to eye-tracking studies of reading, but is also relevant to priming studies more broadly (e.g. see Jacobs, Grainger, & Ferrand, 1995). To date, the extent to which different types of parafoveal masks introduce processing costs is poorly understood. In one study, Hutzler et al. (2013) investigated the effect of parafoveal X masks on reading lists of words. By using fixation-related brain potentials, they found that such masks interfere with foveal word recognition. More recently, Marx et al. (2015) also presented evidence that parafoveal X masks and random letter masks overestimate the preview benefit effect compared to a more neutral baseline where the target word is visually degraded. The search for a new and more neutral mask is indeed a commendable effort. However, regardless of whether a mask that yields a “pure” preview benefit without any interference is found in the future (e.g. visual degradation; Marx et al, 2015), the processing costs associated with each of the masks outlined above still haven’t been quantified. Since the boundary technique has been used for the last 40 years, this question is very pertinent as there are dozens, if not hundreds, of boundary experiments that have used such masks.

A second line of evidence also suggests that the preview benefit effect may not reflect the pure advantage of having a valid preview of the target word. More specifically, Risse and colleagues (Risse & Kliegl, 2012, 2014; Risse & Vasilev, in prep.) manipulated both the validity of the word in parafoveal vision (valid vs invalid) and its difficulty (high vs low). The difficulty was operationalised as the lexical frequency of the target word. They found that what is traditionally known as the preview benefit effect is in fact a combination of both the difficulty and the validity of the parafoveal preview. In other words, the preview validity and preview difficulty are two separate effects that, when combined, make up the preview benefit effect. Therefore, even though readers benefit from having a valid preview of the target word, the nature of this preview is also influenced by the lexical difficulty of the word in parafoveal vision. In this sense, these results suggest that the preview benefit may be inherently confounded with preview frequency effects.

In the light of the evidence reviewed above, it can be argued that the preview benefit may not necessarily be the pure advantage of having a valid preview of the target word in parafoveal vision. For this reason, the present paper purposefully avoids talking about a preview benefit effect. Instead, we will use the slightly more conservative term *N+1 preview effect*, which denotes the effect resulting from manipulating the preview of word N+1 in parafoveal vision. Even though there is no doubt that readers do obtain a net benefit from previewing word N+1 in parafoveal vision, at present it is not known what contributes to this benefit. For example, it could be formed of a large benefit effect moderated by large mask interference effects, or it could be a smaller benefit effect that isn't reduced much by interference. For this reason, we prefer to remain agnostic on this issue until more evidence is available (nevertheless, we will attempt to address this question in the empirical part of this paper). The issue of whether and to what extent

interference effects exist is also theoretically important for computational models of reading. For example, models such as the E-Z Reader (Reichle et al., 1998) and SWIFT (Engbert et al., 2005) assume that parafoveal processing is only benefit and no interference. In this sense, the issue is also important for future model development.

To summarise the discussion so far, it is well known that readers obtain an advantage from previewing word N+1 in parafoveal vision. However, the size of this effect for different measures is less clear. For example, there is a great deal of uncertainty about how the effect differs between alphabetical and Chinese studies. Also, the consequences of using different parafoveal masks are not well-understood. These are all issues that we will return to in the empirical part of the paper. Before that, however, we will also consider the evidence from N+2 preview effects.

N+2 Preview Effects During Reading

Originally, research on parafoveal preview effects focused on word N+1. Interest in N+2 preview effects is more recent and was primarily motivated by the possibility of testing competing predictions of computational models of eye-movement control during reading (Rayner, Juhasz, & Brown, 2007). While N+2 preview effects are expected in guidance-by-attentional-gradient (GAG) models such as SWIFT (Engbert et al., 2005), they are generally not predicted by serial-attention shift (SAS) models such as the E-Z Reader (Reichle et al., 1998). More recent simulations with SWIFT have confirmed that the model is able to simulate N+2 preview effects without directly fitting it to such effects (Risse, Hohenstein, Kliegl, & Engbert, 2014). Interestingly, simulations with the E-Z Reader have also demonstrated that N+2 preview effects are theoretically possible, although they are likely to be small in size (Schotter, Reichle, & Rayner, 2014). However, it is currently not clear how well these simulations fit with the

empirical data due to the mixed findings in the literature. As shown in Figure 1, if word N+1 is short, then the next word (N+2) can also be previewed in parafoveal vision prior to crossing the boundary. However, even though N+2 preview effects have been investigated for the last 10 years, these studies have yielded inconsistent findings. As a result of this, there is currently no clear consensus on whether any processing of word N+2 occurs in parafoveal vision.

Judgments about the existence of such effects have usually been made by comparing statistical significance across studies. However, this practice is problematic because it can result in misleading conclusions (Gelman & Stern, 2006). The use of *p*-values also does not answer the question of real interest- how likely is it that readers obtain an advantage from previewing word N+2? If these studies are examined even more closely, there are other statistical issues that may further obscure the underlying effect. Because word N+2 is situated closer to the limits of parafoveal vision compared to word N+1, visual acuity there should by definition be lower. For this reason, N+2 preview effects, if they exist, should almost certainly be smaller in size than N+1 effects. Even though it is well known that larger sample sizes are needed to achieve sufficient statistical power with smaller effect sizes, this issue has not always been considered in planning studies investigating N+2 preview effects.

Another issue that may create further confusion is the fact that some N+2 studies have also included a N+1 boundary in the same experiment (Rayner et al., 2007; Yang, Rayner, Li, & Wang, 2012a; Yang, Wang, Xu, & Rayner, 2009). The general reasoning behind this was to see if a dissociation can be found between parafoveal preview and boundary location. Unfortunately, such dissociation can be found only by analysing the interaction between boundary location and parafoveal preview, and not by comparing the statistical significance of the two preview effects (see Nieuwenhuis, Forstmann, & Wagenmakers, 2011). As not all experiments have reported

such an interaction, N+2 preview effects at the very least warrant further investigation. The use of two boundary locations in the same experiment may also introduce additional artefacts that may not be found in experiments with a single (N+2) boundary¹. For example, participants may be more likely to detect display changes when there are two types of boundary manipulation in the same experiment. Since preview effects from word N+1 are well established and since the comparison of statistical significance of the two effects is generally not very informative, it may be worthwhile to use only one boundary location in future N+2 studies.

The Present Study

The main goal of the present study was to re-examine N+1 and N+2 preview effects by undertaking a systematic statistical synthesis of previous findings. There were four main questions of interest: 1) what is the size of N+1 and N+2 preview effects?; 2) what is the probability that N+2 preview effects exist given the evidence to date?; 3) Do N+1 and N+2 preview effects differ between alphabetical and Chinese studies?; 4) are N+1 preview effects influenced by the type of parafoveal mask?

Although it may seem that determining the size of the N+1 preview effect is not theoretically important, estimating this effect is crucial, among other things, for evaluating the magnitude of the N+2 effect. Because word N+1 and word N+2 are two points on the continuum of decreasing visual acuity, N+2 preview effects can be meaningfully compared only in relation to the N+1 preview effect. However, as a precise estimate of the latter has not yet been systematically derived, determining the size of this effect was also important. In addition to that, the two effect sizes are very useful for evaluating the theoretical claims and simulations of computational models of reading.

¹ I thank Sarah Risse for bringing this issue to my attention.

Motivation for the meta-analysis. In order to fully answer these questions, the present meta-analysis adopted a Bayesian approach to statistical inference. Although this method of inference is less commonly known in psychology, it confers a number of advantages for the present meta-analysis. First, Bayesian statistics offers an intuitive and rigorous way of handling uncertainty that is based on the laws of probability. Second, the results of the analysis are conditional on the data and thus can be used to calculate the probability of preview effects given the data, i.e. $P(\text{Effect} \mid \text{Data})$. This is in stark contrast to the frequentist approach to statistical inference, which yields the less informative probability of obtaining the data given the effect, i.e. $P(\text{Data} \mid \text{Effect})$. Finally, the results of Bayesian inference, which are expressed as posterior probability distributions, offer much richer information about parafoveal preview effects (e.g. what is the probability that N+2 preview effects are bigger than 5 ms?). Posterior distributions also avoid the use of p -values, which are often misinterpreted as providing information that they do not (Perezgonzalez, 2015). Because null hypothesis significance testing is a mishmash of two incompatible theories, p -values are often ascribed statistical properties that they do not actually possess (for more details, see Kline, 2004).

Although Bayesian statistics is less familiar to psychology researchers, it has recently increased in popularity (Andrews & Baguley, 2013). In light of the well-known limitations of null hypothesis significance testing (Cohen, 1994; Cumming, 2013; Gigerenzer, 2004; Royall, 1997), there has been a call in recent years for the use of Bayesian inference in psychology (e.g., Andraszewicz et al., 2014; Kruschke, 2010; Kruschke & Liddell, 2015; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, 2007; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). Bayesian approaches to meta-analysis have traditionally been used in the biomedical

sciences (e.g. Higgins & Spiegelhalter, 2002; Salpeter, Cheng, Thabane, Buckley, & Salpeter, 2009; see Sutton and Abrams, 2001 for a review), but more recently also in psychology (Rouder & Morey, 2011; Rouder, Morey, & Province, 2013) and psycholinguistics (Engelmann, Jäger, & Vasishth, 2016; Vasishth, Chen, Li, & Guo, 2013). For example, Vasishth et al. (2013) used a Bayesian random-effects meta-analysis to investigate conflicting evidence about the processing of relative clauses in Chinese. Their results showed that the probability that subject-relative clauses are easier to process than object-relative clauses, given the data, is about 80%.

The present meta-analysis is similar in spirit to the one of Vasishth et al. (2013) and it also adopted a random-effects meta-analytical model. This particular type of model was chosen because it assumes that the observed effect sizes vary around a true unknown effect. This makes it possible to model the variability associated with the use of different languages, writing systems, as well as variability due to different experimental designs, lab practices and so forth. The present meta-analysis was mostly exploratory in nature. However, based on the literature reviewed above, it was expected that: 1) N+1 preview effects are somewhat smaller (e.g. around 20-30ms) than what is sometimes assumed; and 2) that the posterior distribution of N+2 preview effects was not centered on 0.

Method

The meta-analysis was conducted by following the guidelines of the PRISMA statement (Moher, Liberati, Tetzlaff, Altman, & the PRISMA group, 2009). The graphical presentation of results was done by following the suggestions of Anzures-Cabrera and Higgins (2010).

Literature Search

The search of the literature was done separately for N+1 and N+2 studies. Figure 2 shows a flowchart of the process. Google Scholar and Web of Science were searched with “word n+2

preview” and “n+2 preview reading” as keywords for N+2 studies, and “word n+1 preview”, “preview benefit reading”, and “parafoveal preview reading” as keywords for N+1 studies. This was done in September 2015. The search process was completed in four stages: 1) recording of all relevant papers in a database; 2) removing of duplicate entries; 3) screening of full-text papers for experiments with a N+1/N+2 boundary manipulation; 4) full-text reading of the papers identified in the previous stage and coding them if they met the inclusion criteria. During the fourth stage, care was also taken to identify cited papers that were missed in the literature search and that may be relevant. Such papers (N=18) were then reviewed using the same process.

The full study inclusion criteria can be found in Appendix A. In short, the experiments had to be methodologically sound and not deviate too much from conditions that are typical for everyday reading. As a result of this, 44.2% of all reviewed N+1 experiments and 64.7% of all N+2 experiments were included in the meta-analysis. Although this may seem like a high exclusion rate, it was necessary to ensure that only studies that have similar parafoveal masks and experimental design are analysed together.

Publication bias. Publication bias is a threat to meta-analyses because the literature consists mostly of studies that report positive findings (Vasilev, 2013) and negative results are less likely to be written up and submitted for publication (Franco, Malhotra, & Simonovits, 2014). In this meta-analysis, two unpublished (at the time of analysis) experiments were included². Furthermore, funnel plots were used to assess the data for publication and other related biases (Sterne et al., 2011; see Appendix B). On the whole, there was no clear evidence of systematic bias in the data, while keeping in mind that a small degree of heterogeneity was expected for reasons discussed below.

² One unpublished N+2 study was identified, but details could not be obtained from the authors.

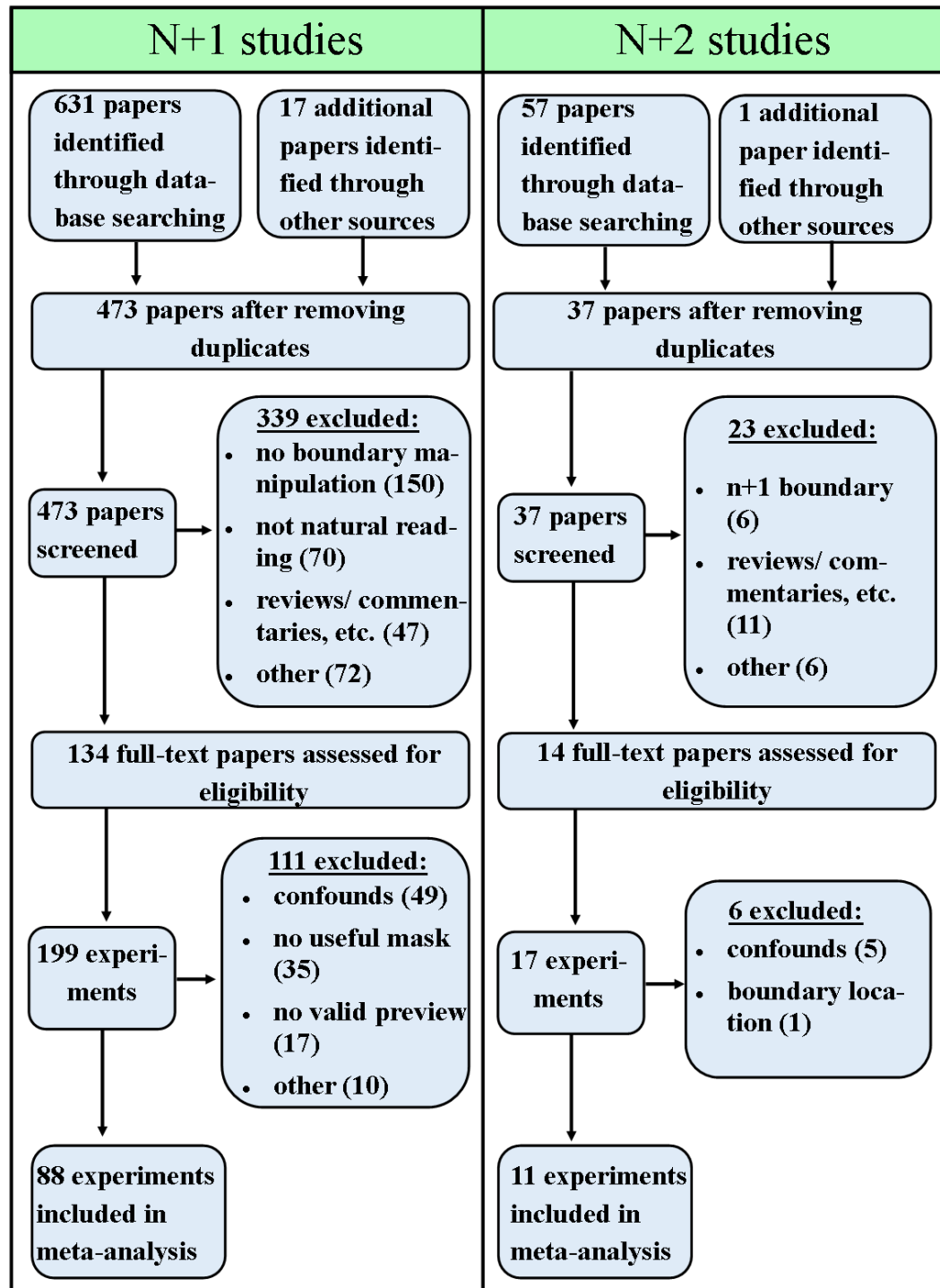


Figure 2. Flowchart of the literature search process for N+1 and N+2 studies. The breakdown of excluded papers is only approximate because some of them were excluded for more than one reason. Six of the included experiments had both a N+1 and N+2 boundary.

Data Coding

For all experiments included in the meta-analysis, the following type of information was extracted: mean fixation durations (FFD, SFD, GD, or TVT), their respective standard deviations, the sample size and the language of the study (a summary of this information is available in Supplementary files #1). This information was further used to combine the experiments into different groups for analysis. For studies with a N+1 boundary, descriptive statistics were coded for the following types of parafoveal masks: random string of letters, a string of Xs, an unrelated word, a pseudo-word, an orthographically-related word/nonword, a semantically-related word, or a phonologically-related word/nonword³. The pseudo-word mask was defined as a nonword that is pronounceable. The orthographical mask was defined as a word/nonword that has at least 50% letter overlap with the target word, regardless of the position of these overlapping letters (in all Chinese studies, the orthographical mask was classified as such by the original authors). Since there was no easy way to quantify the degree of semantic and phonological relatedness of the mask and the target word, the classification of the original authors was taken at face value. Finally, due to the much smaller number of N+2 studies, the effect of the mask type could not be addressed in any meaningful way. Whenever two masks were available, the one that best preserves the consistency across studies was chosen⁴. For example, if experiment one had only unrelated word mask, but experiment two had both unrelated word and pseudo-word masks, then the unrelated word mask was coded in both experiments.

³ In the few cases where there was ambiguity about the type of mask (e.g. unrelated word vs pseudo word), the first author was proficient in the language and could make this judgement. The only exception was studies done in Chinese, for which a native speaker was consulted.

⁴ To test for subjectivity in this decision, the N+2 analyses were repeated by coding the opposite mask to the one that was chosen. This resulted in a mean difference of 0.6 ms across the effects sizes and it did not affect the conclusions from this paper.

Missing data. All N+2 studies reported the full data needed for the analysis. However, 27% of all N+1 studies did not report a measure of variance. This was handled in the following way. For papers published in the last 10 years, the authors were contacted and asked if they could provide the missing data (8 % of all the missing data were recovered in this way). In all other cases, the missing values were imputed by pooling SDs from the remaining studies. This approach has been shown to produce accurate results (Furukawa, Barbui, Cipriani, Brambilla, & Watanabe, 2006). Also, it has a number of advantages over simply excluding such studies (e.g. it increases the sample size and it does not introduce study-selection biases; see Lajeunesse, 2013).

Data Analysis

A random-effects model was chosen because some heterogeneity between studies was expected *a priori* due to the different types of languages (alphabetical vs Chinese). It should be noted, however, that a random-effects model does not account for the heterogeneity per se. Rather, it no longer matters as the model already assumes that such heterogeneity exists (Higgins & Green, 2011).

Meta-analysis. A Bayesian random-effects meta-analysis was defined as follows (Engelmann et al., 2016; Schmid & Mengersen, 2013). Let T_i be the observed effect size (in milliseconds) for study i , where i ranges from 1 to n (the number of experiments in the analysis). Positive values represent evidence for a preview effect and negative values represent evidence against it. Let θ be the true unknown preview effect that is to be estimated by the model. Furthermore, let θ_i be the true preview effect in the i -th study, and σ_i^2 the true variance of the sampling distribution in the i -th study. Finally, let τ^2 be the unknown between-study variance. Therefore, the model was constructed as:

$$T_i | \theta_i, \sigma_i^2 \sim N(\theta_i, \sigma_i^2) \quad i = 1, 2, 3, \dots, n$$

$$\theta_i | \theta, \tau^2 \sim N(\theta, \tau^2), \quad (1)$$

$$\theta \sim Uniform(-200, 200),$$

$$\tau \sim Uniform(0, 200)$$

Each σ_i was estimated from the standard error of study i . In a within-study design such as the experiments in the present meta-analysis, this is calculated in the following way (Borenstein, Hedges, & Higgins, 2009, p.24). If S_1 is the standard deviation of the valid preview condition and S_2 is the standard deviation of the invalid preview condition, the variance of the mean difference T , V_T , is calculated as:

$$V_T = \frac{S_{diff}^2}{n} \quad (2)$$

where n is the sample size and S_{diff} is given by:

$$S_{diff} = \sqrt{S_1^2 + S_2^2 - 2 \times r \times S_1 \times S_2} \quad (3)$$

where $r \times S_1 \times S_2$ denotes the covariance of the two means.

The standard error is then given by:

$$SE_T = \sqrt{V_T} \quad (4)$$

In equation 3, it was not possible to estimate the covariance of the two means due to the fact that we didn't have access to the raw data. Therefore, the calculations assume that the covariance is equal to 0. It should be noted that this has the consequence that the variance parameter is likely overestimated in the analyses. In the meta-analysis, precision was defined as the inverse of the within-study variance of the sampling distribution (i.e. $1 / SE_T^2$). This means that

studies with smaller variance are given greater weight in the analysis. The same was done for the between-study variance τ^2 , as the between-study variance is added to the within-study variance of the sampling distribution in calculating the weight of each study.

Uniform distributions were used as priors in the analysis. These non-informative prior distributions make each value on these intervals equally likely, that is, they do not contain any meaningful assumptions about the size of preview effects. In this way, maximising the posterior probability is equivalent to maximising the likelihood (Edwards, 1974). In other words, the data are allowed to “speak for themselves” and the prior distributions have very little to no influence on the results.

In order to check whether the chosen priors influence the results, sensitivity analyses were also conducted by using different priors (the results are reported in Supplementary files #3). The additional priors used were Gamma(0.001, 0.001) and Normal(0, 100²) I(0,) for τ , and Normal(0, 100²) for θ (note that the Gamma prior was on the precision of τ). The results of the sensitivity analysis showed that the choice of priors had very little effect on the results and had practically no influence on the conclusions from the analyses. The only notable finding was that the Gamma prior tended to underestimate τ (cf. Gelman, 2006), but even that did not result in a dramatic difference in the results.

Meta-regression. As mentioned above, a random-effects meta-analysis can account for the heterogeneity between studies, but it does not help explain what gives rise to this heterogeneity in the first place (Welton, Sutton, & Cooper, 2012). However, it is possible to use random-effects meta-regression to investigate how categorical study characteristics (e.g. whether the language was alphabetical or Chinese) are associated with the effect of interest (Deeks, Higgins, & Altman, 2008). In this case, the model in equation 1 is extended by adding a regression coefficient β for

the underlying effect of the covariate of interest (Welton et al., 2012). Therefore, the meta-regression model was constructed as follows (the terms in bold are what is different from equation 1; Engelmann et al., 2016; Welton et al., 2012):

$$\begin{aligned}
 T_i | \theta_i, \boldsymbol{\beta}, \sigma_i^2 &\sim N(\theta_i + \boldsymbol{\beta} \mathbf{x} \text{ covariate}_i, \sigma_i^2) \quad i = 1, 2, 3, \dots, n \\
 \theta_i | \theta, \tau^2 &\sim N(\theta, \tau^2), \\
 \boldsymbol{\beta} &\sim \text{Uniform}(-200, 200), \\
 \theta &\sim \text{Uniform}(-200, 200), \\
 \tau &\sim \text{Uniform}(0, 200)
 \end{aligned} \tag{5}$$

In this model, a prior on β is also needed. For consistency purposes, and since the possible range of values for β is constrained, we also used a Uniform prior. However, sensitivity analysis with a Normal(0, 100²) prior on β showed that the results were not influenced by this decision (see Supplementary files #3). In this meta-regression model, θ_i is the effect in the i -th study that is adjusted for the covariate effect β ; therefore, the effect of the covariate β becomes the parameter of interest, rather than θ (Engelmann et al., 2016).

Meta-regression was used to investigate two of the main questions identified in the introduction: 1) whether effect sizes differ as a function of the type of language (alphabetical vs Chinese); and 2) whether the N+1 effect size differs between the different parafoveal masks. For the first question, a sum contrast was used where alphabetical languages were coded as -1 and Chinese studies as 1. Therefore, positive values of β indicate that preview effects are larger for Chinese compared to alphabetical languages. For the second question, the same type of contrasts were set up post-hoc after inspecting the effects sizes of different masks; the types of

comparisons are shown in the results section for simplicity. It is important to note, however, that the conclusions from meta-regression are only observational in nature (Thompson & Higgins, 2002). Therefore, the meta-regression results should be considered as exploratory and they need to be verified by future experiments.

Posterior sampling. Sampling from the posterior distribution was done with JAGS (Plummer, 2003) using the R software, v. 3.1.3 (R Core Team, 2015). Three MCMC chains were run with 75,000 iterations each. Checks were performed to ensure that the initial values did not influence the results. The first 3,000 iterations were discarded as burn-in. Chain thinning of 5 was used to reduce the influence of auto-correlation. Convergence was assessed with visual inspection and Gelman and Rubin's (1992) convergence diagnostic. The evidence suggested that all models had converged (for details on MCMC sampling, see Lynch, 2007).

Types of analyses. In the main analysis, we fitted separate models for N+1 and N+2 preview effects. For N+2 effects, all 11 studies were included in the analysis. For N+1 studies, only experiments containing a random letters, unrelated word, X-string or pseudo-word mask were included. This was done because such masks do not contain any useful information about the target word. These effects found in these two analyses will be referred to as N+1 and N+2 preview effects, respectively. In all analyses, separate models were fit for each dependent variable.

