Perceptual Disfluency and Recognition Memory: A Response Time Distributional Analysis

Jason Geller[1], Pablo Gomez[2], Erin Buchanon[3], and Dominique Makowski[4]

[1]Department of Psychology and Neuroscience, Boston College

[2]Department of Psychology, Skidmore College

[3]Department of Cognitive Analytics (Analytics), Harrisburg University of Science and Technology

[4]School of Psychology, University of Sussex

Author Note

Jason Geller https://orcid.org/0000-0002-7459-4505

Pablo Gomez https://orcid.org/0000-0003-4180-1560

Erin Buchanon https://orcid.org/0000-0002-9689-4189

Dominique Makowski https://orcid.org/0000-0002-9689-4189

review & editing, validation, and formal analysis. *Dominique Makowski*: writing – review & editing, validation, and formal analysis

Correspondence concerning this article should be addressed to Jason Geller, Department of Psychology and Neuroscience, Boston College, Mcguinn Hall 405, Chestnut Hill, MA 02467-9991, USA, drjasongeller@gmail.com: jason.geller@bc.edu

Abstract

Perceptual disfluency, induced by blurring or difficult-to-read typefaces, can sometimes enhance memory retention, but the underlying mechanisms remain unclear. To investigate this, we manipulated blurring levels (clear, low blur, high blur) during encoding and assessed recognition performance in a surprise memory test. In Experiments 1A and 1B, response latencies from a lexical decision task were analyzed using ex-Gaussian distribution modeling and drift diffusion modeling. Results showed that blurring differentially influenced parameters oe the model, with high blur affecting both early and late-stage processes, while low blur primarily influenced early-stage processes. Recognition test results further revealed that high-blur words were remembered better than both clear and low-blurred words.Experiment 2 employed a semantic categorization task with a word frequency manipulation to further examine the locus of the perceptual disfluency effect. Similar to Experiments 1A and 1B, high blur influenced both early and late-stage processes, while low blur primarily affected early-stage processes. Low-frequency words exhibited greater shifting and skewing in distributional parameters, yet only high-frequency, highly blurred words demonstrated an enhanced memory effect. These findings suggest that both early and late cognitive processes contribute to the mnemonic benefits associated with perceptual disfluency.Overall, this study demonstrates that distributional and computational analyses provide powerful tools for dissecting encoding mechanisms and their effects on memory, offering valuable insights into models of perceptual disfluency.

Keywords: disfluency, LDT, DDM, ex-Gaussian, distributional analyses, word recognition

Perceptual Disfluency and Recognition Memory: A Response Time Distributional Analysis

We live in a world that, even for adults, is, "blooming and buzzing with confusion" (James, 1890, p. 488). Despite this, we possess the remarkable ability to accomplish perceptual tasks, such as deciphering unclear and unsegmented handwritten cursive words, or actively participating in conversations amidst the chaos of a noisy bar. The ability to deal with a noisy, confusing world, has been a topic of research at the interface between education and cognitive psychology. Decades of work has demonstrated a relationship between encoding difficulty and long-term memory. While making learning something harder and not easier runs counter to people's beliefs, a host of memory and learning effects show that making encoding more effortful (and thus more errorful), under certain circumstances, can be beneficial for memory. This has popularly become known as the "desirable difficulties" principle (Bjork & Bjork, 2011). Well-documented examples of desirable difficulties include spacing out encoding across multiple sessions rather than massing study during a single session (see Carpenter et al., 2022), studying concepts in an interleaved fashion rather than a blocked fashion (Rohrer & Taylor, 2007), and generating or retrieving information rather than simply re-reading or studying the information again (Roediger & Karpicke, 2006; Slamecka & Graf, 1978).

Another desirable difficulty example involves a very simple manipulation: changing the perceptual characteristics of to-be-learned material to make it more difficult to process. A growing literature has shown that manipulating the perceptual characteristics of the to-be-learned material at encoding can improve memory (e.g., Geller et al., 2018; Geller & Peterson, 2021; Halamish, 2018; Rosner et al., 2015). The resulting memory benefit has been called the perceptual disfluency effect (see Geller et al., 2018 for a discussion).

The perceptual disfluency effect

The relationship between perceptual disfluency and long-term memory has a long and storied history. While it is not quite clear where the term perceptual disfluency effect originated

from, the idea behind it goes back to the late 80s with the work of Nairne (1988). Under the term *perceptual-interference effect*, Nairne employed the technique of backward masking with hash marks (e.g., ####) with a quick presentation time to make word encoding noisier during study. Because the word is presented and masked so quickly, considerable effort is needed to identify the word. Since then, different types of perceptual disfluency manipulations have shown to elicit a similar effect, such as high-level blurring (Rosner et al., 2015), word inversion (Sungkhasettee et al., 2011) , small text size (Halamish, 2018), handwritten cursive (Geller et al., 2018), and other unusual or difficult-to-read typefaces (Geller & Peterson, 2021; Weissgerber & Reinhard, 2017; Weltman & Eakin, 2014).

Given the simplicity and ease in which perceptual disfluency can be implemented, it is not surprising researchers began touting the educational implications of such a manipulation. Perceptual disfluency as a possible educational intervention started to garner more support with the publication of Diemand-Yauman et al. (2011). Across two experiments, Diemand-Yauman et al. (2011) showed that placing learning materials in disfluent typefaces (e.g., Comic Sans, Bodoni MT, Haettenschweiler, Monotype Corsiva) improved memory in the lab (when learning about space aliens) and in a high school classroom where students learned about a variety of different content areas (i.e., AP English, Honors English, Honors Physics, Regular Physics, Honors U.S. History, and Honors Chemistry) from PowerPoints with information presented in difficult to read typefaces.

Unfortunately, the evidence for the perceptual disfluency effect is not always as clear as in the example above. A recent case