In addition to these general analyses, we did a separate breakdown of N+1 preview effects by parafoveal mask. The minimum number of studies needed to fit a model was set at six. The purpose of this breakdown was to investigate the effect sizes associated with each mask type. Separate models were again fit for each fixation duration measure when at least six studies were available. Additionally, whenever enough experiments were available, all analyses were completed once for all studies, once for alphabetical-only studies and once for Chinese studies.

Finally, it should be kept in mind that all probabilities reported in this paper are *not* p-values and cannot be interpreted as such; rather, they are the posterior probability of parafoveal preview effects given the data. For a gentle introduction to using and interpreting the results of Bayesian inference, readers are referred to Nicenboim and Vasishth (2016) (see also Kruschke, 2014).

Results

Baseline Fixation Time Measures

First, we analysed the fixation time measures (FFD, SFD, GD, and TVT) on the target word in the valid preview condition. In gaze-contingent display change experiments, this condition provides the baseline that the experimental preview conditions are compared against. The posterior distribution in this analysis gives the fixation time given all the available data. The results of this analysis are presented in Table 1. The purpose of this analysis was to calculate the average duration of each fixation time measure when parafoveal preview is allowed. As the valid preview condition corresponds to normal reading, the results can also be used as benchmark data for designing experiments or computational models of reading.

For this analysis, nine additional experiments were added. These experiments were not included in the meta-analysis of parafoveal preview effects because they did not have a mask that could be added to one of the mask types that were analysed in this paper. However, as these studies did have a valid preview condition, they are suitable for this baseline analysis.

Table 1

Posterior Distribution Mean and 95% Credible Intervals for Fixation Time Measures on the Target Word when Parafoveal Preview Was Allowed (i.e. in the Valid Preview Condition)

Measure	All studies			Alphabetical studies			Chinese studies		
	N	Mean	95% CrI	N	Mean	95% CrI	N	Mean	95% CrI
FFD	91	247.2	[242.5, 251.9]	76	246.2	[240.9, 251.5]	15	252.3	[241.6, 262.8]
SFD	43	250.4	[242.9, 257.9]	35	248.8	[239.6, 258.1]	8	256.7	[248.8, 265.1]
GD	96	284.2	[276.8, 291.7]	78	280.4	[273.5, 287.4]	18	300.2	[272.2, 328.2]
TVT	28	333.9	[314.2, 353.7]	28	334	[314.3, 353.6]	0	N/A	N/A

Note: Only N+1 studies are included in the analysis. N: number of experiments on which the analysis is based. Mean: posterior mean. CrI: credible intervals.

N+2 Preview Effects

Means and 95% credible intervals for the posterior distributions of the fixation time measures for N+2 preview effects are presented in Figure 3. When all the studies are combined, the posterior distributions of FFD and GD showed a 5 ms preview effect compared to the valid preview baseline. When only the alphabetical studies were analysed, the estimates of the two effects were slightly smaller and less certain. Due to the fact that some previous studies (e.g. Kliegl, Risse, & Laubrock, 2007) have reported N+2 preview effects on word N+1, we also ran a separate analysis with the 8 studies that reported descriptive statistics for word N+1. The mean N+2 preview effect on word N+1 was 6.7 ms for FFD [95% Credible interval: -4, 17.2 ms] and 7.4 ms for GD [95% Credible interval: -5.2, 20.2 ms].

When investigating N+2 preview effects, the length of word N+1 is an important factor because it determines whether and to what extent word N+2 can be previewed in parafoveal

vision. Therefore, it can be argued that, for alphabetical languages, N+2 preview effects can be reliably found only when word N+1 is short (e.g. 3-4 letters). Ideally, one would include word N+1 length of alphabetical languages as a covariate in a meta-regression model. However, in the present analysis this was not feasible because there were only 7 alphabetical studies and meta-regression is not recommended when there are less than 10 studies (Deeks et al., 2008).

However, the length of word N+1 for all studies is included in Figure 3 for visual comparison. Also, we fit a separate meta-analytical model for the 5 alphabetical studies that had a word N+1 length of 3 letters. The expectation that N+2 preview effects will be larger when word N+1 is 3-letter long was confirmed for GD (mean effect size: 6.9 ms [95% Credible interval: -19.7, 33.4]), but not for FFD (mean effect size: 2.3 ms [95% Credible interval: -17.2, 21.8]). Therefore, the results suggest that the effect size for GD is larger when word N+1 is 3 letters long, but it should be kept in mind that this finding could be influenced by the small sample size.

In order to take full advantage of the Bayesian approach in the present meta-analysis, we also calculated the probability that N+2 preview effects are bigger than 1 ms. The threshold of 1 ms was taken as a purely statistical measure of the minimum effect size anyone would sensibly consider to indicate that N+2 preview effects exist. Of course, researchers may have their own beliefs about what threshold can be used to determine if N+2 preview effects exist. For example, a sceptical researcher may believe that only effect sizes greater than 4 ms can be taken as evidence that N+2 preview effects are real. Such beliefs can be easily accommodated within the Bayesian framework, as researchers can simply look up the probabilities associated with their beliefs. This information is presented in Figure 4.

When all studies are combined, the probability that N+2 preview effects on word N+2 are bigger than 1 ms was 0.87 and 0.77 for FFD and GD, respectively. When only alphabetical

studies were analysed, these probabilities were 0.70 for FFD and 0.53 for GD. The probability that N+2 preview effects on word N+1 are bigger than 1 ms was 0.85 for both FFD and GD.

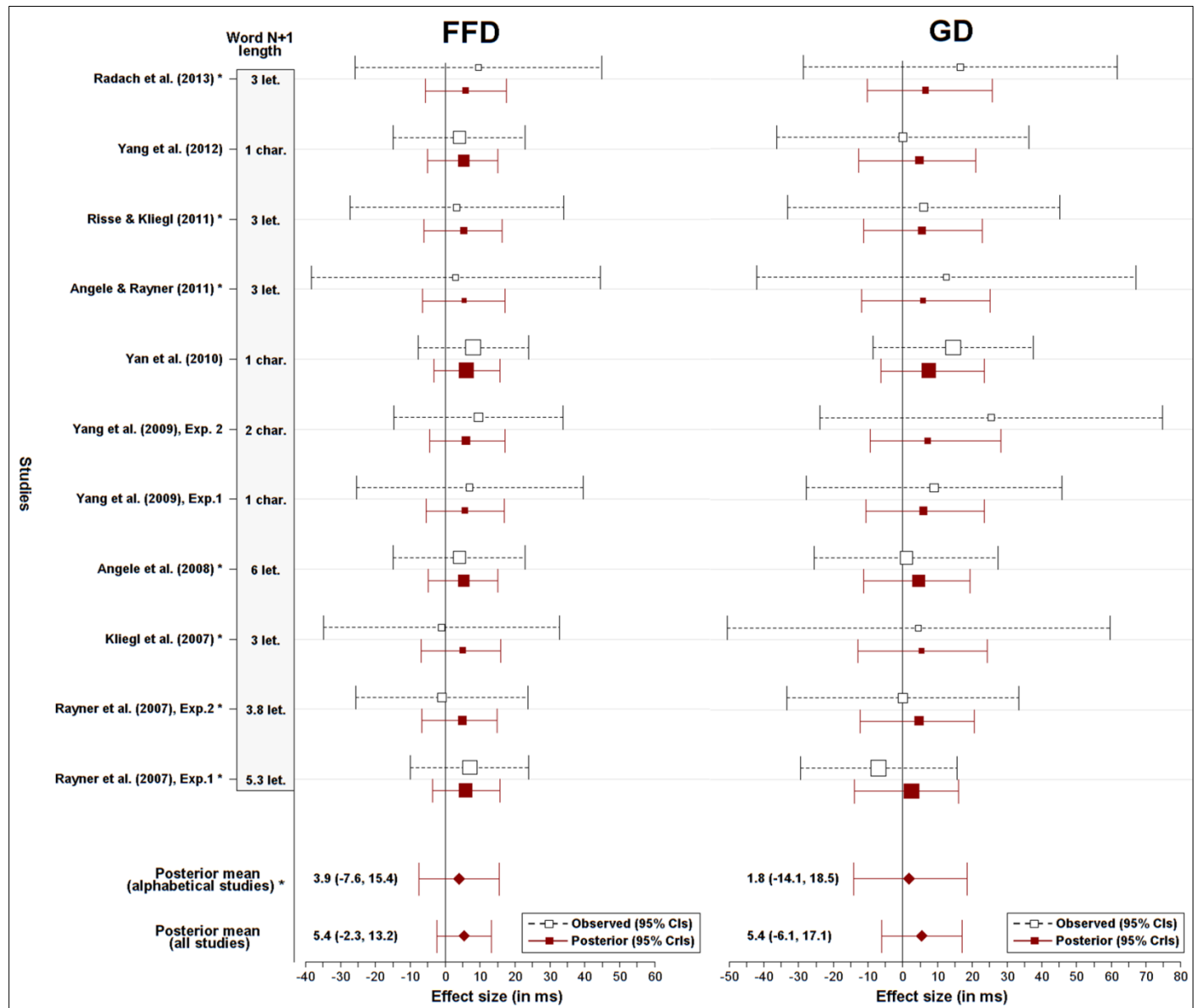


Figure 3. A forest plot of N+2 preview effects on word N+2. Plotted are the observed effect sizes with 95% confidence intervals (CIs) and the posterior estimate of the effect sizes with 95% credible intervals (CrIs). The size of squares is proportional to the weight of the study (i.e. the inverse of the within-study variance of the sampling distribution). Results for individual studies

Given that the credible intervals for both FFD and GD included 0, one could reasonably ask whether this could indicate a null effect. The fact that the intervals include 0 means that this number cannot be rejected as a credible value for N+2 preview effects. However, this also does not prove that the effect is null. To prove a null effect, it would be necessary to show that the 95% credible interval is contained within the null region that was defined in the present analysis (from -1 to 1 ms; cf. Kruschke, 2014). In this sense, a null effect can only be accepted or rejected when the posterior estimate is sufficiently precise.

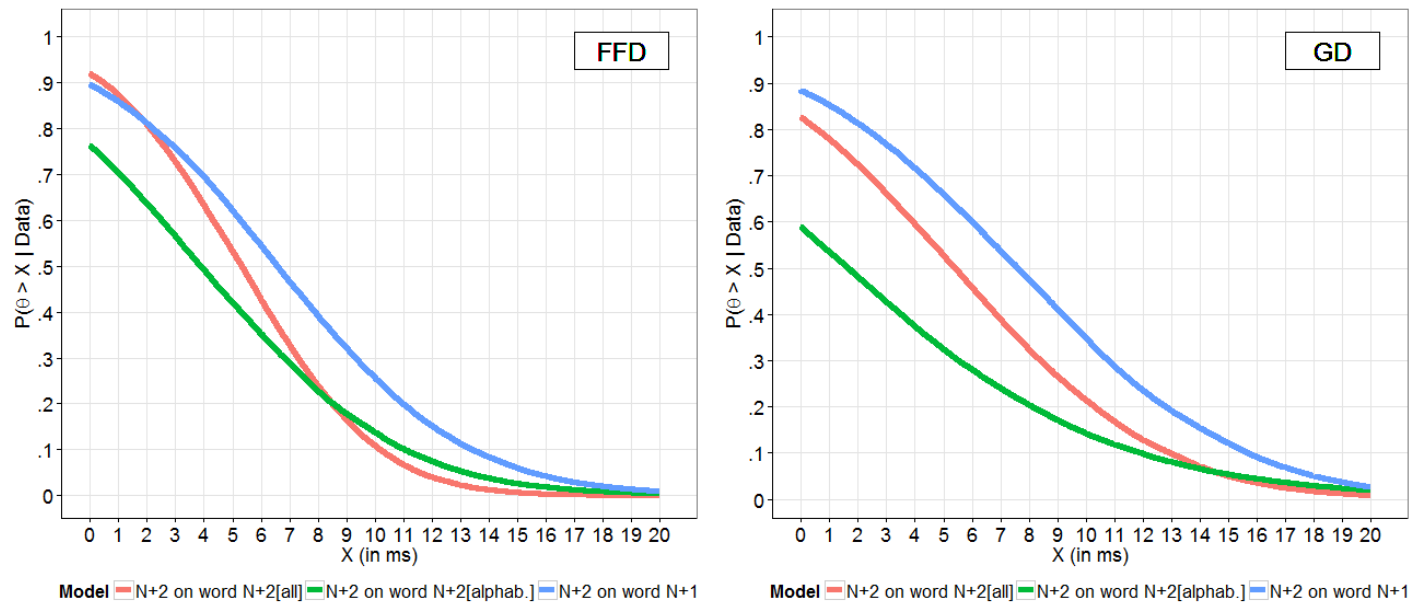


Figure 4. Probability that N+2 preview effects are greater than some number X, given the data.

All: all studies; alphab: alphabetical studies.

Finally, we used meta-regression to investigate whether N+2 preview effects on word N+2 are greater for Chinese compared to alphabetical studies. This was possible due to the larger number of studies ($n=11$). Recall that β here reflects the difference in the effect sizes between alphabetical and Chinese studies, and that positive values indicate bigger effect sizes for Chinese

studies. There was 65 % probability that Chinese studies have bigger effect size for FFD (mean estimate of β : 1.5 [95% Credible interval: -6.2, 9.4]). There was higher probability (80 %) that the effect size is bigger for Chinese studies for GD (mean estimate of β : 5.1 [95% Credible interval: -6.7, 16.7]). Therefore, the results suggest that the N+2 preview effect on word N+2 is bigger for Chinese compared to alphabetical studies, and that this difference is more pronounced for GD compared to FFD.

N+1 Preview Effects

As our analysis included many more studies on N+1 preview effects than N+2 preview effects, the results for the N+1 preview effect are presented in Table 2 (however, see Supplementary files #2 for forest plots). This analysis includes all experiments with an N+1 boundary manipulation that had invalid parafoveal preview (unrelated word, pseudo-word, random letters or X-string mask). The probabilities associated with these effects are presented in Figure 5. As the breakdown shows, the majority of studies were alphabetical (78% for FFD, 73% for SFD, 75% for GD, and 100% for TVT). The effect sizes for all studies increased in the expected direction: FFD was smaller than SFD, which was smaller than GD, which in turn was smaller than TVT.

Table 2

Effect (in ms)

Measure	All studies			Alphabetical studies			Chinese studies		
	N	Mean	95% CrI	N	Mean	95% CrI	N	Mean	95% CrI
FFD	66	25.9	[22.3, 29.6]	52	26.1	[21.8, 30.3]	14	25.6	[17.7, 33.8]
SFD	30	32.6	[25, 40.4]	22	35.7	[26.2, 45.5]	8	25.4	[9.6, 41.2]
GD	70	39.8	[34.6, 44.9]	53	37.7	[31.9, 43.5]	17	46.8	[34.6, 60.1]
TVT	20	47.2	[35.2, 59.1]	20	47.2	[35.1, 58.8]	0	N/A	N/A

Note: N: number of experiments on which the analysis is based. Mean: posterior mean. CrI: credible intervals.

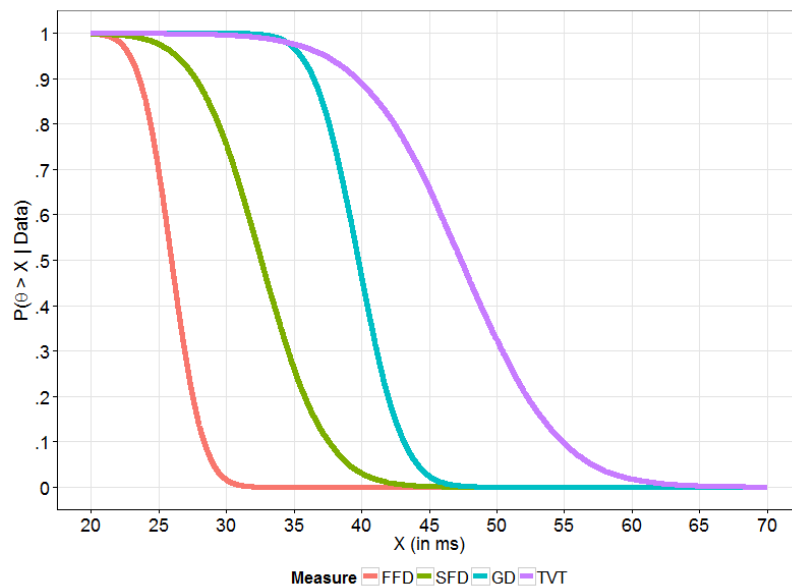


Figure 5. Probability that N+1 preview effects are greater than some number X, given the data.

The analysis was done on all studies from all languages.

Differences between alphabetical and Chinese studies. Similar to the N+2 analysis, we also fitted meta-regression models to investigate differences in the N+1 preview effect between Chinese and alphabetical studies when no useful information from the target word is obtained.

The findings are presented in Table 3. The results suggest that the common assumption that N+1 preview effects are bigger for Chinese studies holds up only for GD. However, this result could be confounded by the type of mask due to the fact that two masks (a string of letters, and a string of Xs) are not possible in Chinese studies. A separate analysis with unrelated word mask alone showed an overall high probability that effects sizes for Chinese studies are bigger. Critically, the difference for GD was almost 10 ms. However, the opposite result was found for the pseudo-word mask: there was relatively low probability that the effects sizes for Chinese are bigger.

Table 3

Mean Estimate of the Difference (β) in the N+1 Preview Effect Between Chinese and Alphabetical Studies

Measure	N _A	N _C	Mean	95% CrI	p(ES _C > ES _A)
FFD [all masks]	52	14	-0.29	[-4.45, 4]	44 %
SFD [all masks]	22	8	-5.14	[-13.49, 3.23]	11 %
GD [all masks]	53	17	4.09	[-1.82, 10.14]	91 %
FFD [unrelated word mask]	18	11	3.47	[-2.21, 8.79]	89 %
SFD [unrelated word mask]	10	7	1.46	[-8.53, 10.28]	64 %
GD [unrelated word mask]	19	13	9.92	[2.45, 17.25]	100 %
FFD [pseudo-word mask]	8	6	-3.83	[-12.99, 5.47]	19 %
GD [pseudo-word mask]	8	7	-0.58	[-13.43, 13.45]	46 %

Note: N_A: number of alphabetical studies. N_C: number of Chinese studies. p(ES_C > ES_A):

probability that effect size of Chinese studies is bigger than the effect size of alphabetical ones.

CrI: credible intervals.

Differences between parafoveal masks. The N+1 preview effect was further analysed by checking how the type of parafoveal mask influences the effect size. This exploratory analysis was done by fitting separate meta-analytical models for each mask type (detailed results for each effect size are available in Supplementary files # 2 [Figures S2.5 and S2.6]). These exploratory

findings were then used to fit meta-regression models that tested for differences in all possible pair-wise comparisons of masks. The results of these analyses are presented in Figure 6. In contrast to all analyses so far, this one also included masks that contain some information about the target word (orthographical, phonological or semantic).

It should be noted that while meta-regression does not solve the problem of the different sample sizes for each mask, the mean difference for each comparison is weighted by the precision of the studies that were included in each analysis. It is also worth remembering that these results should *not* be interpreted as orthographical, phonological or semantic preview *benefit* effects. Rather, this is an analysis of the trade-off between facilitation and interference effects depending on what type of mask is used.

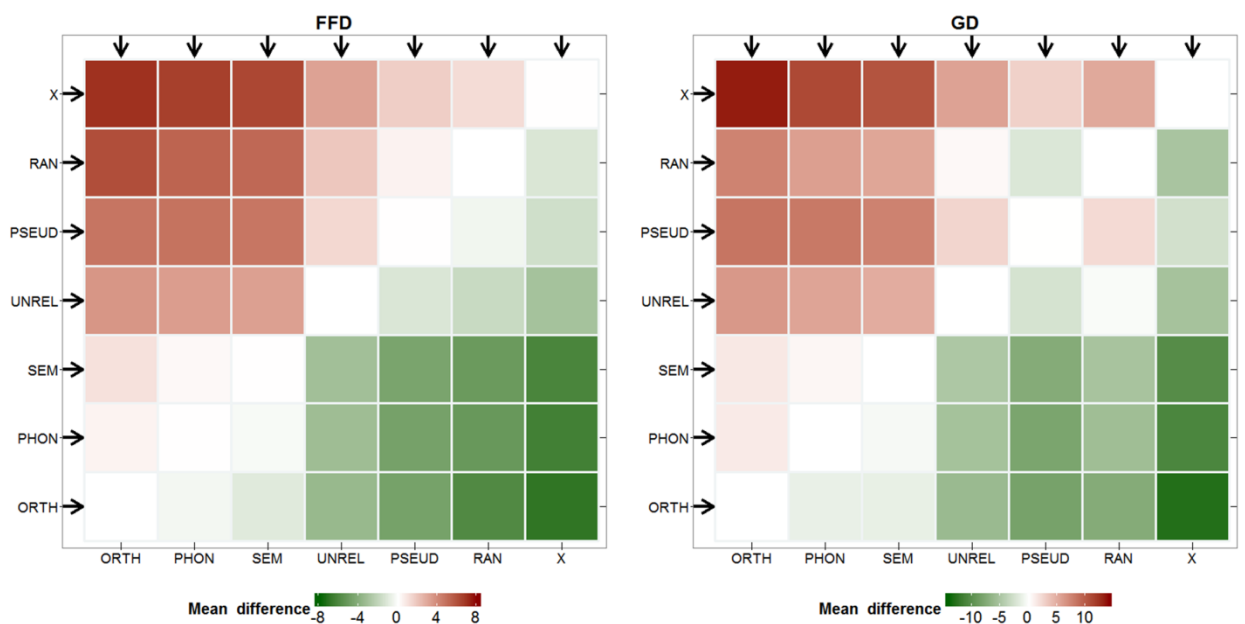


Figure 6. A heat map of facilitation (green) versus interference effects (red) associated with the type of parafoveal mask (based on all studies). The masks on the y axis are compared to the masks on the x axis (arrows are added to show the direction in which masks can be compared). ORTH: orthographical mask. PHON: phonological mask. SEM: semantic mask. UNREL:

unrelated word mask. PSEUD: pseudo-word mask. RAN: random letters mask. X: string of Xs mask.

There were three main findings. First, as expected, parafoveal masks that contain some information about the target word (orthographical, phonological, or semantic) resulted in the greatest facilitation. Conversely, unrelated word, pseudo-word, random letters, X-string masks led to the greatest interference in parafoveal processing since they do not contain any useful information about the target word. Second, there were also interference effects between masks that do not provide any information about the target word. In this analysis, unrelated word mask resulted in the least interference compared to the other three masks, and a string of Xs resulted in the greatest interference. Finally, the amount of facilitation of the masks that provide some information about the target word depended on the baseline that they are compared to.

To check whether the results are influenced by including all Chinese studies, the analysis was also repeated with alphabetical-only studies (fewer mask comparisons were possible due to the smaller sample sizes). Although there were minor differences, the main results remained: unrelated word mask led to the least interference, while the x-string mask led to the greatest interference (4.6ms larger for FFD and 10.5 ms larger for GD).

Robustness of the Main Results

We checked the robustness of the main results in order to make sure that they are not sensitive to the exclusion of any individual study (Deeks et al., 2008). To do this, the leave-one-out method was used, where the same analysis was repeated by omitting one different study each time (e.g. see Greenhouse & Iyengar, 2009, pp. 423-424). For the main N+2 results (presented in Figure 3), the effect sizes ranged from 4.7 to 5.9 ms for FFD (SD= 0.32) and from 3.3 to 7.9 ms for GD (SD=1.16). Therefore, the results for FFD were influenced very little by individual

studies; the results for GD were influenced slightly more, but this was still within a reasonable range for this number of studies. The results from the main N+1 analysis presented in Table 2 were also largely unaffected by this analysis. When studies from all languages were analysed together, the effect size ranged from 24.9 to 26.4 ms for FFD (SD= 0.23), 30.7 to 33.7 ms for SFD (SD= 0.71), 38.6 to 40.5 ms for GD (SD= 0.32), and 44 to 49.1 ms for TVT (SD= 1.29).

Discussion

The present results provide a new, comprehensive perspective on what can be believed about parafoveal preview effects given the available data. We will use this new perspective to revisit the open questions we identified in our review of the literature: (1) the existence of N+2 effects, (2) how the N+1 preview effect differs as a function of parafoveal mask, and (3) type of language. Finally, we will consider issues related to interpreting parafoveal preview effects. Before we address all these issues, we will consider the size of the N+1 preview effect very briefly.

How Big is the N+1 Preview Effect?

The N+1 preview effect has been investigated for the past 40 years. However, there is considerable uncertainty about the size of this effect reported in previous literature reviews on the topic (Rayner, 2009; Rayner et al., 2003). By taking into account all the available data, the present meta-analysis has greatly reduced this uncertainty. Based on the present results, it can be concluded that there is 95% probability that the N+1 preview effect is smaller than 29 ms for FFD and 45 ms for GD. As the results show, the effect was constrained not only to the so-called early measures (FFD, SFD, and GD), but also increased for TVT. The precise effect size for N+1 preview may not seem highly theoretically important, but as we will show in the next section, it actually has important theoretical implications when evaluating model predictions about the

Do N+2 Preview Effects Exist?

Perhaps the most controversial question addressed in this paper is whether N+2 preview effects exist. The present results demonstrate that the N+2 preview effect on word N+2 is mostly constrained to FFD and that its size is about 5 ms. Furthermore, there was a high probability that this effect is not centered on 0. Additionally, it was found that the effect is larger for Chinese compared to alphabetical studies, and that this difference is bigger for GD compared to FFD. This is consistent with the view that preview effects are larger for non-alphabetical languages such as Chinese due to the greater proximity of parafoveal words to the current fixation point (Yang et al., 2012b). However, it is important to keep in mind that this analysis was based on a small number of studies. In this sense, to get a better estimate of how N+2 preview effects differ between alphabetic and Chinese studies, more research is needed. Studies comparing the magnitude of N+2 preview effects in different languages within bilingual participants may be particularly helpful. In the meanwhile, it can be concluded that, based on the available evidence, the N+2 preview effect on word N+2 across languages is very likely not centred on 0 and that this effect is modest in size. Therefore, to answer the question posed at the beginning of this section, it does appear that N+2 preview effects exist.

One important question is whether the magnitude of this effect is of any practical or theoretical significance. Realistically, large N+2 preview effects should be prevented by the distance of word N+2 from the fixation point and the resulting loss of visual acuity. Therefore, as it was argued previously, the size of the N+2 effect can be meaningfully compared only in relation to the size of the N+1 effect. This comparison is of course limited by the different number of studies in each analysis. However, the best available evidence to date suggests that the

size of the N+2 preview effect is a non-trivial 20% of the N+1 effect size for FFD and 14% for GD. Therefore, even though the N+2 effect is rather small, it is not negligible when considered in relation to the N+1 effect.

On a more practical level, the present results are very useful for planning future N+2 studies. However, they also have theoretical implications for computational models of reading. Models such as the E-Z Reader (Reichle et al., 1998) and SWIFT (Engbert et al., 2005) make different predictions about parafoveal processing of word N+2. However, recent simulations with the two models show that they can both generate N+2 preview effects of the magnitude observed in this meta-analysis (Risse et al., 2014; Schotter et al., 2014). This suggests that modest N+2 effects can be accounted for by both serial-attention and parallel-attention models. Even so, the two models explain these effects in a different way: in parallel-attention models, they are generated because the attention gradient usually includes word N+2; in serial-attention models, they can be generated when the previous words are recognized prior to fixating word N+2.

The present results, taken together with the simulations, suggest that N+2 preview effects are not likely to distinguish between the two models. More importantly, however, the present findings greatly constrain what is possible with regards to parafoveal processing of word N+2. First, the results do not favour models that cannot explain N+2 preview effects. Second, they also constrain the size of N+2 effects that should be simulated. In fact, any model that simulates N+2 effects greater than 15 ms would be at odds with the present findings.

The precise estimates of both N+1 and N+2 preview effects are also critically important for evaluating the simulations of computational models of reading. As an example, we will use our estimates of N+1 and N+2 preview effects to evaluate the claims made by Schotter et al. (2014) about N+2 effects. Because of the difficulties inherent in simulating the gaze-contingent

boundary paradigm, Schotter et al. estimated the “preview time” available for parafoveal processing of word N+1 and word N+2 in the E-Z Reader model rather than directly simulating the actual preview effects. Using our estimates, we can link these simulated preview times to our numerical estimates of the N+1 and N+2 preview effects. Specifically, the key result of Schotter et al.’s Simulation 2 was that word N+2 is only previewed on 20% of trials for a preview time of 69 ms (and for 0 ms on 80% of trials). In our meta-analysis, we observed a mean N+2 effect size of 5.4 ms in FFD. Using Schotter et al.’s simulation results, we can assume that our estimate for the N+2 effect of 5.4 ms is a weighted average of the 0 ms effect in 80% of the trials and an unknown, nonzero effect in 20% of the trials. It follows that this unknown N+2 preview effect in the trials in which N+2 was actually previewed must be $5.4 \times 1/0.2 = 27$ ms.

We can do the same calculation for the N+1 preview effect. Simulation 1 of Schotter et al. (2014) showed that word N+1 is previewed on 89% of trials and that, if it is previewed, the internal “preview time” in the E-Z Reader model is 98 ms. If we assume, following Schotter et al.’s simulation results, that our estimate for the N+1 effect of 25.9 ms is a weighted average of the 0 ms effect in 11% of the trials and an unknown, nonzero effect in 89% of the trials, it follows that this unknown effect in the trials in which N+1 was actually previewed must be $25.9 \times 1/0.89 = 29.1$ ms. This leads us to a surprising conclusion: For N+1, every 1 ms of “preview time” corresponds to 0.297 ms of observed N+1 preview effect in FFD, while, for N+2, every 1 ms of “preview time” corresponds to 0.391 ms of observed N+2 preview effect in FFD. Given that it is usually assumed that pre-processing of word N+2 is less efficient due to visual acuity constraints, this result is quite unexpected and, at the very least, demands additional theoretical explanation. Without numerical estimates for N+1 and N+2 effects, it would be impossible to detect this discrepancy between the theoretical claim of N+2 preview occurring on only 20% of

trials and our empirically observed effect sizes. In this sense, the numerical values of both N+2 and N+1 preview effects are critically important for evaluating computational models against the empirical data.

Finally, the results of the N+2 analysis are also very clear in one other aspect: there is a considerable difference in the precision of some of the studies (see Figure 3). While the robustness analysis showed that this had little effect on the present results, it does illustrate very well the need for high-power N+2 studies in the future. Although meta-analysis is a good tool for summarising evidence, it is no substitute for replication studies (Van Elk et al., 2015). In this sense, it is recommended that high-power, registered replications of N+2 studies are undertaken in the future. Also, due to the lack of consistently reported data, it was not possible to examine whether the N+2 results are influenced by skipping probability. This remains an issue that needs to be explored in future research.

Does the Type of Parafoveal Mask Influence the Size of the N+1 Preview Effect?

The present data clearly demonstrate that the N+1 preview effect is influenced by the type of parafoveal mask. On the surface, this conclusion is hardly surprising. However, it is of interest to examine two types of masks: the ones that provide some information about the target word and the ones that do not. If the parafoveal mask during invalid preview contains orthographical, phonological, or semantic information about the target word, it can potentially facilitate word recognition times. Previous studies have suggested that both orthographical and phonological overlap between the mask and the target word can reduce the N+1 preview effect and thus reduce target word processing times (see Schotter et al., 2012). Similarly, semantic overlap has also been shown to result in such a benefit in word processing times in German (Hohenstein & Kliegl, 2014; Hohenstein, Laubrock, & Kliegl, 2010), but so far has been observed only under

certain conditions in English (cf. Rayner & Schotter, 2014; Rayner, Schotter, & Drieghe, 2014; Schotter, Lee, Reiderman, & Rayner, 2015). Interestingly, there are also individual studies in Chinese that have reported numerically greater effect sizes for phonological masks compared to semantic ones (e.g. Pan, Laubrock, & Yan, 2016, Exp2; Tsai, Kliegl, & Yan, 2012; Yan, Richter, Shu, & Kliegl, 2009), thus suggesting that semantic information in Chinese may be more readily accessible in parafoveal vision compared to phonological information.

The approach taken in the present paper was simply to quantify the amount of facilitation that semantic, phonological, and orthographic masks induce. A more fine-grained analysis of how this facilitation differs between alphabetical and Chinese languages was not possible due to the small number of studies that had such masks. The findings were consistent with the previous literature in indicating that all three masks led to facilitation effects compared the masks that contain no useful information about the target word. Moreover, orthographic masks led to the greatest facilitation effect and semantic masks led to the least facilitation. However, it is also notable that the difference between the three masks was very small.

A second question of perhaps greater importance is whether N+1 preview effects also differ between invalid masks that contain no useful information about the target word. Until now, no systematic investigation of this has been undertaken. The results from the present study suggest that the type of mask indeed influences the size of the N+1 preview effect. The analysis indicated that the unrelated word mask results in the smallest N+1 effect sizes. All other masks (pseudo-word, string of letters, string of Xs) resulted in greater effect sizes, which indicates that there are interference effects associated with them.

A conceptual illustration of the results is presented in Figure 7. Specifically, we assume that there is a “pure” preview benefit effect of unknown size, and that different masks used as

invalid preview can introduce interference effects that increase the overall size of the observed effect. In contrast to that, parafoveal masks that contain some information about the target word lead to facilitation effects that speed up word processing. The current findings suggest that these interference effects are not very big and amount to up to several milliseconds for first-pass measures. However, as the present findings are only exploratory in nature, it is important that this issue is further investigated experimentally.

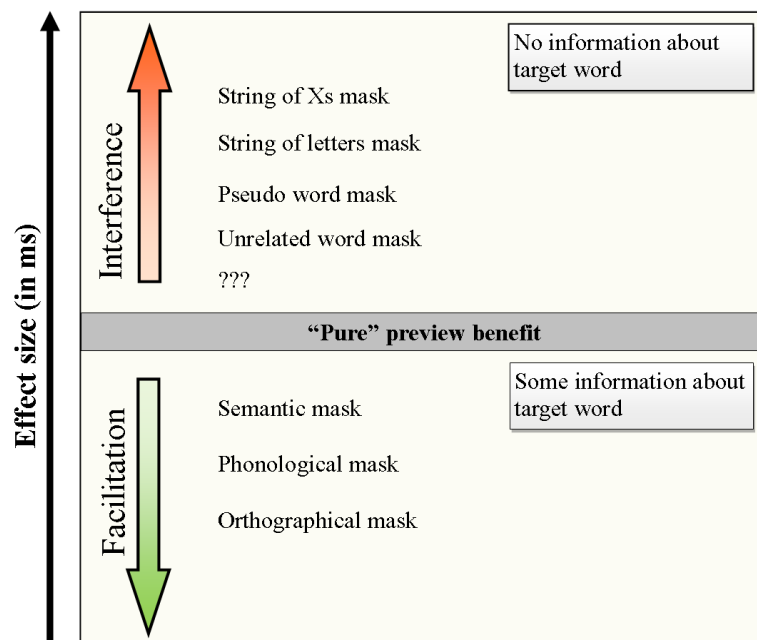


Figure 7. An illustration of the effect of mask type on the N+1 preview effect (based on all studies). Masks that contain no information about the target word can lead to interference, while masks that contain information about the target word can lead to facilitation.

One possible explanation of the present results is that masks such as random letters and a string of Xs may look more “unnatural” and may be more likely to be noticed by participants (consciously or not). This would explain how unrelated word masks result in the smallest N+1 effects sizes, since these masks are actual words that occur in the language. Indeed, there is some

evidence that preview effects are larger for participants who noticed display changes compared to those who did not (Angele, Slattery, & Rayner, 2016; Slattery, Angele, & Rayner, 2011; White, Rayner, & Liversedge, 2005). The present results showed that interference effects roughly increased as the parafoveal mask becomes less “word-like” (see Figure 7). This is consistent with the results of Angele et al. (2016) that participants are more sensitive to changes after nonword-like previews compared to word-like previews. Therefore, it could be the case that the interference effects observed in this analysis are due to differences in display-change sensitivity. However, this conclusion is still in the realm of speculation, and more research is needed to understand why such interference effects exist.

Currently, there is very scarce data about the processing costs associated with different parafoveal masks. In one recent study, Hutzler et al. (2013) argued that parafoveal X-string masks interfere with word recognition. Interestingly, however, it appears that very few studies (6 in this analysis) have used such masks. Nevertheless, the results of this study support Hutzler et al.’s (2013) claim, although it should be noted that this evidence is suggestive at best due to the different sample sizes. More recently, Marx et al. (2015) have provided more evidence for the processing costs associated with such masks and have argued in favour of a more neutral mask (visual degradation)⁵. Clearly, the jury is still out on what the best mask is; however, this is an area that deserves more attention from researchers.

Finally, it should be noted that the present findings also have a couple of limitations. First, since the different masks also had different sample sizes, this could have influenced the results. However, this is not a critical problem, because the Bayesian approach adopted in this paper makes it possible to easily update the present results as more evidence becomes available.

⁵ Hutzler et al.’s (2013) and Marx et al.’s (2015) studies were not included in the meta-analysis as they did not meet all inclusion criteria.

Second, all effect sizes in the present analysis had the confound that invalid preview conditions contain a display change, but the valid preview condition does not. This limitation, however, is not unique to the present analysis and is indeed present in much of the research on reading that utilises gaze-contingent display change techniques. Therefore, there is a need for better understanding of how display changes influence the underlying effects, particularly when using different types of masks.

Are Preview Effects Bigger for Chinese Studies?

There has been an ongoing effort to determine how much reading in Chinese differs from reading in alphabetical languages, and whether models of reading in alphabetical languages are also appropriate for reading in Chinese. Given that parafoveal processing is a major component of most models of skilled reading, it is important to determine whether parafoveal preview effects are different for Chinese compared to alphabetical languages. Overall, the results of the meta-regression analyses showed that both N+1 and N+2 effects are greater in Chinese studies when measured with GD. Although there were also analyses in which a similar result was found for FFD, the mean difference was not very large. Interestingly, the difference in the N+1 preview effect was influenced by the type of mask. For unrelated word mask (the most common one), the effect sizes for Chinese studies were reliably higher. However, the opposite result was found for pseudo-word masks. This discrepancy could be due to the different sample size of the two analyses. Thus, a better estimate of the difference can be obtained when more studies are available. It also wasn't possible to investigate differences in "late" measures such as TVT because Chinese studies did not report descriptive statistics for this measure. Therefore, this needs to be addressed by future research.

The present findings are therefore consistent with the assumption that readers of Chinese

make more efficient use of parafoveal processing. The intuitive explanation of this is that information is more densely packed in Chinese and it is situated close to the point of fixation. Indeed, as the majority of Chinese words consist of one or two characters (Li, Zang, Liversedge, & Pollatsek, 2015; Zang et al., 2015), word N+1 may sometimes even fall in foveal vision. Therefore, this should facilitate parafoveal processing as there will be smaller loss of visual acuity. However, it could also be speculated that some of this advantage in parafoveal processing may be confounded by display change awareness. Indeed, Slattery et al.'s (2011) display-change detection paradigm showed that display change awareness is greater when the subjects' fixation was closer to the invalid preview before triggering the display change. Therefore, as the boundary will generally be closer to the point of fixation in Chinese, it would be interesting for future research to look into how display change awareness affects preview effects in Chinese.

Limitations

One potential criticism of the present findings is that they do not tell researchers much beyond what they did not know already. However, this argument misses the point of the present analysis. Even though researchers may have their own intuitions (i.e. prior beliefs) about parafoveal preview effects, the present analysis makes it possible to update these beliefs in the light of all the available evidence. Also, it could be argued that the N+2 preview effects could be explained by misallocated fixations or the reliability of measures. While this possibility cannot be excluded, such an argument assumes that these sources of error selectively affect N+2 effects, but not all other preview effects. Such a scenario is, of course, unlikely. Even though this is a valid concern, we argue that such systematic sources of error will be averaged out in the long run as more and more data becomes available.

On Being Cautious When Interpreting Parafoveal Preview Effects

The present paper has purposefully avoided talking about preview *benefit* effects. The rationale behind this was that there is accumulating evidence that N+1 preview effects are not just the result of being able to pre-process the target but also depend on properties of the mask. The present results reinforce this notion. Despite the fact that unrelated word, pseudo-word, random letters and X-string masks all do not provide any useful information about the target word, they result in different effect sizes. This difference was not trivial in some of the cases and it is not likely to be explained by sample size differences or sampling variability alone, as differences between mask types were found even in analyses done on comparable sample sizes. Therefore, it is important that the differences between parafoveal masks are further investigated experimentally. Also, future studies should take care to consider the type of mask that was used when interpreting the results.

The current findings reflect our best understanding of parafoveal preview effects in the light of the available data. However, they are not the last word on this topic. The beauty of the Bayesian approach is that the present results can be updated in a natural and intuitive way as more evidence becomes available. This will further reduce the uncertainty and give us an even better understanding of what can be believed about parafoveal preview effects given the data.

The present results also show that the true complexity of parafoveal preview effects may not always be captured by simply comparing statistical significance across studies. Even if no statistically-significant result is found when comparing two parafoveal masks, this does not by necessity mean that the effect is null or that it has absolutely no practical importance (see Kirk, 1996). To truly capture such complexity in the data, it is recommended that researchers use Bayes factors as a supplementary analysis (Rouder et al., 2009; Rouder, Morey, Speckman, & Province, 2012; for a recent application, see Abbott & Staub, 2015). The present study did not

use Bayes factors as it was mostly concerned with the estimation of effect sizes. However, Bayes factors are very useful for hypothesis testing as they give a ratio of how likely the null and alternative hypotheses are given the data. In this way, it is possible to test whether the data support the null hypothesis that there is no difference between two parafoveal preview conditions or the alternative hypothesis that such difference exists.

Conclusion

Parafoveal processing during reading has been an active area of research for the past four decades and will likely remain so in the future. The present study is one of the first attempts to make a statistical synthesis of previous findings. The results showed that there is high probability that N+2 effects are not centred on 0 and that they are modest in size. The N+1 preview effect was perhaps smaller than what has sometimes been assumed, and it was influenced by the type of parafoveal mask: both by masks that provide information about the target word and the ones that do not. The type of language also had an effect on the results, with Chinese studies generally increasing the size of the effects. The present results are also relevant for computational models of eye-movement control during reading. Actual model simulations are needed to determine whether existing models can accommodate the effects that were observed in this meta-analysis. It is hoped, however, that the present data will be useful both for computational modelling and for planning new experiments.

Study inclusion criteria

- The study used the boundary paradigm (N+1, N+2, or both) to manipulate parafoveal preview. Studies with the moving window paradigm (or modifications of it) were excluded. Studies with multiple boundary manipulations were also excluded.
- The study included whole-sentence reading (and not lists of words, etc.).
- Reading was silent.
- Participants were native speakers of the language
- The population was healthy, young adults (usually university students).
- The sentences themselves were “normal” (e.g. neutral tone, nothing that deviates too much from everyday reading, no potential garden-path effects, syntactic violations/ambiguities).
- The study reported descriptive statistics for at least one of the standard fixation duration measures (FFD, SFD, GD or TVT).
- Any additional manipulations other than parafoveal preview should not be too disruptive to the reading process. Such manipulations were averaged out.
- The researchers took measures to ensure reasonable eye-tracking data quality (e.g. calibration, drift check/correction, efforts to minimize artefacts due to head movements).
- Only one word was masked (and not e.g. the whole remaining sentence)
- The condition(s) in which the target word was masked had a whole-word mask (with the exception of orthographical masks).
- Studies with compound target words were coded only if there was a condition in which the whole compound word was masked.

- There was a condition with valid preview of the target word. This serves as the baseline.

Studies without such condition were excluded.

Appendix B

Funnel plots are a type of a scatterplot that compares the observed effect size against the precision of the study (usually defined as the inverse of the standard error). In the absence of bias, studies with smaller sample size should scatter at the bottom of the graph, while studies with larger sample size should be more narrowly spread towards the top of the graph (Sterne et al., 2011). It should be kept in mind that existing heterogeneity between studies (e.g. whether a study is alphabetical or Chinese) can lead to asymmetry in the funnel plot. All funnel plots were generated with the “funnel” function of the “meta” package in R (Schwarzer, 2015).

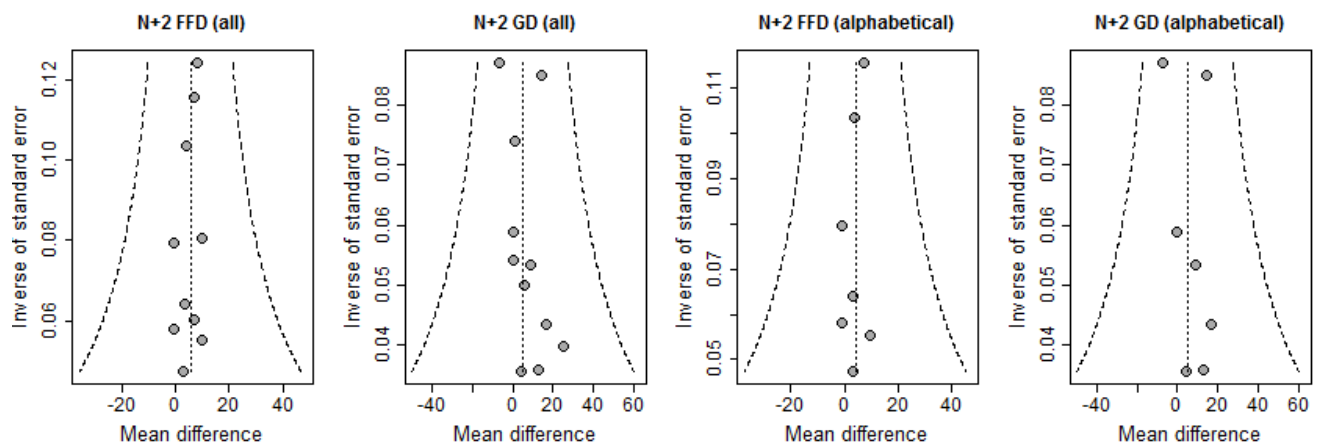


Figure B.1. Funnel plots for the N+2 preview effect (with 95% confidence intervals).

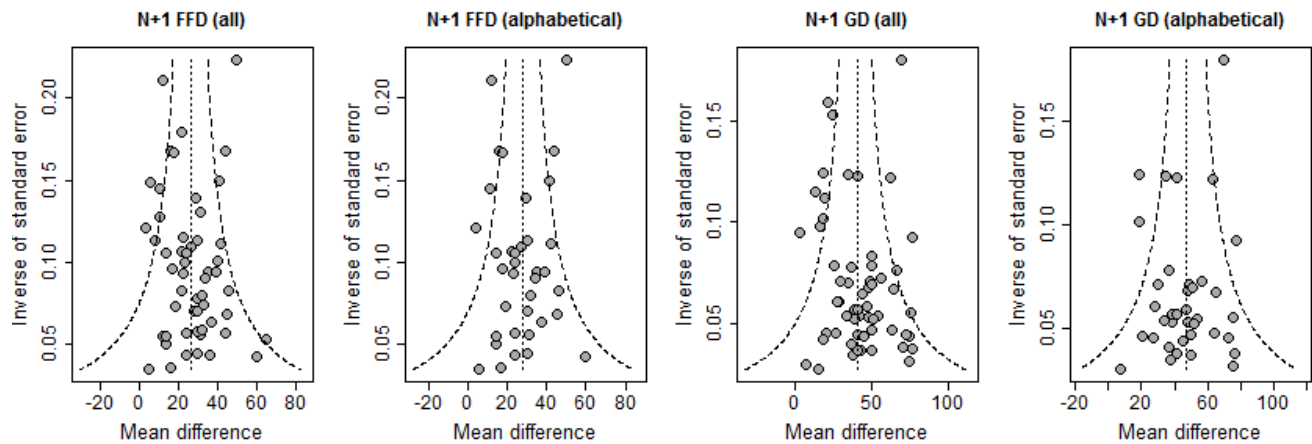


Figure B.2. Funnel plots for the N+1 preview effect (with 95% confidence intervals). Studies with imputed variance are not included.

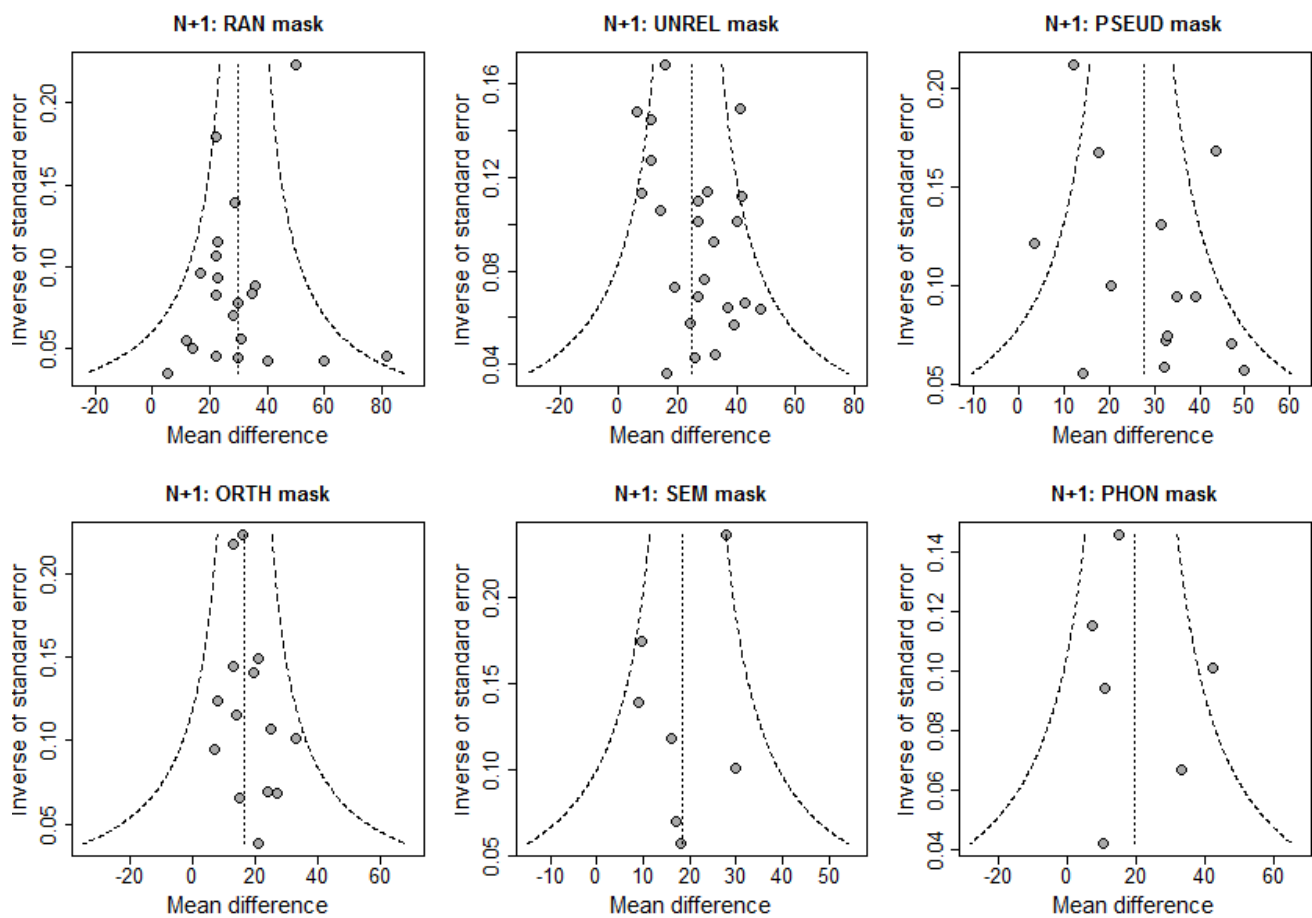


Figure B.3. Funnel plots for all N+1 studies broken down by mask type (with 95% confidence intervals). Studies with imputed variance are not included. The measure is FFD.

References

References marked with an asterisk indicate studies included in the meta-analysis.

- Abbott, M. J., & Staub, A. (2015). The effect of plausibility on eye movements in reading: Testing E-Z Reader's null predictions. *Journal of Memory and Language*, 85, 76-87.
doi:10.1016/j.jml.2015.07.002
- Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E. J. (2014). An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, 41(2), 521-543. doi:10.1177/0149206314560412
- Andrews, M., & Baguley, T. (2013). Prior approval: The growth of Bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology*, 66(1), 1-7.
doi:10.1111/bmsp.12004
- *Angele, B., Laishley, A. E., Rayner, K., & Liversedge, S. P. (2014). The effect of high-and low-frequency previews and sentential fit on word skipping during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(4), 1181-1203.
doi:10.1037/a0036396
- Angele, B., Slattery, T. J., & Rayner, K. (2016). Two stages of parafoveal processing during reading: Evidence from a display change detection task. *Psychonomic Bulletin & Review*, 1-9. doi:10.3758/s13423-015-0995-0
- *Angele, B., Tran, R., & Rayner, K. (2013c). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 526-538.
doi:10.1037/a0029492

- *Angele, B., & Rayner, K. (2011). Parafoveal processing of word n+ 2 during reading: Do the preceding words matter? *Journal of Experimental Psychology: Human Perception and Performance*, 37(4), 1210-1220. doi:10.1037/a0023096
- *Angele, B., & Rayner, K. (2013a). Processing the in the parafovea: Are articles skipped automatically?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(2), 649-662. doi:10.1037/a0029294
- *Angele, B., & Rayner, K. (2013b). Eye movements and parafoveal preview of compound words: Does morpheme order matter?. *The Quarterly Journal of Experimental Psychology*, 66(3), 505-526. doi:10.1080/17470218.2011.644572
- *Angele, B., Slattery, T. J., Yang, J., Kliegl, R., & Rayner, K. (2008). Parafoveal processing in reading: Manipulating n+ 1 and n+ 2 previews simultaneously. *Visual Cognition*, 16(6), 697-707. doi: 10.1080/13506280802009704
- Anzures-Cabrera, J., & Higgins, J. P. (2010). Graphical displays for meta-analysis: An overview with suggestions for practice. *Research Synthesis Methods*, 1(1), 66-80.
doi:10.1002/jrsm.6
- *Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17(3), 364-390.
doi:10.1016/0010-0285(85)90013-1
- *Bélanger, N. N., Mayberry, R. I., & Rayner, K. (2013). Orthographic and phonological preview benefits: Parafoveal processing in skilled and less-skilled deaf readers. *The Quarterly Journal of Experimental Psychology*, 66(11), 2237-2252.
doi:10.1080/17470218.2013.780085
- *Binder, K. S., Pollatsek, A., & Rayner, K. (1999). Extraction of information to the left of the

- fixated word in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 25(4), 1162-1172. doi:10.1037/0096-1523.25.4.1162
- Borenstein, M., Hedges, L. V., & Higgins, J. P. T. (2009). *Introduction to meta-analysis*. Hoboken, NJ, USA: Wiley.
- *Briihl, D., & Inhoff, A. W. (1995). Integrating information across fixations during reading: The use of orthographic bodies and of exterior letters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1), 55-67. doi:10.1037/0278-7393.21.1.55
- *Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: effects of reading skill. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, 59(3), 209-217. doi:10.1037/h0087476
- *Choi, W., & Gordon, P. C. (2014). Word skipping during sentence reading: Effects of lexicality on parafoveal processing. *Attention, Perception, & Psychophysics*, 76(1), 201-213. doi:10.3758/s13414-013-0494-1
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12), 997-1003. doi:10.1037/0003-066X.49.12.997
- *Cui, L., Drieghe, D., Bai, X., Yan, G., & Liversedge, S. P. (2014). Parafoveal preview benefit in unspaced and spaced Chinese reading. *The Quarterly Journal of Experimental Psychology*, 67(11), 2172-2188. doi:10.1080/17470218.2014.909858
- Cumming, G. (2013). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29. doi:10.1177/0956797613504966
- *Dare, N., & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *The Quarterly Journal of Experimental Psychology*, 66(3), 487-504.

- Deeks, J. J., Higgins, J. P. T., & Altman, D.G. (Eds.). (2008). Analysing data and undertaking meta-analyses. In J. P. T Higgins & S. Green, *Cochrane handbook for systematic reviews of interventions* (pp. 243-296). United Kingdom: Wiley-Blackwell.
- *Deutsch, A., Frost, R., Pelleg, S., Pollatsek, A., & Rayner, K. (2003). Early morphological effects in reading: Evidence from parafoveal preview benefit in Hebrew. *Psychonomic Bulletin & Review*, 10(2), 415-422. doi:10.3758/BF03196500
- *Deutsch, A., Frost, R., Pollatsek, A., & Rayner, K. (2005). Morphological parafoveal preview benefit effects in reading: Evidence from Hebrew. *Language and Cognitive Processes*, 20(1-2), 341-371. doi:10.1080/01690960444000115
- *Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 954- 969. doi:10.37/0096-1523.31.5.954
- Edwards, A. W. (1974). The history of likelihood. *International Statistical Review/Revue Internationale de Statistique*, 42(1) 9-15.
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112(4), 777–813. doi:10.1037/0033-295X.112.4.777
- Engelmann, F., Jäger, L. A., & Vasishth, S. (2016). The determinants of retrieval interference in dependency resolution: Review and computational modeling. *Manuscript submitted for publication*.
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502-1505. doi:10.1126/science.1255484

Furukawa, T. A., Barbui, C., Cipriani, A., Brambilla, P., & Watanabe, N. (2006). Imputing

missing standard deviations in meta-analyses can provide accurate results. *Journal of*

Clinical Epidemiology, 59(1), 7-10. doi:10.1016/j.jclinepi.2005.06.006

Gagl, B., Hawelka, S., Richlan, F., Schuster, S., & Hutzler, F. (2014). Parafoveal preprocessing

in reading revisited: Evidence from a novel preview manipulation. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 40(2), 588-595.

doi:10.1037/a0034408

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment

on article by Browne and Draper). *Bayesian Analysis*, 1(3), 515-534.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple

sequences. *Statistical Science*, 7(4), 457-472.

Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not

itself statistically significant. *The American Statistician*, 60(4), 328-331.

doi:10.1198/000313006X152649

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33(5), 587-606.

doi:10.1016/j.socec.2004.09.033

Greenhouse, J. B. & Iyengar, S. (2009). Sensitivity analysis and diagnostics. In H. M. Cooper, L.

V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 418-434). New York: Russell Sage Foundation.

*Gu, J., Li, X., & Liversedge, S. P. (2015). Character order processing in Chinese reading.

Journal of Experimental Psychology: Human Perception and Performance, 41(1), 127-

137. doi:10.1037/a0038639

*Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual

- span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3), 417-429.
doi:10.1037/0278-7393.16.3.417
- Higgins J. P. T. & Green, S. (Eds.) (2011). Cochrane handbook for systematic reviews of interventions (version 5.1.0 [updated March 2011]). The Cochrane Collaboration.
Retrieved from www.cochrane-handbook.org.
- Higgins, J. P., & Spiegelhalter, D. J. (2002). Being skeptical about meta-analyses: A Bayesian perspective on magnesium trials in myocardial infarction. *International Journal of Epidemiology*, 31(1), 96-104. doi:10.1093/ije/31.1.96
- *Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(1), 166-190.
doi:10.1037/a0033670
- Hohenstein, S., Laubrock, J., & Kliegl, R. (2010). Semantic preview benefit in eye movements during reading: A parafoveal fast-priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1150-1170. doi:10.1037/a0020233
- Hutzler, F., Fuchs, I., Gagl, B., Schuster, S., Richlan, F., Braun, M., & Hawelka, S. (2013). Parafoveal X-masks interfere with foveal word recognition: Evidence from fixation-related brain potentials. *Frontiers in Systems Neuroscience*, 7, 33.
doi:10.3389/fnsys.2013.00033
- *Hyönä, J., & Häikiö, T. (2005). Is emotional content obtained from parafoveal words during reading? An eye movement analysis. *Scandinavian Journal of Psychology*, 46(6), 475-483. doi:10.1111/j.1467-9450.2005.00479.x
- *Inhoff, A. W., Starr, M., Liu, W., & Wang, J. (1998). Eye-movement-contingent display changes

- are not compromised by flicker and phosphor persistence. *Psychonomic Bulletin & Review*, 5(1), 101-106. doi:10.3758/BF03209463
- *Inhoff, A. W., Starr, M., & Shindler, K. L. (2000). Is the processing of words during eye fixations in reading strictly serial?. *Perception & Psychophysics*, 62(7), 1474-1484. doi:10.3758/BF03212147
- *Inhoff, A. W., & Tousman, S. (1990). Lexical integration across saccades in reading. *Psychological Research*, 52(4), 330-337. doi:10.1007/BF00868065
- Jacobs, A. M., Grainger, J., & Ferrand, L. (1995). The incremental priming technique: A method for determining within-condition priming effects. *Perception & Psychophysics*, 57(8), 1101-1110. doi:10.3758/BF03208367
- *Johnson, R. L., & Dunne, M. D. (2012). Parafoveal processing of transposed-letter words and nonwords: Evidence against parafoveal lexical activation. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1), 191-212. doi:10.1037/a0025983
- *Johnson, R. L., Perea, M., & Rayner, K. (2007). Transposed-letter effects in reading: Evidence from eye movements and parafoveal preview. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1), 209-229. doi:10.1037/0096-1523.33.1.209.
- *Kambe, G. (2004). Parafoveal processing of prefixed words during eye fixations in reading: Evidence against morphological influences on parafoveal preprocessing. *Perception & Psychophysics*, 66(2), 279-292. doi:10.3758/BF03194879
- *Kennison, S. M., & Clifton, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1), 68-81. doi:10.1037/0278-7393.21.1.68

- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746-759. doi:10.1177/0013164496056005002
- Kliegl, R., Hohenstein, S., Yan, M., & McDonald, S. A. (2013). How preview space/time translates into preview cost/benefit for fixation durations during reading. *The Quarterly Journal of Experimental Psychology*, 66(3), 581-600.
doi:10.1080/17470218.2012.658073
- *Kliegl, R., Risse, S., & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word n+ 2. *Journal of Experimental Psychology: Human Perception and Performance*, 33(5), 1250-1255. doi:10.1037/0096-1523.33.5.1250
- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.
- Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, 14(7), 293-300. doi:10.1016/j.tics.2010.05.001
- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). London: Academic Press.
- Kruschke, J. K., & Liddell, T. M. (2015). The Bayesian new statistics: Two historical trends converge. *Available at SSRN 2606016*.
- Lajeunesse, M. J. (2013). Recovering missing or partial data from studies: A survey of conversions and imputations for meta-analysis. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *The handbook of meta-analysis in ecology and evolution* (pp. 195-206). Princeton University Press: Princeton, New Jersey.
- Li, X., Zang, C., Liversedge, S. P., & Pollatsek, A. (2015). The role of words in Chinese reading. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 232-244).

- *Lima, S. D. (1987). Morphological analysis in sentence reading. *Journal of Memory and Language*, 26(1), 84-99. doi:10.1016/0749-596X(87)90064-7
- *Liu, W., Inhoff, A. W., Ye, Y., & Wu, C. (2002). Use of parafoveally visible characters during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1213-1227. doi:10.1037/0096-1523.28.5.1213
- Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. Springer Science & Business Media.
- Marx, C., Hawelka, S., Schuster, S., & Hutzler, F. (2015). An incremental boundary study on parafoveal preprocessing in children reading aloud: Parafoveal masks overestimate the preview benefit. *Journal of Cognitive Psychology*, 27(5), 1-13.
doi:10.1080/20445911.2015.1008494
- *Masserang, K. M., & Pollatsek, A. (2012). Transposed letter effects in prefixed words: Implications for morphological decomposition. *Journal of Cognitive Psychology*, 24(4), 476-495. doi:10.1080/20445911.2012.658037
- *McDonald, S. A. (2006). Parafoveal preview benefit in reading is only obtained from the saccade goal. *Vision Research*, 46(26), 4416-4424. doi:10.1016/j.visres.2006.08.027
- *Miellet, S., & Sparrow, L. (2004). Phonological codes are assembled before word fixation: Evidence from boundary paradigm in sentence reading. *Brain and Language*, 90(1), 299-310. doi:10.1016/S0093-934X(03)00442-5
- *Miller, B., Juhasz, B. J., & Rayner, K. (2006). The orthographic uniqueness point and eye movements during reading. *British Journal of Psychology*, 97(2), 191-216.
doi:10.1348/000712605X66845

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & the PRISMA group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement.

Annals of Internal Medicine, 151(4), 264-269. doi:10.1371/journal.pmed.1000097

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas-Part II. *arXiv preprint*:1602.00245.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E. J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, 14(9), 1105-1107. doi:10.1038/nn.2886

Pan, J., Laubrock, J., & Yan, M. (2016). Parafoveal processing in silent and oral reading: Reading mode influences the relative weighting of phonological and semantic information in Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication.

*Payne, B. R., & Stine-Morrow, E. A. (2012). Aging, parafoveal preview, and semantic integration in sentence processing: Testing the cognitive workload of wrap-up. *Psychology and Aging*, 27(3), 638-649. doi:10.1037/a0026540

Perezgonzalez, J. D. (2015). The meaning of significance in data testing. *Frontiers in Psychology*, 6, 1293. doi:10.3389/fpsyg.2015.01293

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, p. 125). Technische Universität at Wien.

*Plummer, P., & Rayner, K. (2012). Effects of parafoveal word length and orthographic features on initial fixation landing positions in reading. *Attention, Perception, & Psychophysics*, 74(5), 950-963. doi:10.3758/s13414-012-0286-z

- *Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1), 148-162.
doi:10.1037/0096-1523.18.1.148
- R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/>.
- *Radach, R., Inhoff, A. W., Glover, L., & Vorstius, C. (2013). Contextual constraint and N+2 preview effects in reading. *The Quarterly Journal of Experimental Psychology*, 66(3), 619-633. doi: 10.1080/17470218.2012.761256
- Radach, R., & Kennedy, A. (2013). Eye movements in reading: Some theoretical context. *The Quarterly Journal of Experimental Psychology*, 66(3), 429-452.
doi:10.1080/17470218.2012.750676
- Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 7(1), 65-81. doi:[10.1016/0010-0285\(75\)90005-5](https://doi.org/10.1016/0010-0285(75)90005-5)
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422. doi:10.1037/0033-2909.124.3.372
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457-1506.
doi:10.1080/17470210902816461
- *Rayner, K., Balota, D. A., & Pollatsek, A. (1986). Against parafoveal semantic preprocessing during eye fixations in reading. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, 40(4), 473-483. doi:10.1037/h0080111
- *Rayner, K., Castelhana, M. S., & Yang, J. (2010). Preview benefit during eye fixations in

reading for older and younger readers. *Psychology and Aging*, 25(3), 714-718.

doi:10.1037/a0019199.

*Rayner, K., Juhasz, B. J., & Brown, S. J. (2007). Do readers obtain preview benefit from word n+ 2? A test of serial attention shift versus distributed lexical processing models of eye movement control in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1), 230-245. doi:10.1037/0096-1523.33.1.230

*Rayner, K., & Schotter, E. R. (2014a). Semantic preview benefit in reading English: The effect of initial letter capitalization. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1617-1628. doi:10.1037/a0036763

*Rayner, K., Schotter, E. R., & Drieghe, D. (2014b). Lack of semantic parafoveal preview benefit in reading revisited. *Psychonomic Bulletin & Review*, 21(4), 1067-1072. doi:10.3758/s13423-014-0582-9

Rayner, K., White, S. J., Kambe, G., Miller, B., & Liversedge, S. P. (2003). On the processing of meaning from parafoveal vision during eye fixations in reading. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 213-234). Amsterdam: Elsevier.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1), 125-157. doi:10.1037/0033-295X.105.1.125

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, 7(1), 34-55. doi:10.1016/j.cogsys.2005.07.006

*Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of

eye movements in reading: Evidence from a survival analysis of fixation durations.

Cognitive Psychology, 65(2), 177-206. doi:10.1016/j.cogpsych.2012.03.001

*Risse, S., & Kliegl, R. (2011). Adult age differences in the perceptual span during reading.

Psychology and Aging, 26(2), 451-460. doi:10.1037/a0021616

Risse, S., & Kliegl, R. (2012). Evidence for delayed parafoveal-on-foveal effects from word n+2

in reading. *Journal of Experimental Psychology: Human Perception and Performance*,

38, 1026-1042. doi:10.1037/a0027735

Risse, S., & Kliegl, R. (2014). Dissociating preview validity and preview difficulty in parafoveal

processing of word n+1 during reading. *Journal of Experimental Psychology: Human*

Perception and Performance, 40(2), 653-668. doi:10.1037/a0034997

Risse, S., & Vasilev, M. R. (in preparation). The time course of parafoveal processing during

reading.

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim.

Psychonomic Bulletin & Review, 18(4), 682-689. doi:10.3758/s13423-011-0088-7

Rouder, J. N., Morey, R. D., & Province, J. M. (2013). A Bayes factor meta-analysis of recent

extrasensory perception experiments: Comment on Storm, Tressoldi, and Di Risio (2010).

Psychological Bulletin, 139(1), 241-247. doi:10.1037/a0029008

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors

for ANOVA designs. *Journal of Mathematical Psychology*, 56(5), 356-374.

doi:10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for

accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225-

237. doi:10.3758/PBR.16.2.225

Royall, R. (1997). *Statistical evidence: A likelihood paradigm* (Vol. 71). London: Chapman & Hall.

Salpeter, S. R., Cheng, J., Thabane, L., Buckley, N. S., & Salpeter, E. E. (2009). Bayesian meta-analysis of hormone therapy and mortality in younger postmenopausal women. *The American Journal of Medicine*, 122(11), 1016-1022. doi:10.1016/j.amjmed.2009.05.021

Schmid, C. H., & Mengersen, K. (2013). Bayesian meta-analysis. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *The handbook of meta-analysis in ecology and evolution* (pp. 145-173). Princeton, NJ: Princeton University Press.

*Schotter, E. R. (2013). Synonyms provide semantic preview benefit in English. *Journal of Memory and Language*, 69(4), 619-633. doi:10.1016/j.jml.2013.09.002

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, 74(1), 5-35. doi:10.3758/s13414-011-0219-2

*Schotter, E. R., Lee, M., Reiderman, M., & Rayner, K. (2015). The effect of contextual constraint on parafoveal processing in reading. *Journal of Memory and Language*, 83, 118-139. doi:10.1016/j.jml.2015.04.005

Schwarzer, G. (2015). *meta: General package for meta-analysis*. R package version 4.3-2. Retrieved from <http://CRAN.R-project.org/package=meta>

*Serenó, S. C., & Rayner, K. (2000). Spelling-sound regularity effects on eye fixations in reading. *Perception & Psychophysics*, 62(2), 402-409. doi:10.3758/BF03205559

*Shahid, A. (2014). Using word frequency and parafoveal preview to determine the locus of contextual predictability and imageability effects: Evidence from eye movements during reading and lexical decision. Doctoral dissertation: University of Glasgow.

Slattery, T. J., Angele, B., & Rayner, K. (2011). Eye movements and display change detection

- during reading. *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), 1924-1938. doi:10.1037/a0024322
- *Starr, M., & Inhoff, A. (2004). Attention allocation to the right and left of a fixated word: Use of orthographic information from multiple words during reading. *European Journal of Cognitive Psychology*, 16(1-2), 203-225. doi:10.1080/09541440340000150
- Sterne, J. A., Sutton, A. J., Ioannidis, J., Terrin, N., Jones, D. R., Lau, J., ... & Tetzlaff, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343. doi:10.1136/bmj.d4002
- Sutton, A. J., & Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, 10(4), 277-303. doi:10.1177/096228020101000404
- Thompson, S. G., & Higgins, J. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21(11), 1559-1573. doi:10.1002/sim.1187
- *Tiffin-Richards, S. P., & Schroeder, S. (2015). Children's and adults' parafoveal processes in German: Phonological and orthographic effects. *Journal of Cognitive Psychology*, 27(5) 1-18. doi:10.1080/20445911.2014.999076
- Trukenbrod, H. A., & Engbert, R. (2014). ICAT: A computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review*, 21(4), 907-934. doi:10.3758/s13423-013-0575-0
- *Tsai, J. L., Kliegl, R., & Yan, M. (2012). Parafoveal semantic information extraction in traditional Chinese reading. *Acta Psychologica*, 141(1), 17-23. doi:10.1016/j.actpsy.2012.06.004
- Van Elk, M., Matzke, D., Gronau, Q. F., Guan, M., Vandekerckhove, J., & Wagenmakers, E. J.

- (2015). Meta-analyses are no substitute for registered replications: A skeptical perspective on religious priming. *Frontiers in Psychology*, 6, 1365.
doi:10.3389/fpsyg.2015.01365
- Vasilev, M. R. (2013). Negative results in European psychology journals. *Europe's Journal of Psychology*, 9(4), 717-730. doi:10.5964/ejop.v9i4.590
- Vasishth, S., Chen, Z., Li, Q., & Guo, G. (2013). Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PloS one*, 8(10), e77006.
doi:10.1371/journal.pone.0077006
- *Veldre, A., & Andrews, S. (2015a). Parafoveal preview benefit is modulated by the precision of skilled readers' lexical representations. *Journal of Experimental Psychology: Human Perception and Performance*, 41(1), 219-232. doi:10.1037/xhp0000017
- *Veldre, A., & Andrews, S. (2015b). Parafoveal lexical activation depends on skilled reading proficiency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2), 586-595. doi:10.1037/xlm0000039
- von der Malsburg, T., & Angele, B. (2015). False positive rates in standard analyses of eye movements in reading. *arXiv preprint*. arXiv:1504.06896.
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5), 779-804. doi:10.3758/BF03194105
- Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, 60(3), 158-189. doi:10.1016/j.cogpsych.2009.12.001
- *Wakeford, L., & Murray, W. (n.d.). Effects of parafoveal plausibility during reading. *Unpublished manuscript*.

- *Wang, C. A., & Inhoff, A. W. (2010). The influence of visual contrast and case changes on parafoveal preview benefits during reading. *The Quarterly Journal of Experimental Psychology*, 63(4), 805-817. doi:10.1080/17470210903147494
- *Wang, C. A., & Inhoff, A. W. (2013). Extraction of linguistic information from successive words during reading: Evidence for spatially distributed lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3), 662-677. doi:10.1037/a0030242
- *Wang, C. A., Inhoff, A. W., & Radach, R. (2009). Is attention confined to one word at a time? The spatial distribution of parafoveal preview benefits during reading. *Attention, Perception, & Psychophysics*, 71(7), 1487-1494. doi:10.3758/APP.71.7.1487
- Welton, N. J., Sutton, A. J., & Cooper, N. (2012). *Statistics in practice: Evidence synthesis for decision making in healthcare*. Somerset, UK: Wiley.
- *White, S. J., Rayner, K., & Liversedge, S. P. (2005). Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A reexamination. *Psychonomic Bulletin & Review*, 12(5), 891-896. doi:10.3758/BF03196782
- *White, S. J., Warren, T., & Reichle, E. D. (2011). Parafoveal preview during reading: Effects of sentence position. *Journal of Experimental Psychology: Human Perception and Performance*, 37(4), 1221-1238. doi:10.1037/a0022190
- *Williams, C. C., Perea, M., Pollatsek, A., & Rayner, K. (2006). Previewing the neighborhood: The role of orthographic neighbors as parafoveal previews in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4), 1072-1082. doi:10.1037/0096-1523.32.4.1072
- *Winkel, H. (2011). Orthographic and phonological parafoveal processing of consonants,

vowels, and tones when reading Thai. *Applied Psycholinguistics*, 32(04), 739-759.

doi:10.1017/S014271641100004X

- *Winskel, H., & Perea, M. (2013). Consonant/vowel asymmetries in letter position coding during normal reading: Evidence from parafoveal previews in Thai. *Journal of Cognitive Psychology*, 25(1), 119-130. doi:10.1080/20445911.2012.753077
- *Winskel, H., & Salehuddin, K. (2014). Morphological parafoveal preview benefit effects when reading derived words in Malay. *Kajian Malaysia*, 32(2), 23-40.
- *Yan, M. (2015). Visually complex foveal words increase the amount of parafoveal information acquired. *Vision Research*, 111, 91-96. doi:10.1016/j.visres.2015.03.025
- *Yan, M., Kliegl, R., Shu, H., Pan, J., & Zhou, X. (2010). Parafoveal load of word N+1 modulates preprocessing effectiveness of word N+2 in Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 36(6), 1669-1676. doi:10.1037/a0019329
- *Yan, M., Luo, Y., & Inhoff, A. W. (2014). Syllable articulation influences foveal and parafoveal processing of words during the silent reading of Chinese sentences. *Journal of Memory and Language*, 75, 93-103. doi:10.1016/j.jml.2014.05.007
- *Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review*, 16(3), 561-566. doi:10.3758/PBR.16.3.561
- Yan, M., Risse, S., Zhou, X., & Kliegl, R. (2012). Preview fixation duration modulates identical and semantic preview benefit in Chinese reading. *Reading and Writing*, 25(5), 1093-1111. doi:10.1007/s11145-010-9274-7
- *Yan, M., Zhou, W., Shu, H., & Kliegl, R. (2012). Lexical and sublexical semantic preview

- benefits in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 1069-1075. doi:10.1037/a0026935.
- *Yang, J. (2013). Preview effects of plausibility and character order in reading Chinese transposed words: Evidence from eye movements. *Journal of Research in Reading*, 36(S1), S18-S34. doi:10.1111/j.1467-9817.2013.01553.x
- *Yang, J., Li, N., Wang, S., Slattery, T. J., & Rayner, K. (2014). Encoding the target or the plausible preview word? The nature of the plausibility preview benefit in reading Chinese. *Visual Cognition*, 22(2), 193-213. doi:10.1080/13506285.2014.890689
- *Yang, J., Rayner, K., Li, N., & Wang, S. (2012a). Is preview benefit from word n+2 a common effect in reading Chinese? Evidence from eye movements. *Reading and Writing*, 25(5), 1079-1091. doi:10.1007/s11145-010-9282-7
- *Yang, J., Wang, S., Tong, X., & Rayner, K. (2012b). Semantic and plausibility effects on preview benefit during eye fixations in Chinese reading. *Reading and Writing*, 25(5), 1031-1052. doi:10.1007/s11145-010-9281-8
- *Yang, J., Wang, S., Xu, Y., & Rayner, K. (2009). Do Chinese readers obtain preview benefit from word n+ 2? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4), 1192-1204. doi:10.1037/a0013554
- *Yen, M. H., Tsai, J. L., Tzeng, O. J., & Hung, D. L. (2008). Eye movements and parafoveal word processing in reading Chinese. *Memory & Cognition*, 36(5), 1033-1045. doi:10.3758/MC.36.5.1033
- Zang, C., Wang, Y., Bai, X., Yan, G., Drieghe, D., & Liversedge, S. P. (2015). The use of probabilistic lexicality cues for word segmentation in Chinese reading. *The Quarterly Journal of Experimental Psychology*, 69(3), 1-13. doi:10.1080/17470218.2015.1061030

Table S1.1

Summary of N+1 Studies and Their Descriptive Statistics for FFD

Paper	N	Language	ORTH		PHON		SEM		UNREL		PSEUD		RAN		X	
			T	S.E	T	S.E	T	S.E	T	S.E	T	S.E	T	S.E	T	S.E
Gu et al. (2015)	48	Chinese	—	—	—	—	—	—	—	—	35	10.6	—	—	—	—
Schotter et al. (2015), Exp.2	72	English	—	—	—	—	9.5	5.7	16	5.9	—	—	—	—	—	—
Tiffin-Richards & Schroeder (2015)	23	German	7	10.6	11	10.6	—	—	—	—	—	—	—	—	—	—
Veldre & Andrews (2015a)	107	English	—	—	—	—	—	—	—	—	—	—	50	4.5	—	—
Veldre & Andrews (2015b)	94	English	13	4.6	—	—	—	—	—	—	12	4.7	—	—	—	—
Yan (2015)	42	Chinese	—	—	—	—	—	—	27	9.9	20.5	10.0	—	—	—	—
Angele et al. (2014)	35	English	—	—	—	—	—	—	—	—	—	—	31	17.9	—	—
Choi & Gordon (2014)	48	English	8	8.1	15	6.9	—	—	—	—	—	—	—	—	—	—
Cui et al. (2014)	96	Chinese	—	—	—	—	—	—	—	—	17.5	6.0	—	—	—	—
Hohenstein & Kliegl (2014), Exp.3	48	German	—	—	—	—	18	17.6	39	17.5	50	17.5	—	—	—	—
Rayner & Schotter (2014a)	60	English	—	—	—	—	9	7.2	6	6.7	—	—	—	—	—	—
Shahid (2014), Exp.4	80	English	—	—	—	—	—	—	—	—	43.7	6.0	—	—	—	—
Winkel & Salehuddin (2014) [†]	30	Malay	—	—	—	—	—	—	—	—	—	—	—	—	33	12.3
Yan et al. (2014)	36	Chinese	—	—	—	—	—	—	—	—	3.5	8.3	—	—	—	—
Yang et al. (2014)	39	Chinese	—	—	—	—	—	—	8	8.8	—	—	—	—	—	—
Angele & Rayner (2013a)	30	English	—	—	—	—	—	—	—	—	—	—	22	5.6	—	—
Angele & Rayner (2013b), Exp. 2	40	English	—	—	—	—	—	—	—	—	—	—	12	18.2	—	—
Angele et al. (2013c)	40	English	—	—	—	—	—	—	33	22.6	—	—	40	23.3	—	—
Angele et al. (2013c), Exp.2	40	English	—	—	—	—	—	—	—	—	14	18.0	—	—	—	—
Belanger et al. (2013)	20	English	21	26.2	10.5	23.8	—	—	16.5	27.8	—	—	—	—	—	—
Dare & Shillcock (2013), Exp.2	30	English	—	—	—	—	—	—	41.8	8.9	—	—	—	—	—	—
Schotter (2013)	36	English	—	—	—	—	—	—	11	6.9	—	—	—	—	—	—
Schotter (2013), Exp.2	40	English	—	—	—	—	16	8.5	11	7.8	—	—	—	—	—	—
Wang & Inhoff	29	English	—	—	—	—	—	—	—	—	—	—	29	7.2	—	—

[illegible]

Rayner et al. (2007)	30	English	—	—	—	—	—	—	32	10.8	—	—	36	11.2	—	—
Rayner et al. (2007), Exp. 2	36	English	—	—	—	—	—	—	29	13.1	—	—	35	11.9	—	—
McDonald (2006)	16	English	—	—	—	—	—	—	—	—	—	—	22	12.0	—	—
Miller et al. (2006)	42	English	—	—	—	—	—	—	—	—	—	—	22	9.4	—	—
Miller et al. (2006), Exp.2	30	English	—	—	—	—	—	—	—	—	—	—	—	—	46	12.0
Williams et al. (2006)	18	English	15	15.3	—	—	—	—	—	—	—	—	—	—	—	—
Williams et al. (2006), Exp.2	24	English	27	14.5	—	—	—	—	—	—	—	—	—	—	—	—
Chace et al. (2005)	23	English	14	8.7	7	8.7	—	—	—	—	—	—	23	8.7	—	—
Deutsch et al. (2005)	36	Hebrew	16	4.5	—	—	—	—	—	—	—	—	—	—	—	—
Hyönä & Häikiö (2005)	22	Finnish	—	—	—	—	—	—	19	13.7	—	—	—	—	—	—
White et al. (2005)	48	English	—	—	—	—	—	—	—	—	—	—	60	23.2	—	—
Kambe (2004) [†]	36	English	—	—	—	—	—	—	—	—	—	—	22.6	14.5	—	—
Kambe (2004), Exp.2 [†]	24	English	—	—	—	—	—	—	—	—	—	—	—	—	44	13.8
Mielliet & Sparrow (2004) [†]	13	French	—	—	11	21.0	—	—	—	—	—	—	—	—	—	—
Starr & Inhoff (2004)	24	English	—	—	—	—	—	—	26	23.3	—	—	22	21.9	—	—
Deutsch et al. (2003)	30	Hebrew	13	6.9	—	—	—	—	—	—	—	—	—	—	—	—
Inhoff et al. (2000)	20	English	—	—	—	—	—	—	48	15.8	—	—	82	21.9	—	—
Sereno & Rayner (2000) [†]	32	English	—	—	—	—	—	—	—	—	—	—	19.7	15.4	—	—
Binder et al. (1999) [†]	24	English	—	—	—	—	—	—	15	16.7	—	—	—	—	—	—
Inhoff et al. (1998)	15	English	—	—	—	—	—	—	—	—	—	—	28	14.2	—	—
Brihl & Inhoff (1995), Exp.2	24	English	—	—	—	—	—	—	—	—	—	—	—	—	24	9.5
Kennison & Clifton (1995) [†]	48	English	—	—	—	—	—	—	—	—	—	—	23	12.6	—	—
Pollatsek et al. (1992), Exp. 2 [†]	40	English	35	10.8	15	12.0	—	—	41	12.9	—	—	—	—	—	—
Henderson & Ferreira (1990) [†]	12	English	—	—	—	—	—	—	—	—	—	—	2	25.1	—	—
Inhoff & Tousman (1990)	16	English	—	—	—	—	—	—	—	—	—	—	—	—	29.5	17.2
Lima (1987) [†]	18	English	—	—	—	—	—	—	—	—	—	—	—	—	4	15.9
Lima (1987), Exp.2 [†]	24	English	—	—	—	—	—	—	—	—	—	—	18.5	17.8	—	—
Rayner et al. (1986) [†]	24	English	1	14.0	—	—	16	15.5	20	16.7	—	—	—	—	—	—
Balota et al.	30	English	2	12.5	—	—	—	—	19	14.9	—	—	—	—	—	—

(1985) [†]																
Wakeford & Murray (n.d) [†]	28	English	—	—	—	—	—	—	10	15.4	—	—	—	—	—	—

Note: The study is experiment one, unless otherwise indicated. N: number of participants. T: effect size in milliseconds. S.E: Standard error. ORTH: orthographical mask. PHON: phonological mask. SEM: semantic mask. UNREL: unrelated word mask. PSEUD: pseudo-word mask. RAN: random letters mask. X: string of Xs mask.

[†] imputed variance

Table S1.2

Summary of N+1 Studies and Their Descriptive Statistics for SFD

Paper	N	Language	ORTH		PHON		SEM		UNREL		PSEUD		RAN		X	
			T	S.E	T	S.E	T	S.E	T	S.E	T	S.E	T	S.E	T	S ²
Schotter et al. (2015), Exp.2	72	English	—	—	—	—	13	6.7	21.5	7.1	—	—	—	—	—	—
Tiffin-Richards & Schroeder (2015)	23	German	14	12.0	11	12.0	—	—	—	—	—	—	—	—	—	—
Veldre & Andrews (2015a)	107	English	—	—	—	—	—	—	—	—	—	—	64	4.9	—	—
Veldre & Andrews (2015b)	94	English	13	5.2	—	—	—	—	—	—	15	5.5	—	—	—	—
Yan (2015)	42	Chinese	—	—	—	—	—	—	33.5	11.6	23	11.1	—	—	—	—
Choi & Gordon (2014)	48	English	5	7.7	10	7.8	—	—	—	—	—	—	—	—	—	—
Hohenstein & Kliegl (2014), Exp.3	48	German	—	—	—	—	20	17.9	46	17.7	56	17.6	—	—	—	—
Rayner & Schotter (2014a)	60	English	—	—	—	—	13	8.1	7	7.1	—	—	—	—	—	—
Rayner et al. (2014b)	24	English	26	14.1	—	—	41	14.9	55	15.7	—	—	—	—	—	—
Shahid (2014), Exp.4	80	English	—	—	—	—	—	—	—	—	53.5	7.1	—	—	—	—
Yan et al. (2014)	36	Chinese	—	—	—	—	—	—	—	—	5	8.6	—	—	—	—
Yang et al. (2014)	39	Chinese	—	—	—	—	—	—	8.5	11.4	—	—	—	—	—	—
Angele et al. (2013c)	40	English	—	—	—	—	—	—	41	22.4	—	—	46	22.5	—	—
Angele et al. (2013c), Exp.2	40	English	—	—	—	—	—	—	—	—	14	17.8	—	—	—	—
Belanger et al. (2013)	20	English	20.5	26.1	13	24.0	—	—	13	28.0	—	—	—	—	—	—
Schotter (2013)	36	English	—	—	—	—	—	—	17	7.9	—	—	—	—	—	—

Schotter (2013), Exp.2	40	English	—	—	—	—	20	9.1	14	8.2	—	—	—	—	—	—
Winkel & Perea (2013) [†]	33	Thai	15	11.8	—	—	—	—	—	—	—	—	—	—	—	—
Johnson & Dunne (2012)	30	English	23	7.3	—	—	—	—	—	—	—	—	—	—	—	—
Johnson & Dunne (2012), Exp.2	30	English	19.5	10.6	—	—	—	—	—	—	—	—	—	—	—	—
Payne & Stine-Morrow (2012)	24	English	—	—	—	—	—	—	—	—	—	—	39	15.5	—	—
Plummer & Rayner (2012)	36	English	20.5	8.3	—	—	—	—	—	—	—	—	—	—	—	—
Reingold et al. (2012)	60	English	—	—	—	—	—	—	—	—	44	9.0	—	—	—	—
Tsai et al. (2012)	50	Chinese	34	11.3	50	11.3	28	11.3	50	11.3	—	—	—	—	—	—
Yan et al. (2012)	48	Chinese	—	—	—	—	30	5.6	46	7.2	—	—	—	—	—	—
Yang et al. (2012a)	36	Chinese	—	—	—	—	—	—	19	13.4	—	—	—	—	—	—
Yang et al. (2012b)	24	Chinese	—	—	—	—	25	12.5	28	13.1	—	—	—	—	—	—
Yang et al. (2012b), Exp.2	48	Chinese	—	—	—	—	7	8.3	17	8.9	—	—	—	—	—	—
Angele & Rayner (2011)	32	English	—	—	—	—	—	—	—	—	—	—	16	19.5	—	—
Rayner et al. (2010)	36	English	—	—	—	—	—	—	—	—	—	—	24	12.2	—	—
Angele et al. (2008)	32	English	—	—	—	—	—	—	—	—	—	—	37	12.0	—	—
Johnson et al. (2007) [†]	30	English	22	12.4	—	—	—	—	—	—	—	—	—	—	—	—
Johnson et al. (2007), Exp.2 [†]	35	English	25	11.5	—	—	—	—	—	—	—	—	—	—	—	—
Johnson et al. (2007), Exp.3 [†]	30	English	31	12.4	—	—	—	—	—	—	—	—	—	—	—	—
Miller et al. (2006)	42	English	—	—	—	—	—	—	—	—	—	—	43	13.5	—	—
Miller et al. (2006), Exp.2	30	English	—	—	—	—	—	—	—	—	—	—	—	—	59	12.8
Williams et al. (2006)	18	English	14	18.3	—	—	—	—	—	—	—	—	—	—	—	—
Williams et al. (2006), Exp.2	24	English	29	15.5	—	—	—	—	—	—	—	—	—	—	—	—
Drieghe et al. (2005) [†]	24	English	17	13.8	—	—	—	—	43	16.1	—	—	43	18.6	—	—
White et al. (2005)	48	English	—	—	—	—	—	—	—	—	—	—	66.7	23.4	—	—
Inhoff et al. (2000)	20	English	—	—	—	—	—	—	50	14.4	—	—	90	21.5	—	—

Note: The study is experiment one, unless otherwise indicated. N: number of participants. T: effect size in milliseconds. S.E: Standard error. ORTH: orthographical mask. PHON: phonological mask. SEM: semantic mask. UNREL: unrelated word mask. PSEUD: pseudo-word mask. RAN: random letters mask. X: string of Xs mask.

[†] imputed variance

Table S1.3

Summary of N+1 Studies and Their Descriptive Statistics for GD

Paper	N	Language	ORTH		PHON		SEM		UNREL		PSEUD		RAN		X	
			T	S.E	T	S.E	T	S.E	T	S.E	T	S.E	T	S.E	T	S.E
Gu et al. (2015)	48	Chinese	—	—	—	—	—	—	—	—	72.5	22.3	—	—	—	—
Schotter et al. (2015), Exp.2	72	English	—	—	—	—	12.5	8.0	19	8.0	—	—	—	—	—	—
Tiffin-Richards & Schroeder (2015)	23	German	13	14.1	15	14.1	—	—	—	—	—	—	—	—	—	—
Veldre & Andrews (2015a)	107	English	—	—	—	—	—	—	—	—	—	—	70	5.6	—	—
Veldre & Andrews (2015b)	94	English	28	6.6	—	—	—	—	—	—	22	6.3	—	—	—	—
Yan (2015)	42	Chinese	—	—	—	—	—	—	56	14.8	45.5	14.1	—	—	—	—
Angele et al. (2014)	35	English	—	—	—	—	—	—	—	—	—	—	37	24.9	—	—
Choi & Gordon (2014)	48	English	15	9.3	21	10.7	—	—	—	—	—	—	—	—	—	—
Cui et al. (2014)	96	Chinese	—	—	—	—	—	—	—	—	35	8.1	—	—	—	—
Hohenstein & Kliegl (2014), Exp.3	48	German	—	—	—	—	20	20.0	46	19.3	57	19.2	—	—	—	—
Rayner & Schotter (2014a)	60	English	—	—	—	—	12	9.3	14	8.7	—	—	—	—	—	—
Rayner et al. (2014b)	24	English	25	16.3	—	—	44	17.1	47	17.1	—	—	—	—	—	—
Shahid (2014), Exp.4	80	English	—	—	—	—	—	—	—	—	62.75	8.2	—	—	—	—
Winkel & Salehuddin (2014) [†]	30	Malay	—	—	—	—	—	—	—	—	—	—	—	—	57.5	16.8
Yan et al. (2014)	36	Chinese	—	—	—	—	—	—	—	—	3.5	10.6	—	—	—	—
Yang et al. (2014)	39	Chinese	—	—	—	—	—	—	39	17.7	—	—	—	—	—	—
Angele & Rayner (2013a)	30	English	—	—	—	—	—	—	—	—	—	—	25	6.5	—	—
Angele & Rayner (2013b), Exp. 2	40	English	—	—	—	—	—	—	—	—	—	—	43	27.2	—	—
Angele et al. (2013c)	40	English	—	—	—	—	—	—	39	27.0	—	—	43	26.9	—	—
Angele et al. (2013c), Exp.2	40	English	—	—	—	—	—	—	—	—	19	23.6	—	—	—	—
Belanger et al. (2013)	20	English	17.5	34.1	11.5	32.3	—	—	16	36.5	—	—	—	—	—	—
Dare & Shillcock (2013), Exp.2	30	English	—	—	—	—	—	—	50	12.7	—	—	—	—	—	—
Schotter (2013)	36	English	—	—	—	—	—	—	20	9.0	—	—	—	—	—	—
Schotter (2013),	40	English	—	—	—	—	20	10.3	17	10.2	—	—	—	—	—	—

Exp.2																
Wang & Inhoff (2013)	29	English	—	—	—	—	—	—	—	—	—	—	41	8.2	—	—
Winskel & Perea (2013) [†]	33	Thai	38.5	17.2	—	—	—	—	—	—	—	—	—	—	—	—
Yang (2013)	48	Chinese	—	—	—	—	—	—	48	14.7	—	—	—	—	—	—
Yang (2013), Exp.2	54	Chinese	—	—	—	—	—	—	49	14.2	—	—	—	—	—	—
Johnson & Dunne (2012)	30	English	24	8.4	—	—	—	—	—	—	—	—	—	—	—	—
Johnson & Dunne (2012), Exp.2	30	English	25	11.0	—	—	—	—	—	—	—	—	—	—	—	—
Masserang & Pollatsek (2012) [†]	32	English	31	17.5	—	—	—	—	—	—	—	—	—	—	—	—
Masserang & Pollatsek (2012), Exp.2 [†]	58	English	34.5	13.0	—	—	—	—	—	—	—	—	—	—	—	—
Payne & Stine-Morrow (2012)	24	English	—	—	—	—	—	—	—	—	—	—	41	22.6	—	—
Plummer & Rayner (2012)	36	English	21.7	9.5	—	—	—	—	—	—	—	—	—	—	—	—
Reingold et al. (2012)	60	English	—	—	—	—	—	—	—	—	56.5	13.9	—	—	—	—
Tsai et al. (2012)	50	Chinese	56	14.1	69	14.1	45	14.1	65	14.9	—	—	—	—	—	—
Yan et al. (2012)	48	Chinese	—	—	—	—	58.5	8.5	77	10.8	—	—	—	—	—	—
Yang et al. (2012a)	36	Chinese	—	—	—	—	—	—	75	22.9	—	—	—	—	—	—
Yang et al. (2012b)	24	Chinese	—	—	—	—	31	14.8	30	14.0	—	—	—	—	—	—
Yang et al. (2012b), Exp.2	48	Chinese	—	—	—	—	13	9.8	18.5	9.8	—	—	—	—	—	—
Angele & Rayner (2011)	32	English	—	—	—	—	—	—	—	—	—	—	21	21.9	—	—
White et al. (2011)	32	English	—	—	—	—	—	—	—	—	—	—	7	33.7	—	—
White et al. (2011), Exp.2	36	English	—	—	—	—	—	—	—	—	—	—	50	27.0	—	—
Winskel (2011) [†]	36	Thai	37	16.5	54	19.2	—	—	—	—	—	—	—	—	—	—
Rayner et al. (2010)	36	English	—	—	—	—	—	—	—	—	—	—	37	12.8	—	—
Wang & Inhoff (2010)	25	English	—	—	—	—	—	—	—	—	54	18.4	—	—	—	—
Wang & Inhoff (2010), Exp.2	16	English	—	—	—	—	—	—	—	—	34	18.6	—	—	—	—
Wang et al. (2009)	44	English	—	—	—	—	—	—	—	—	50	12.0	—	—	—	—
Yan et al. (2009)	48	Chinese	53	28.6	56	28.2	49	27.9	75	31.6	—	—	—	—	—	—
Yang et al. (2009)	42	Chinese	—	—	—	—	—	—	39	19.1	—	—	—	—	—	—
Yang et al. (2009), Exp.2	66	Chinese	—	—	—	—	—	—	58	25.7	82.5	26.2	—	—	—	—
Angele et al. (2008)	32	English	—	—	—	—	—	—	—	—	—	—	48	19.0	—	—
Yen et al. (2008)	30	Chinese	—	—	—	—	—	—	61	21.9	67	20.6	—	—	—	—

Johnson et al. (2007) [†]	30	English	34.5	18.0	—	—	—	—	—	—	—	—	—	—	—	—
Johnson et al. (2007), Exp.2 [†]	35	English	31.5	16.7	—	—	—	—	—	—	—	—	—	—	—	—
Johnson et al. (2007), Exp.3 [†]	30	English	28	18.0	—	—	—	—	—	—	—	—	—	—	—	—
Rayner et al. (2007)	30	English	—	—	—	—	—	—	47	15.5	—	—	42	15.5	—	—
Rayner et al. (2007), Exp. 2	36	English	—	—	—	—	—	—	18	16.0	—	—	37	17.0	—	—
McDonald (2006)	16	English	—	—	—	—	—	—	—	—	—	—	35	14.3	—	—
Miller et al. (2006)	42	English	—	—	—	—	—	—	—	—	—	—	28.5	16.5	—	—
Miller et al. (2006), Exp.2	30	English	—	—	—	—	—	—	—	—	—	—	—	—	67	13.0
Williams et al. (2006)	18	English	24.5	21.6	—	—	—	—	—	—	—	—	—	—	—	—
Williams et al. (2006), Exp.2	24	English	24	14.3	—	—	—	—	—	—	—	—	—	—	—	—
Chace et al. (2005)	23	English	21	12.7	12	12.7	—	—	—	—	—	—	26	12.7	—	—
Deutsch et al. (2005)	36	Hebrew	21	5.6	—	—	—	—	—	—	—	—	—	—	—	—
Hyönä & Häikiö (2005)	22	Finnish	—	—	—	—	—	—	27	22.0	—	—	—	—	—	—
White et al. (2005)	48	English	—	—	—	—	—	—	—	—	—	—	76.7	26.7	—	—
Kambe (2004) [†]	36	English	—	—	—	—	—	—	—	—	—	—	25	19.0	—	—
Kambe (2004), Exp.2 [†]	24	English	—	—	—	—	—	—	—	—	—	—	—	—	52.6	18.8
Miellet & Sparrow (2004) [†]	13	French	—	—	12	31.9	—	—	—	—	—	—	—	—	—	—
Starr & Inhoff (2004)	24	English	—	—	—	—	—	—	37	28.4	—	—	39	30.0	—	—
Deutsch et al. (2003)	30	Hebrew	22	12.2	—	—	—	—	—	—	—	—	—	—	—	—
Liu et al. (2002), Exp.2	20	Chinese	—	—	18	20.6	—	—	—	—	50	21.4	—	—	—	—
Inhoff et al. (2000)	20	English	—	—	—	—	—	—	60	15.6	—	—	91	20.6	—	—
Sereno & Rayner (2000) [†]	32	English	—	—	—	—	—	—	—	—	—	—	26.7	20.1	—	—
Binder et al. (1999) [†]	24	English	—	—	—	—	—	—	21.5	23.7	—	—	—	—	—	—
Inhoff et al. (1998)	15	English	—	—	—	—	—	—	—	—	—	—	42	18.6	—	—
Briihl & Inhoff (1995), Exp.2	24	English	—	—	—	—	—	—	—	—	—	—	—	—	41	17.7
Kennison & Clifton (1995) [†]	48	English	—	—	—	—	—	—	—	—	—	—	27	16.4	—	—
Pollatsek et al. (1992), Exp. 2 [†]	40	English	25	15.6	11	18.2	—	—	43	18.3	—	—	—	—	—	—
Henderson & Ferreira (1990) [†]	12	English	—	—	—	—	—	—	—	—	—	—	5	32.9	—	—

Inhoff & Tousman (1990)	16	English	—	—	—	—	—	—	—	—	—	—	—	—	45.5	—
Lima (1987) [†]	18	English	—	—	—	—	—	—	—	—	—	—	—	—	18	21.7
Lima (1987), Exp.2 [†]	24	English	—	—	—	—	—	—	—	—	—	—	29	23.2	—	—
Rayner et al. (1986) [†]	24	English	5	20.2	—	—	40	21.3	44	23.7	—	—	—	—	—	—
Balota et al. (1985) [†]	30	English	8	18.0	—	—	—	—	43	21.2	—	—	—	—	—	—
Wakeford & Murray (n.d) [†]	28	English	—	—	—	—	—	—	1	21.9	—	—	—	—	—	—

Note: The study is experiment one, unless otherwise indicated. N: number of participants. T: effect size in milliseconds. S.E: Standard error. ORTH: orthographical mask. PHON: phonological mask. SEM: semantic mask. UNREL: unrelated word mask. PSEUD: pseudo-word mask. RAN: random letters mask. X: string of Xs mask.

[†] imputed variance

Table S1.4

Summary of N+1 Studies and Their Descriptive Statistics for TVT

Paper	N	Language	ORTH		PHON		SEM		UNREL		PSEUD		RAN		X	
			T	S.E	T	S.E	T	S.E	T	S.E	T	S.E	T	S.E	T	S.E
Schotter et al. (2015), Exp.2	72	English	—	—	—	—	31.5	11.4	32	10.5	—	—	—	—	—	—
Veldre & Andrews (2015a)	107	English	—	—	—	—	—	—	—	—	—	—	79.5	8.5	—	—
Veldre & Andrews (2015b)	94	English	19	13.8	—	—	—	—	—	—	22	13.8	—	—	—	—
Angele et al. (2014)	35	English	—	—	—	—	—	—	—	—	—	—	36.5	35.9	—	—
Choi & Gordon (2014)	48	English	24	21.4	37	22.4	—	—	—	—	—	—	—	—	—	—
Rayner & Schotter (2014a)	60	English	—	—	—	—	30	13.9	36	13.6	—	—	—	—	—	—
Shahid (2014), Exp.4	80	English	—	—	—	—	—	—	—	—	76.7	10.4	—	—	—	—
Angele & Rayner (2013b), Exp.2	40	English	—	—	—	—	—	—	—	—	—	—	41	47.2	—	—
Angele et al. (2013c)	40	English	—	—	—	—	—	—	78	39.2	—	—	84	40.4	—	—
Angele et al. (2013c), Exp.2	40	English	—	—	—	—	—	—	—	—	35	40.9	—	—	—	—
Schotter (2013)	36	English	—	—	—	—	—	—	34	15.2	—	—	—	—	—	—

Schotter (2013), Exp.2	40	English	—	—	—	—	28	19.1	25	17.7	—	—	—	—	—	—
Wang & Inhoff (2013)	29	English	—	—	—	—	—	—	—	—	—	—	41	10.2	—	—
Masserang & Pollatsek (2012) [†]	32	English	-7.5	23.8	—	—	—	—	—	—	—	—	—	—	—	—
Masserang & Pollatsek (2012), Exp.2 [†]	58	English	27	17.7	—	—	—	—	—	—	—	—	—	—	—	—
Plummer & Rayner (2012)	36	English	37	18.6	—	—	—	—	—	—	—	—	—	—	—	—
Wang & Inhoff (2010)	25	English	—	—	—	—	—	—	—	—	48	18.6	—	—	—	—
Wang & Inhoff (2010), Exp.2	16	English	—	—	—	—	—	—	—	—	36	20.6	—	—	—	—
Wang et al. (2009)	44	English	—	—	—	—	—	—	—	—	50	13.4	—	—	—	—
Rayner et al. (2007)	30	English	—	—	—	—	—	—	52	26.5	—	—	43	24.9	—	—
Rayner et al. (2007), Exp. 2	36	English	—	—	—	—	—	—	31	20.7	—	—	44	22.2	—	—
McDonald (2006)	16	English	—	—	—	—	—	—	—	—	—	—	34	35.8	—	—
Williams et al. (2006)	18	English	44	41.1	—	—	—	—	—	—	—	—	—	—	—	—
Williams et al. (2006), Exp.2	24	English	35.5	17.4	—	—	—	—	—	—	—	—	—	—	—	—
Hyönä & Häikiö (2005)	22	Finnish	—	—	—	—	—	—	32	24.0	—	—	—	—	—	—
Deutsch et al. (2003)	30	Hebrew	28	23.5	—	—	—	—	—	—	—	—	—	—	—	—
Inhoff et al. (2000)	20	English	—	—	—	—	—	—	81	18.6	—	—	139	28.2	—	—

Note: The study is experiment one, unless otherwise indicated. N: number of participants. T: effect size in milliseconds. S.E: Standard error. ORTH: orthographical mask. PHON: phonological mask. SEM: semantic mask. UNREL: unrelated word mask. PSEUD: pseudo-word mask. RAN: random letters mask. X: string of Xs mask.

[†] imputed variance

Table S1.5

Summary of N+2 Studies and Their Descriptive Statistics for Word N+2

Paper	N	Language	FFD		GD	
			T	S.E	T	S.E
Radach et al. (2013)	36	English	9.5	18.1	16.5	23.0
Yang et al. (2012a)	36	Chinese	4	9.7	0	18.5
Angele & Rayner (2011), Exp. 2	32	English	3	21.1	12.5	27.9
Risse & Kliegl (2011)	40	German	3.2	15.6	6	20.0
Yan et al. (2010)	68	Chinese	8	8.1	14.5	11.8
Yang et al. (2009)	42	Chinese	7	16.6	9	18.7
Yang et al. (2009), Exp.2	66	Chinese	9.5	12.4	25.5	25.1
Angele et al. (2008)	32	English	4	9.6	1	13.5
Kliegl et al. (2007)	30	German	-1	17.2	4.5	28.1
Rayner et al. (2007)	30	English	7	8.6	-7	11.5
Rayner et al. (2007), Exp. 2	36	English	-1	12.6	0	17.0

Note: The study is experiment one, unless otherwise indicated. N: number of participants. T: effect size in milliseconds. S.E: Standard error.

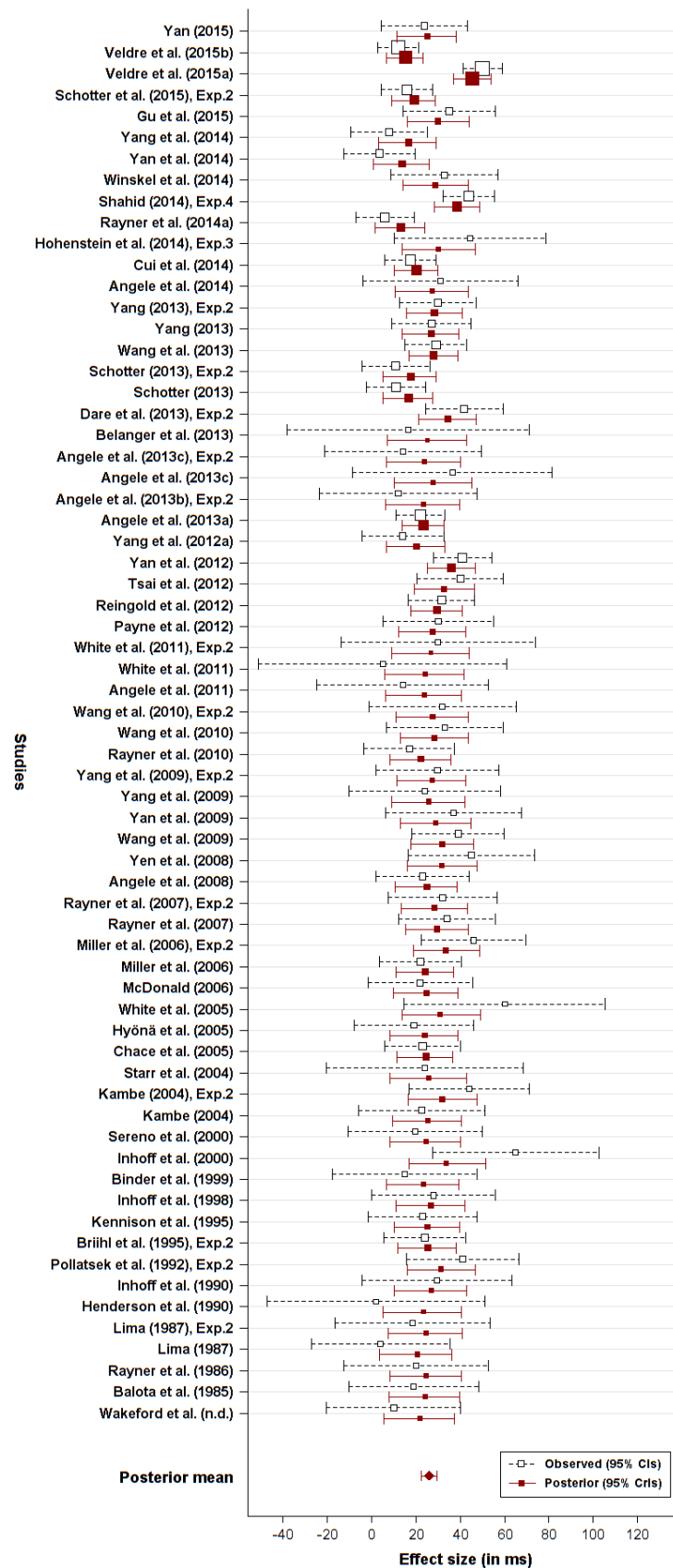


Figure S2.1. Forest plot for the N+1 preview effect for all studies measured with FFD. CIs: confidence intervals. CRIs: credible intervals. The size of squares is proportional to the weight of the study (i.e. the inverse of the within-study variance of the sampling distribution).

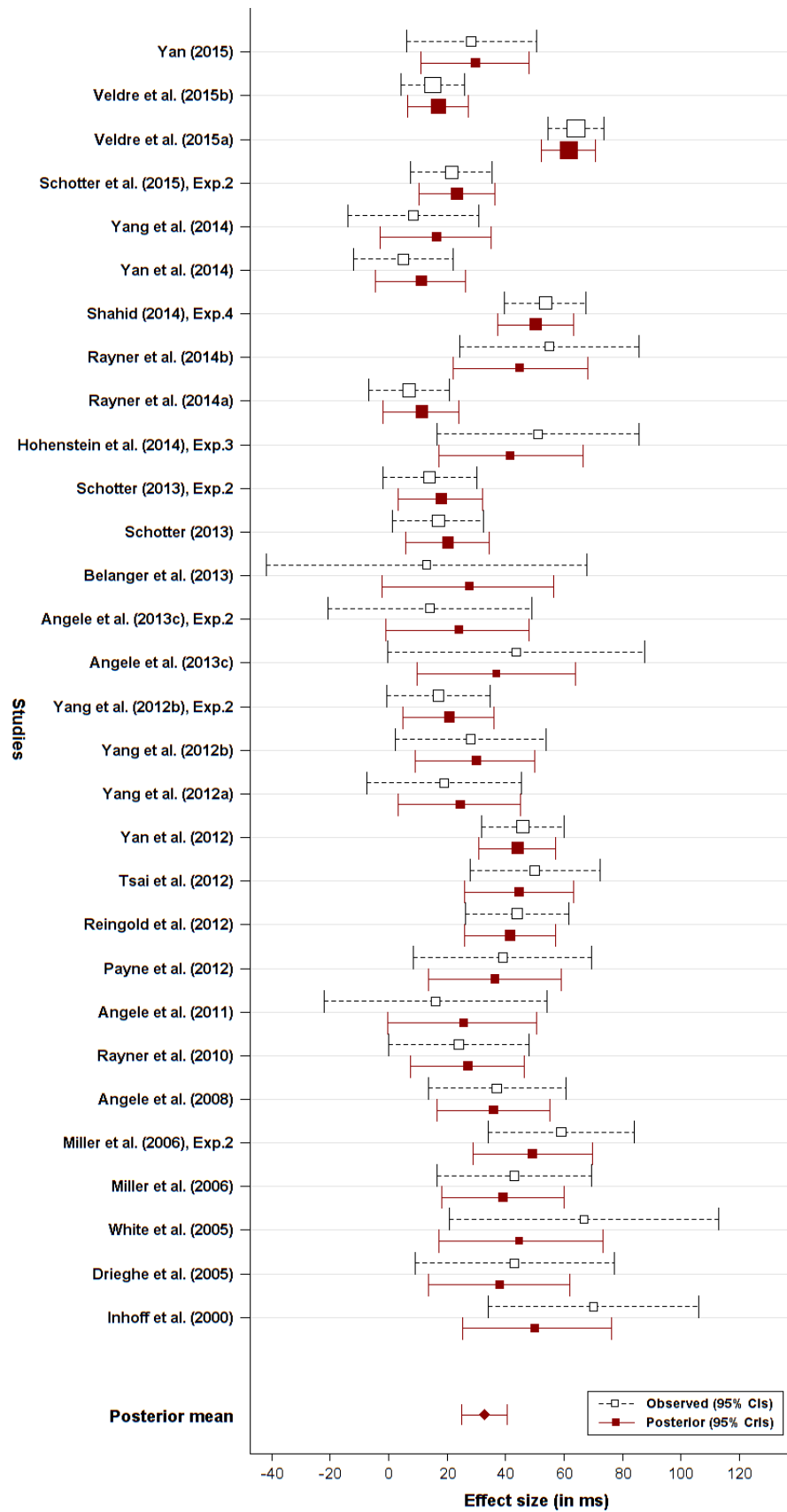


Figure S2.2. Forest plot for the N+1 preview effect for all studies measured with SFD. CIs: confidence intervals. CrIs: credible intervals. The size of squares is proportional to the weight of the study (i.e. the inverse of the within-study variance of the sampling distribution).

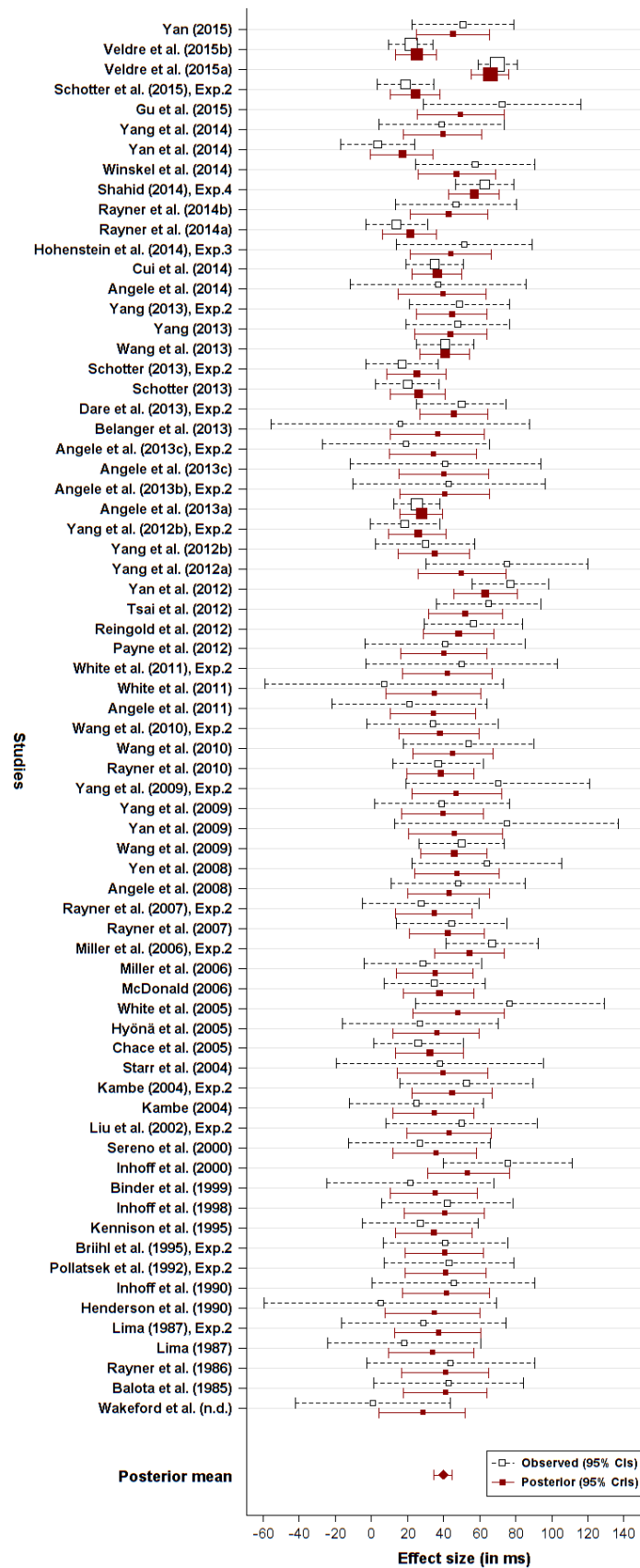


Figure S2.3. Forest plot for the N+1 preview effect for all studies measured with GD. CIs: confidence intervals. CrIs: credible intervals. The size of squares is proportional to the weight of the study (i.e. the inverse of the within-study variance of the sampling distribution).

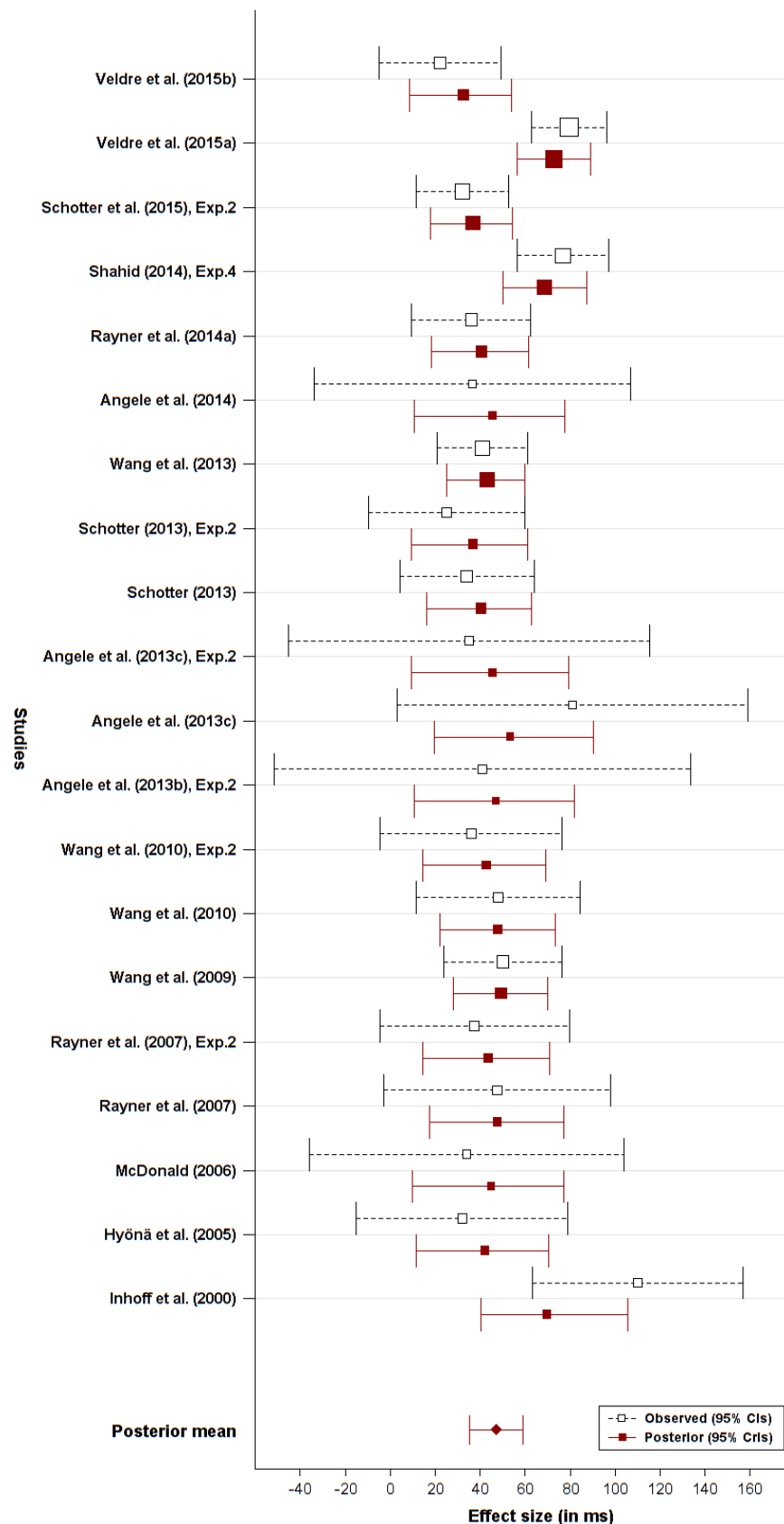


Figure S2.4. Forest plot for the N+1 preview effect for all studies measured with TVT. CIs: confidence intervals. CrIs: credible intervals. The size of squares is proportional to the weight of the study (i.e. the inverse of the within-study variance of the sampling distribution).

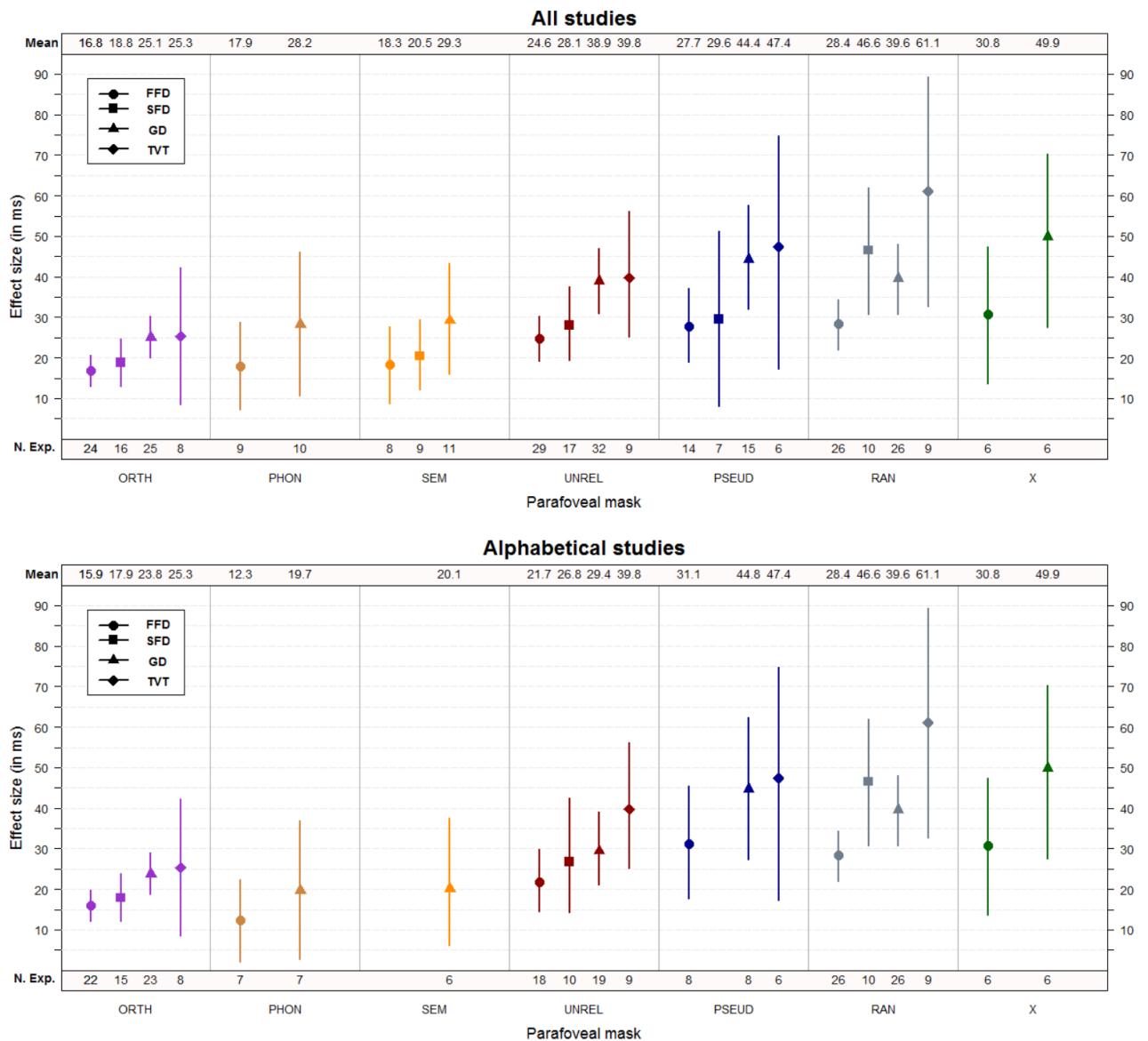


Figure S2.5. N+1 preview effects for all studies and alphabetical-only studies, broken down by mask type. ORTH: orthographical mask. PHON: phonological mask. SEM: semantic mask. UNREL: unrelated word mask. PSEUD: pseudo-word mask. RAN: random letters mask. X: string of Xs mask. N. Exp.: number of experiments on which the analysis was done. P. Mean: posterior mean. Error bars show 95% credible intervals.

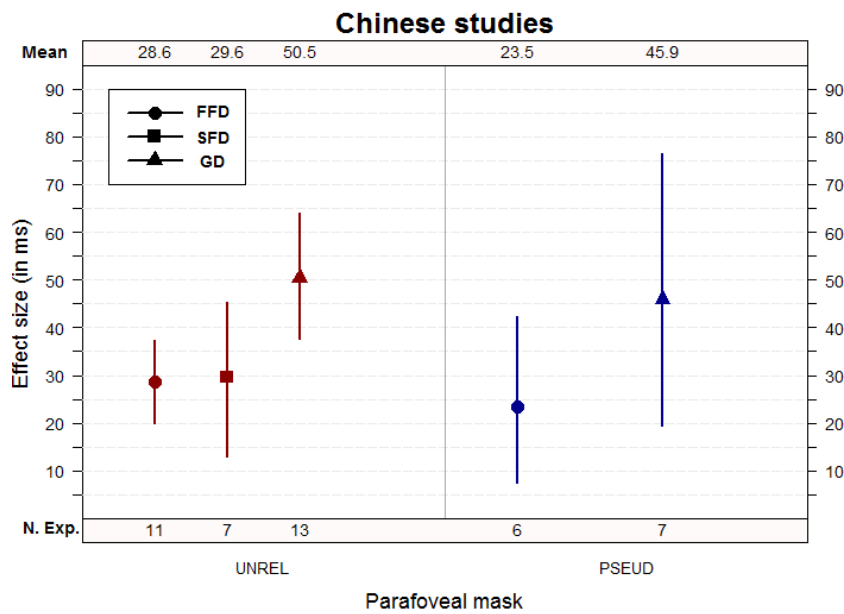


Figure S2.6. N+1 preview effects for Chinese studies, broken down by mask type. N. Exp.: number of experiments on which the analysis was done. P. Mean: posterior mean. UNREL: unrelated word mask. PSEUD: pseudo-word mask. Error bars show 95% credible intervals.

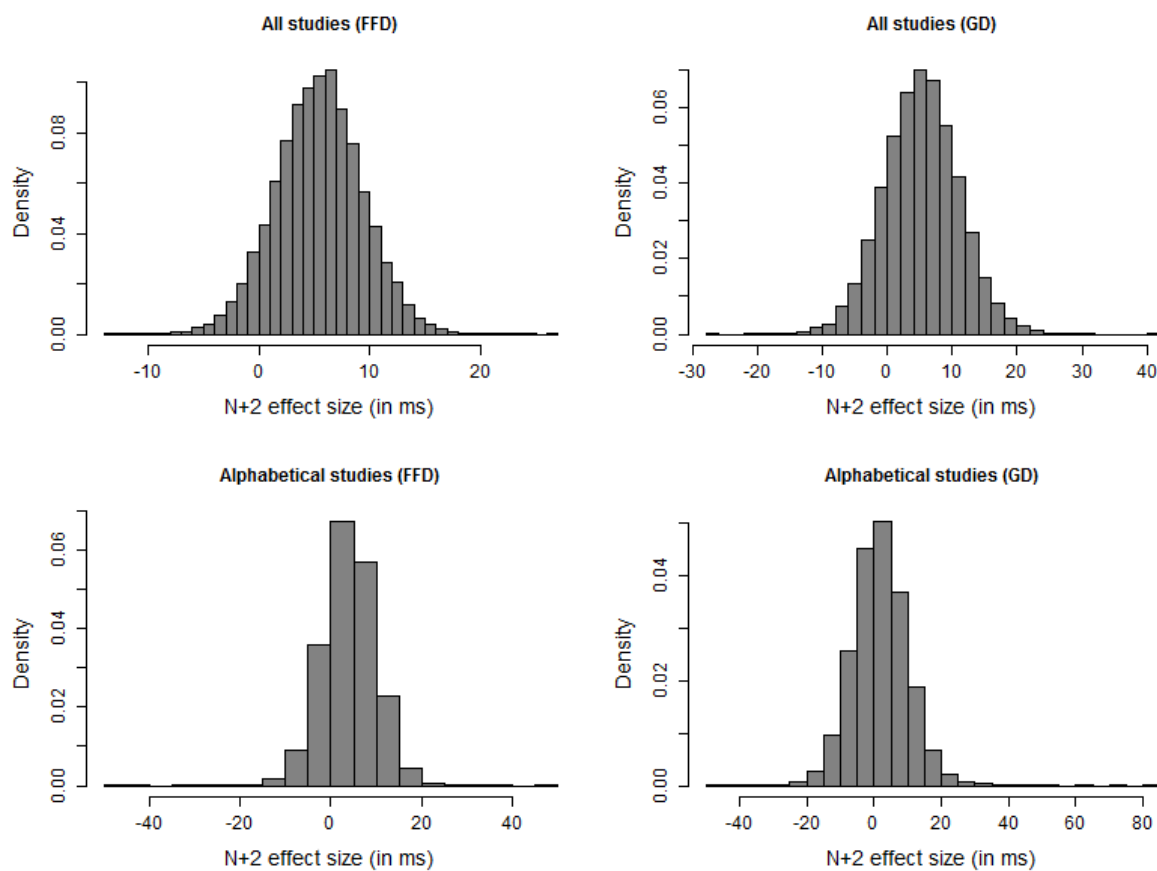


Figure S2.7. Histograms of the posterior distribution for the N+2 preview effects meta-analysis. Number of bins: 30.

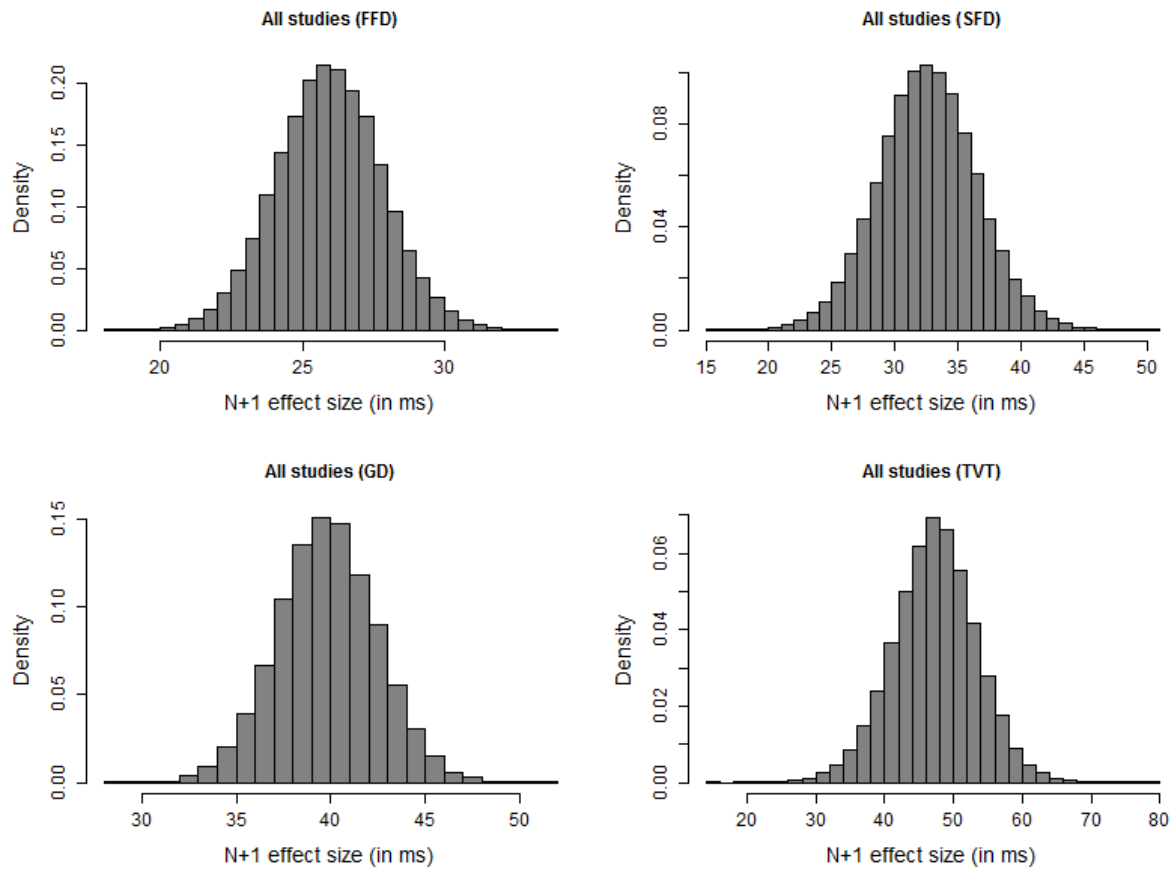


Figure S2.8. Histograms of the posterior distribution for the N+1 preview effect meta-analysis for all studies. Number of bins: 30.

N+2 Preview Effects

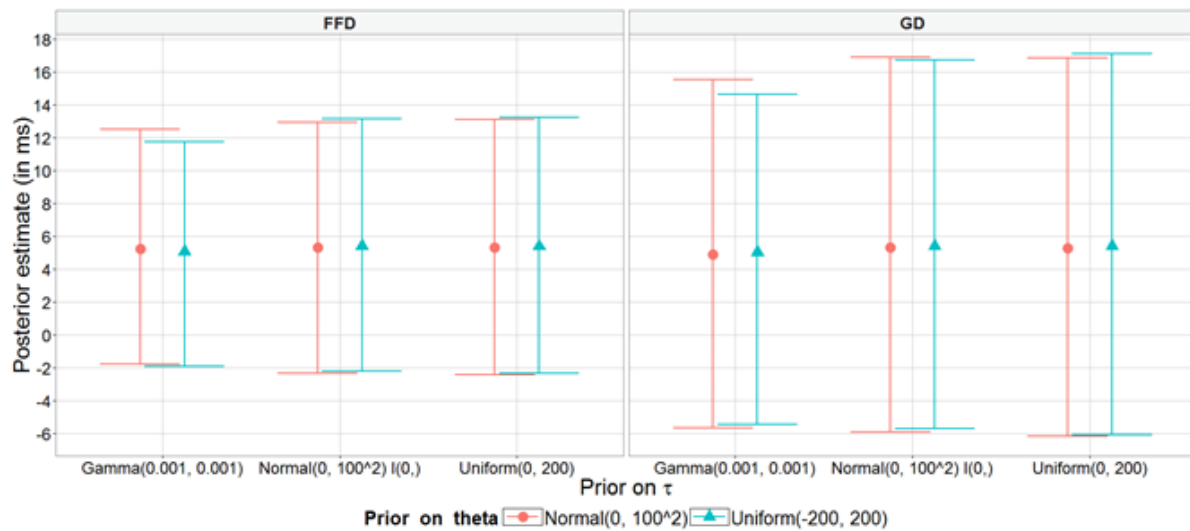


Figure 1. A visual illustration of the effect of different priors on the results from the main N+2 analysis (Tables S3.1 and S3.3). Error bars show 95% credible intervals.

Table S3.1

Mean N+2 Preview Effect on Word N+2 for All Studies [FFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	5.39 (3.96) [-2.34, 13.22]	3.75 (3.19) [0.13, 11.86]
Normal(0, 100^2)	Uniform(0, 200)	5.34 (3.96) [-2.42, 13.1]	3.84 (3.24) [0.16, 12.18]
Uniform(-200, 200)	Gamma(0.001, 0.001)	5.06 (3.55) [-1.91, 11.76]	1.14 (1.74) [0.03, 6.17]
Normal(0, 100^2)	Gamma(0.001, 0.001)	5.25 (3.63) [-1.78, 12.5]	1.16 (1.74) [0.03, 6.22]
Uniform(-200, 200)	Normal(0, 100^2) I(0,.)	5.41 (3.92) [-2.23, 13.17]	3.72 (3.14) [0.1, 11.63]
Normal(0, 100^2)	Normal(0, 100^2) I(0,.)	5.33 (3.9) [-2.34, 12.94]	3.84 (3.29) [0.16, 11.97]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.2

Mean N+2 Preview Effect on Word N+2 Only for Alphabetical Studies [FFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	3.92 (5.92) [-7.6, 15.41]	6.15 (5.7) [0.23, 20.76]
Normal(0, 100^2)	Uniform(0, 200)	3.89 (5.93) [-7.61, 15.68]	6.21 (5.68) [0.25, 21.01]
Uniform(-200, 200)	Gamma(0.001, 0.001)	4.45 (4.95) [-5.2, 14.34]	1.67 (2.71) [0.03, 9.56]
Normal(0, 100^2)	Gamma(0.001, 0.001)	3.85 (4.86) [-5.84, 13.33]	1.62 (2.79) [0.03, 9.4]
Uniform(-200, 200)	Normal(0, 100^2) I(0,.)	4.03 (5.9) [-7.5, 15.46]	6.01 (5.62) [0.18, 20.68]
Normal(0, 100^2)	Normal(0, 100^2) I(0,.)	3.85 (5.92) [-7.85, 15.31]	6.12 (5.63) [0.21, 20.8]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.3

Mean N+2 Preview Effect on Word N+2 for All Studies [GD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	5.4 (5.92) [-6.08, 17.12]	6.42 (5.19) [0.26, 19.3]
Normal(0, 100^2)	Uniform(0, 200)	5.3 (5.81) [-6.18, 16.86]	6.4 (5.22) [0.26, 19.49]
Uniform(-200, 200)	Gamma(0.001, 0.001)	5.02 (5.13) [-5.45, 14.65]	1.83 (2.92) [0.03, 10.51]
Normal(0, 100^2)	Gamma(0.001, 0.001)	4.91 (5.43) [-5.67, 15.54]	1.71 (2.79) [0.03, 9.96]
Uniform(-200, 200)	Normal(0, 100^2) I(0,)	5.4 (5.73) [-5.68, 16.71]	6.36 (5.27) [0.17, 19.48]
Normal(0, 100^2)	Normal(0, 100^2) I(0,)	5.32 (5.81) [-5.92, 16.9]	6.35 (5.32) [0.21, 19.65]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.4

Mean N+2 Preview Effect on Word N+2 Only for Alphabetical Studies [GD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	1.77 (8.26) [-14.1, 18.47]	8.85 (8.02) [0.31, 29.55]
Normal(0, 100^2)	Uniform(0, 200)	1.72 (8.25) [-13.92, 18.39]	8.99 (8.21) [0.28, 30.42]
Uniform(-200, 200)	Gamma(0.001, 0.001)	1.45 (6.51) [-10.74, 14.6]	2.26 (3.81) [0.03, 13.55]
Normal(0, 100^2)	Gamma(0.001, 0.001)	0.99 (6.74) [-12.12, 14.23]	2.16 (3.74) [0.03, 13.17]
Uniform(-200, 200)	Normal(0, 100^2) I(0,)	1.7 (8.14) [-13.99, 18.02]	8.75 (7.94) [0.32, 29.21]
Normal(0, 100^2)	Normal(0, 100^2) I(0,)	1.68 (8.25) [-14.15, 18.32]	8.92 (8.12) [0.31, 29.61]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.5

Mean N+2 Preview Effect on Word N+1 [FFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	6.7 (5.41) [-4, 17.21]	5.53 (4.82) [0.2, 17.67]
Normal(0, 100^2)	Uniform(0, 200)	6.74 (5.41) [-3.91, 17.44]	5.57 (4.89) [0.24, 18.05]
Uniform(-200, 200)	Gamma(0.001, 0.001)	6.5 (4.83) [-2.54, 16.11]	1.66 (2.63) [0.03, 9.23]
Normal(0, 100^2)	Gamma(0.001, 0.001)	6.55 (4.89) [-3, 16]	1.57 (2.5) [0.03, 8.76]
Uniform(-200, 200)	Normal(0, 100^2) I(0,)	6.71 (5.51) [-4.09, 17.46]	5.54 (4.91) [0.2, 17.8]
Normal(0, 100^2)	Normal(0, 100^2) I(0,)	6.69 (5.47) [-4.08, 17.51]	5.56 (4.88) [0.21, 17.88]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.6

Mean N+2 Preview Effect on Word N+1 [GD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	7.44 (6.48) [-5.27, 20.21]	7.01 (6.01) [0.26, 22.09]
Normal(0, 100 ²)	Uniform(0, 200)	7.46 (6.39) [-4.87, 20.19]	6.94 (6.09) [0.24, 22.17]
Uniform(-200, 200)	Gamma(0.001, 0.001)	7.64 (5.57) [-3.44, 18.21]	1.89 (3.11) [0.03, 11.03]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	7.45 (5.62) [-3.72, 18.53]	1.86 (3.06) [0.03, 10.81]
Uniform(-200, 200)	Normal(0, 100 ²) I(0,.)	7.54 (6.5) [-5.27, 20.43]	6.98 (6.02) [0.28, 22.19]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,.)	7.58 (6.41) [-4.77, 20.1]	6.96 (6.08) [0.24, 22.56]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

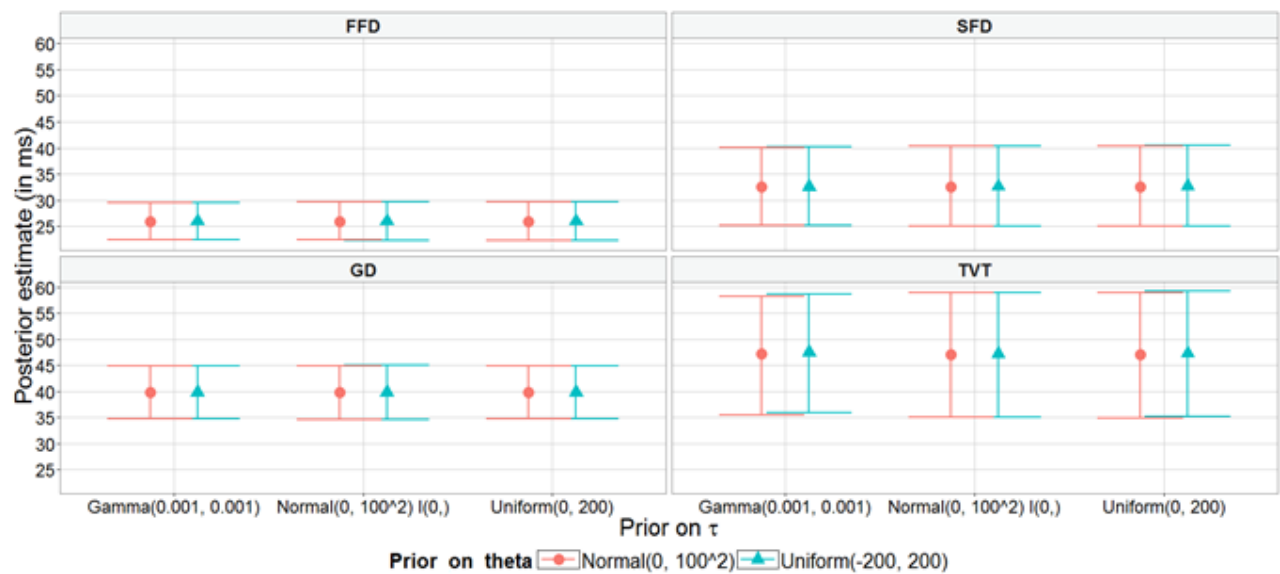
N+1 Preview Effects

Figure 2. A visual illustration of the effect of different priors on the results from the main N+1 analysis of all studies (Tables S3.7 - S3.10). Error bars show 95% credible intervals.

Table S3.7

Mean N+1 Preview Effect for All Languages [FFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	25.97 (1.85) [22.31, 29.64]	9.3 (1.61) [6.39, 12.73]
Normal(0, 100 ²)	Uniform(0, 200)	25.96 (1.86) [22.32, 29.6]	9.32 (1.61) [6.41, 12.76]
Uniform(-200, 200)	Gamma(0.001, 0.001)	25.95 (1.83) [22.36, 29.54]	9.04 (1.58) [6.18, 12.37]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	25.94 (1.86) [22.35, 29.55]	9.04 (1.59) [6.19, 12.42]
Uniform(-200, 200)	Normal(0, 100 ²) I(0,)	25.95 (1.86) [22.29, 29.61]	9.31 (1.61) [6.43, 12.73]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	25.95 (1.85) [22.34, 29.63]	9.3 (1.62) [6.4, 12.76]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.8

Mean N+1 Preview Effect for All Languages [SFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	32.66 (3.91) [25.03, 40.47]	16.99 (3.25) [11.54, 24.25]
Normal(0, 100 ²)	Uniform(0, 200)	32.57 (3.9) [25.03, 40.4]	16.97 (3.27) [11.5, 24.36]
Uniform(-200, 200)	Gamma(0.001, 0.001)	32.58 (3.81) [25.17, 40.17]	16.39 (3.14) [11.14, 23.46]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	32.54 (3.83) [25.11, 40.05]	16.35 (3.15) [11.13, 23.39]
Uniform(-200, 200)	Normal(0, 100 ²) I(0,)	32.61 (3.89) [25.05, 40.42]	16.95 (3.23) [11.54, 24.09]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	32.58 (3.87) [25.09, 40.4]	16.94 (3.24) [11.53, 24.18]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.9

Mean N+1 Preview Effect for All Languages [GD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	39.8 (2.61) [34.68, 44.94]	14.08 (2.22) [10.1, 18.78]
Normal(0, 100 ²)	Uniform(0, 200)	39.76 (2.61) [34.7, 44.92]	14.09 (2.22) [10.11, 18.83]
Uniform(-200, 200)	Gamma(0.001, 0.001)	39.78 (2.59) [34.73, 44.87]	13.74 (2.18) [9.77, 18.33]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	39.74 (2.65) [34.75, 44.85]	13.73 (2.2) [9.79, 18.37]
Uniform(-200, 200)	Normal(0, 100 ²) I(0,)	39.79 (2.63) [34.65, 44.99]	14.09 (2.22) [10.1, 18.82]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	39.76 (2.61) [34.64, 44.94]	14.06 (2.22) [10.08, 18.78]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.10

Mean N+1 Preview Effect for All Languages [TVT]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	47.26 (6.02) [35.24, 59.17]	17.94 (5.61) [8.65, 30.59]
Normal(0, 100 ²)	Uniform(0, 200)	47.12 (6.08) [34.91, 58.91]	18.1 (5.68) [8.72, 30.98]
Uniform(-200, 200)	Gamma(0.001, 0.001)	47.49 (5.72) [35.96, 58.67]	16.07 (5.57) [6.23, 28.19]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	47.26 (5.74) [35.53, 58.18]	16.08 (5.58) [6.11, 28.17]
Uniform(-200, 200)	Normal(0, 100 ²) I(0,)	47.22 (6.02) [35.04, 58.95]	17.99 (5.65) [8.66, 30.84]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	47.12 (6) [35.04, 58.89]	17.95 (5.61) [8.72, 30.76]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.11

Mean N+1 Preview Effect for Alphabetical Studies [FFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	26.12 (2.17) [21.87, 30.38]	9.47 (1.84) [6.2, 13.42]
Normal(0, 100 ²)	Uniform(0, 200)	26.09 (2.18) [21.8, 30.35]	9.47 (1.84) [6.2, 13.44]
Uniform(-200, 200)	Gamma(0.001, 0.001)	26.09 (2.12) [21.91, 30.23]	9.13 (1.8) [5.91, 12.98]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	26.09 (2.16) [21.94, 30.29]	9.14 (1.81) [5.92, 13.02]
Uniform(-200, 200)	Normal(0, 100 ²) I(0,)	26.11 (2.17) [21.85, 30.34]	9.48 (1.85) [6.21, 13.49]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	26.08 (2.17) [21.84, 30.36]	9.49 (1.84) [6.22, 13.44]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.12

Mean N+1 Preview Effect for Alphabetical Studies [SFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	35.71 (4.89) [26.27, 45.59]	18.11 (4.1) [11.57, 27.58]
Normal(0, 100 ²)	Uniform(0, 200)	35.64 (4.86) [26.17, 45.34]	18.04 (4.05) [11.54, 27.25]
Uniform(-200, 200)	Gamma(0.001, 0.001)	35.66 (4.68) [26.57, 44.99]	17.21 (3.85) [10.98, 26.02]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	35.55 (4.72) [26.35, 44.91]	17.21 (3.84) [10.98, 25.96]
Uniform(-200, 200)	Normal(0, 100 ²) I(0,)	35.73 (4.88) [26.24, 45.66]	18.06 (4.09) [11.53, 27.45]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	35.6 (4.83) [26.05, 45.25]	18.02 (4.07) [11.49, 27.44]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.13

Mean N+1 Preview Effect for Alphabetical Studies [GD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	37.74 (2.93) [31.98, 43.51]	13.17 (2.44) [8.87, 18.42]
Normal(0, 100 ²)	Uniform(0, 200)	37.71 (2.92) [31.97, 43.44]	13.17 (2.42) [8.87, 18.37]
Uniform(-200, 200)	Gamma(0.001, 0.001)	37.75 (2.87) [32.08, 43.38]	12.71 (2.39) [8.46, 17.88]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	37.75 (2.91) [32.17, 43.4]	12.69 (2.38) [8.5, 17.81]
Uniform(-200, 200)	Normal(0, 100 ²) I(0,)	37.76 (2.94) [31.96, 43.59]	13.15 (2.43) [8.91, 18.39]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	37.73 (2.91) [32, 43.45]	13.15 (2.42) [8.89, 18.4]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.14

Mean N+1 Preview Effect for Alphabetical Studies [TVT]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	47.26 (6.01) [35.08, 58.84]	18 (5.68) [8.62, 30.88]
Normal(0, 100 ²)	Uniform(0, 200)	47.12 (6.03) [35.04, 58.91]	18.02 (5.67) [8.6, 30.9]
Uniform(-200, 200)	Gamma(0.001, 0.001)	47.46 (5.65) [35.98, 58.42]	16.02 (5.61) [5.76, 28.2]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	47.33 (5.78) [35.61, 58.33]	16.12 (5.59) [6.05, 28.34]
Uniform(-200, 200)	Normal(0, 100 ²) I(0,)	47.23 (6.04) [35.05, 58.88]	17.99 (5.61) [8.77, 30.67]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	47.1 (6.03) [35.08, 58.9]	17.93 (5.59) [8.73, 30.7]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.15

Mean N+1 Preview Effect for Chinese Studies [FFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	25.6 (4.05) [17.73, 33.85]	10.35 (4.31) [2.54, 19.94]
Normal(0, 100 ²)	Uniform(0, 200)	25.55 (4.06) [17.67, 33.84]	10.36 (4.37) [2.58, 20.16]
Uniform(-200, 200)	Gamma(0.001, 0.001)	25.35 (3.57) [18.38, 32.61]	7.18 (4.84) [0.06, 17.12]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	25.3 (3.53) [18.5, 32.5]	7.08 (4.8) [0.07, 16.96]
Uniform(-200, 200)	Normal(0, 100 ²) I(0,)	25.61 (4.06) [17.72, 33.89]	10.35 (4.27) [2.83, 19.83]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	25.59 (4.07) [17.71, 33.9]	10.25 (4.33) [2.12, 19.68]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.16

Mean N+1 Preview Effect for Chinese Studies [SFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	25.46 (7.88) [9.66, 41.23]	17.64 (8.36) [6.08, 38.35]
Normal(0, 100 ²)	Uniform(0, 200)	25.24 (7.81) [9.47, 40.65]	17.53 (8.2) [6.12, 37.8]
Uniform(-200, 200)	Gamma(0.001, 0.001)	25.53 (6.6) [12.13, 38.64]	13.52 (7.3) [0.31, 30.12]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	25.4 (6.52) [12.1, 38.24]	13.55 (7.4) [0.28, 30.3]
Uniform(-200, 200)	Normal(0, 100 ²) I(0,)	25.42 (7.81) [9.74, 41.02]	17.49 (8.34) [6.05, 37.68]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	25.34 (7.63) [9.96, 40.42]	17.28 (7.93) [6.04, 36.54]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.17

Mean N+1 Preview Effect for Chinese Studies [GD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(-200, 200)	Uniform(0, 200)	46.83 (6.46) [34.63, 60.15]	19.5 (5.73) [10.12, 32.75]
Normal(0, 100 ²)	Uniform(0, 200)	46.66 (6.42) [34.37, 59.76]	19.52 (5.84) [10.09, 32.91]
Uniform(-200, 200)	Gamma(0.001, 0.001)	46.57 (6.11) [35.08, 59.17]	17.88 (5.47) [8.71, 30.24]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	46.34 (6.17) [34.81, 58.97]	17.78 (5.53) [8.4, 30.08]
Uniform(-200, 200)	Normal(0, 100 ²) I(0,)	46.82 (6.41) [34.5, 59.99]	19.44 (5.73) [10.12, 32.39]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	46.64 (6.39) [34.49, 59.8]	19.37 (5.72) [10.03, 32.57]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

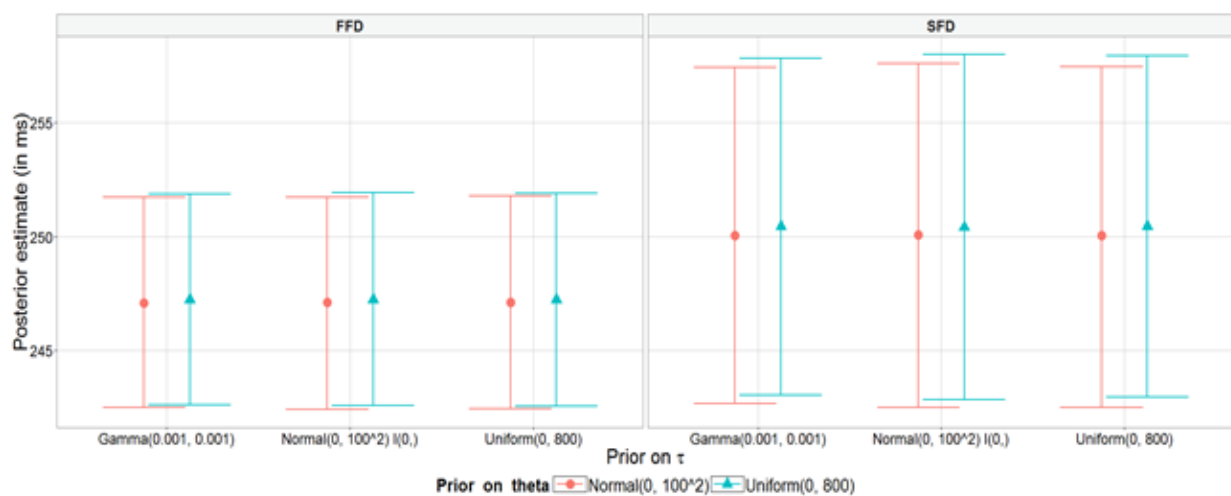
Baseline Fixation Durations

Figure 3a. A visual illustration of the effect of different priors on the results from baseline fixation duration analysis (Tables S3.18 - S3.19). Error bars show 95% credible intervals.

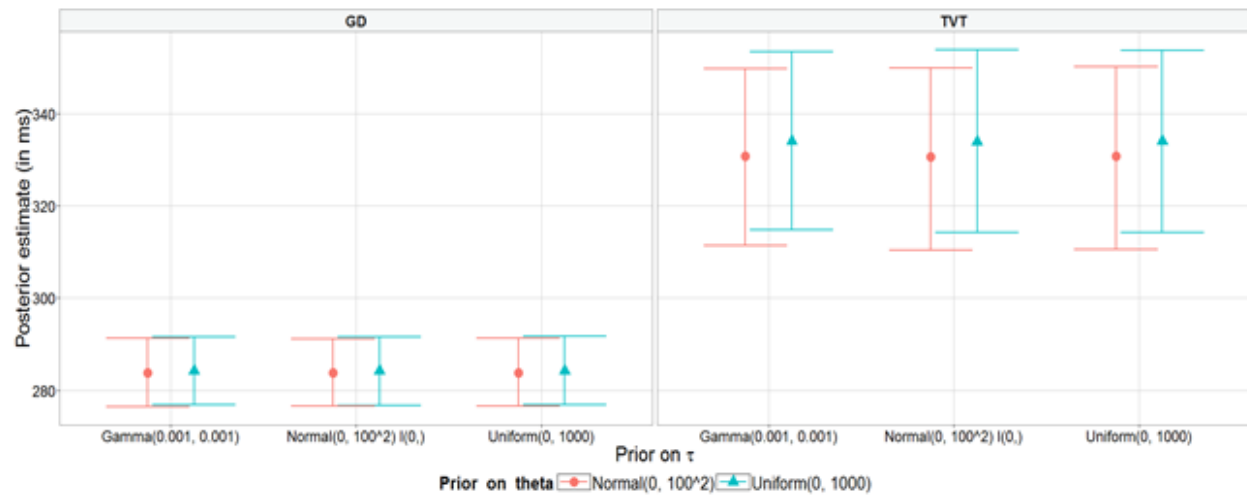


Figure 3b. A visual illustration of the effect of different priors on the results from baseline fixation duration analysis (Tables S3.20 - S3.21). Error bars show 95% credible intervals.

Table S3.18

Mean Baseline Fixation Durations for All Studies [FFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(0, 800)	Uniform(0, 800)	247.23 (2.38) [242.54, 251.9]	21.54 (1.81) [18.28, 25.42]
Normal(0, 100 ²)	Uniform(0, 800)	247.1 (2.37) [242.43, 251.77]	21.53 (1.8) [18.3, 25.33]
Uniform(0, 800)	Gamma(0.001, 0.001)	247.23 (2.37) [242.6, 251.88]	21.39 (1.77) [18.21, 25.16]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	247.07 (3.07) [242.49, 251.73]	21.39 (1.78) [18.23, 25.14]
Uniform(0, 800)	Normal(0, 100 ²) I(0)	247.23 (2.38) [242.57, 251.91]	21.53 (1.8) [18.3, 25.37]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0)	247.1 (2.37) [242.4, 251.72]	21.53 (1.8) [18.3, 25.38]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.19

Mean Baseline Fixation Durations for All Studies [SFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(0, 800)	Uniform(0, 800)	250.45 (3.82) [242.96, 257.93]	23.89 (2.94) [18.89, 30.38]
Normal(0, 100 ²)	Uniform(0, 800)	250.05 (3.8) [242.51, 257.45]	23.87 (2.92) [18.85, 30.31]
Uniform(0, 800)	Gamma(0.001, 0.001)	250.44 (3.74) [243.03, 257.82]	23.52 (2.84) [18.69, 29.79]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	250.05 (4.25) [242.68, 257.42]	23.55 (2.86) [18.69, 29.81]
Uniform(0, 800)	Normal(0, 100 ²) I(0)	250.4 (3.83) [242.83, 257.99]	23.86 (2.91) [18.89, 30.32]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0)	250.07 (3.82) [242.51, 257.59]	23.86 (2.93) [18.86, 30.42]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.20

Mean Baseline Fixation Durations for All Studies [GD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(0, 1000)	Uniform(0, 1000)	284.22 (3.77) [276.81, 291.71]	35.89 (2.83) [30.76, 41.93]
Normal(0, 100 ²)	Uniform(0, 1000)	283.82 (3.76) [276.49, 291.24]	35.91 (2.83) [30.84, 41.97]
Uniform(0, 1000)	Gamma(0.001, 0.001)	284.2 (3.77) [276.81, 291.62]	35.69 (2.8) [30.66, 41.66]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	283.8 (4.4) [276.37, 291.23]	35.7 (2.82) [30.69, 41.66]
Uniform(0, 1000)	Normal(0, 100 ²) I(0,)	284.18 (3.78) [276.75, 291.59]	35.89 (2.81) [30.84, 41.8]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	283.8 (3.74) [276.53, 291.12]	35.88 (2.83) [30.84, 41.88]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.21

Mean Baseline Fixation Durations for All Studies [TVT]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(0, 1000)	Uniform(0, 1000)	333.99 (9.99) [314.23, 353.7]	51.48 (7.69) [38.88, 69]
Normal(0, 100 ²)	Uniform(0, 1000)	330.72 (10.03) [310.46, 350.15]	51.5 (7.78) [38.88, 69.24]
Uniform(0, 1000)	Gamma(0.001, 0.001)	334.06 (9.79) [314.76, 353.44]	50.34 (7.39) [38.3, 67.04]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	330.8 (10.09) [311.32, 349.72]	50.33 (7.41) [38.16, 67.02]
Uniform(0, 1000)	Normal(0, 100 ²) I(0,)	333.92 (9.99) [314.24, 353.86]	51.12 (7.58) [38.74, 68.15]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	330.61 (9.92) [310.44, 349.82]	51.23 (7.59) [38.85, 68.55]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.22

Mean Baseline Fixation Durations for Alphabetical Studies [FFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(0, 800)	Uniform(0, 800)	246.25 (2.69) [240.97, 251.52]	22.29 (2.05) [18.64, 26.65]
Normal(0, 100 ²)	Uniform(0, 800)	246.07 (2.68) [240.85, 251.31]	22.28 (2.03) [18.65, 26.62]
Uniform(0, 800)	Gamma(0.001, 0.001)	246.24 (2.66) [241.07, 251.46]	22.09 (2.01) [18.53, 26.39]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	246.06 (3.31) [240.86, 251.29]	22.08 (2.01) [18.52, 26.42]
Uniform(0, 800)	Normal(0, 100 ²) I(0,)	246.26 (2.68) [240.98, 251.5]	22.24 (2.01) [18.69, 26.54]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	246.07 (2.68) [240.78, 251.37]	22.27 (2.02) [18.67, 26.62]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.23

Mean Baseline Fixation Durations for Alphabetical Studies [SFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(0, 800)	Uniform(0, 800)	248.86 (4.65) [239.63, 258.08]	26.34 (3.57) [20.35, 34.35]
Normal(0, 100^2)	Uniform(0, 800)	248.31 (4.64) [239.15, 257.54]	26.32 (3.6) [20.3, 34.43]
Uniform(0, 800)	Gamma(0.001, 0.001)	248.9 (4.56) [239.87, 257.93]	25.86 (3.47) [20.01, 33.57]
Normal(0, 100^2)	Gamma(0.001, 0.001)	248.32 (4.97) [239.32, 257.22]	25.87 (3.5) [20.04, 33.75]
Uniform(0, 800)	Normal(0, 100^2) I(0,)	248.85 (4.64) [239.61, 258.01]	26.32 (3.6) [20.28, 34.32]
Normal(0, 100^2)	Normal(0, 100^2) I(0,)	248.33 (4.62) [239.18, 257.32]	26.31 (3.57) [20.33, 34.29]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.24

Mean Baseline Fixation Durations for Alphabetical Studies [GD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(0, 1000)	Uniform(0, 1000)	280.47 (3.53) [273.52, 287.44]	30.05 (2.69) [25.26, 35.8]
Normal(0, 100^2)	Uniform(0, 1000)	280.15 (3.54) [273.22, 287.09]	30.03 (2.67) [25.27, 35.75]
Uniform(0, 1000)	Gamma(0.001, 0.001)	280.5 (3.52) [273.57, 287.37]	29.81 (2.64) [25.12, 35.48]
Normal(0, 100^2)	Gamma(0.001, 0.001)	280.13 (4.16) [273.24, 287.04]	29.79 (2.65) [25.09, 35.41]
Uniform(0, 1000)	Normal(0, 100^2) I(0,)	280.48 (3.54) [273.52, 287.49]	30.03 (2.67) [25.31, 35.72]
Normal(0, 100^2)	Normal(0, 100^2) I(0,)	280.12 (3.53) [273.16, 287.1]	30.01 (2.67) [25.3, 35.68]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.25

Mean Baseline Fixation Durations for All Studies [TVT]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(0, 1000)	Uniform(0, 1000)	334 (9.94) [314.36, 353.62]	51.44 (7.71) [38.85, 68.91]
Normal(0, 100^2)	Uniform(0, 1000)	330.75 (10.04) [310.52, 350.21]	51.55 (7.73) [38.96, 69.1]
Uniform(0, 1000)	Gamma(0.001, 0.001)	333.93 (9.77) [314.61, 353.09]	50.37 (7.4) [38.21, 67.04]
Normal(0, 100^2)	Gamma(0.001, 0.001)	330.78 (10.16) [311.23, 349.99]	50.42 (7.42) [38.36, 67.38]
Uniform(0, 1000)	Normal(0, 100^2) I(0,)	333.98 (9.91) [314.34, 353.44]	51.12 (7.57) [38.78, 68.32]
Normal(0, 100^2)	Normal(0, 100^2) I(0,)	330.74 (9.89) [311.18, 349.99]	51.09 (7.56) [38.88, 68.31]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.26

Mean Baseline Fixation Durations for Chinese Studies [FFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(0, 800)	Uniform(0, 800)	252.34 (5.34) [241.62, 262.86]	18.94 (4.68) [11.87, 29.86]
Normal(0, 100 ²)	Uniform(0, 800)	251.6 (5.36) [240.79, 261.97]	18.96 (4.71) [11.87, 30.03]
Uniform(0, 800)	Gamma(0.001, 0.001)	252.35 (5.07) [242.17, 262.34]	17.93 (4.29) [11.3, 27.9]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	251.68 (5.5) [241.35, 261.66]	17.95 (4.32) [11.27, 28.1]
Uniform(0, 800)	Normal(0, 100 ²) I(0,)	252.35 (5.27) [241.8, 262.75]	18.89 (4.65) [11.74, 29.94]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	251.63 (5.35) [240.77, 262.02]	18.99 (4.7) [11.88, 30.3]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.27

Mean Baseline Fixation Durations for Chinese Studies [SFD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(0, 800)	Uniform(0, 800)	256.79 (4.1) [248.85, 265.14]	8.19 (4.92) [0.78, 19.92]
Normal(0, 100 ²)	Uniform(0, 800)	256.39 (4.08) [248.16, 264.56]	8.22 (4.9) [0.83, 20.03]
Uniform(0, 800)	Gamma(0.001, 0.001)	256.51 (2.96) [250.74, 262.57]	3.79 (4.07) [0.04, 13.58]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	256.31 (3.66) [250.4, 262.43]	3.92 (4.11) [0.04, 13.96]
Uniform(0, 800)	Normal(0, 100 ²) I(0,)	256.83 (4.07) [248.86, 265.14]	8.21 (4.91) [0.82, 19.86]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	256.37 (4.1) [248.1, 264.44]	8.2 (5) [0.76, 19.84]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.28

Mean Baseline Fixation Durations for Chinese Studies [GD]

Prior on θ	Prior on τ	Posterior θ (SD) [95% CrI]	Posterior τ (SD) [95% CrI]
Uniform(0, 1000)	Uniform(0, 1000)	300.27 (14.02) [272.27, 328.2]	57.76 (11.32) [40.24, 84.37]
Normal(0, 100 ²)	Uniform(0, 1000)	294.41 (14.05) [265.1, 321.07]	58.02 (11.55) [40.3, 84.95]
Uniform(0, 1000)	Gamma(0.001, 0.001)	300.26 (13.45) [273.75, 327.11]	55.65 (10.67) [39.24, 80.68]
Normal(0, 100 ²)	Gamma(0.001, 0.001)	294.91 (13.87) [266.92, 321.08]	55.96 (10.83) [39.44, 81.43]
Uniform(0, 1000)	Normal(0, 100 ²) I(0,)	300.22 (13.8) [272.97, 327.64]	56.91 (10.89) [39.96, 82.25]
Normal(0, 100 ²)	Normal(0, 100 ²) I(0,)	294.68 (13.99) [265.93, 321.27]	57.15 (11.05) [40.15, 82.96]

Note: For Gamma distributions, the prior is on the precision (i.e. $1/\tau^2$). The priors in bold are the ones on which the results in the paper are based. CrI: credible intervals.

Table S3.29

Mean Difference of the N+1 Preview Effect Between Alphabetical and Chinese Studies

Measure	Prior on β	Posterior β (SD) [95% CrI]
FFD [all masks]	Uniform(-200, 200)	-0.29 (2.15) [-4.45, 4]
FFD [all masks]	Normal(0, 100 ²)	-0.3 (2.16) [-4.52, 3.96]
SFD [all masks]	Uniform(-200, 200)	-5.14 (4.25) [-13.49, 3.23]
SFD [all masks]	Normal(0, 100 ²)	-5.11 (4.25) [-13.53, 3.26]
GD [all masks]	Uniform(-200, 200)	4.09 (3.04) [-1.82, 10.14]
GD [all masks]	Normal(0, 100 ²)	4.03 (3.03) [-1.89, 10.04]
FFD [unrelated word mask]	Uniform(-200, 200)	3.47 (2.78) [-2.21, 8.79]
FFD [unrelated word mask]	Normal(0, 100 ²)	3.46 (2.75) [-2.11, 8.73]
SFD [unrelated word mask]	Uniform(-200, 200)	1.46 (4.73) [-8.53, 10.28]
SFD [unrelated word mask]	Normal(0, 100 ²)	1.45 (4.73) [-8.48, 10.12]
GD [unrelated word mask]	Uniform(-200, 200)	9.92 (3.74) [2.45, 17.25]
GD [unrelated word mask]	Normal(0, 100 ²)	9.9 (3.72) [2.56, 17.2]
FFD [pseudo-word mask]	Uniform(-200, 200)	-3.83 (4.63) [-12.99, 5.47]
FFD [pseudo-word mask]	Normal(0, 100 ²)	-3.8 (4.66) [-12.91, 5.54]
GD [pseudo-word mask]	Uniform(-200, 200)	-0.58 (6.76) [-13.43, 13.45]
GD [pseudo-word mask]	Normal(0, 100 ²)	-0.46 (6.77) [-13.4, 13.42]

Note: The prior in bold is the one on which the results in the paper are based. CrI: credible intervals.

Table S3.30

Mean Difference of the N+2 Preview Effect Between Alphabetical and Chinese Studies

Measure	Prior on β	Posterior β (SD) [95% CrI]
FFD	Uniform(-200, 200)	1.5 (4.02) [-6.29, 9.48]
FFD	Normal(0, 100 ²)	1.49 (4) [-6.36, 9.36]
GD	Uniform(-200, 200)	5.09 (6.02) [-6.75, 16.78]
GD	Normal(0, 100 ²)	5.10 (6.02) [-6.87, 16.8]

Note: The prior in bold is the one on which the results in the paper are based. CrI: credible intervals.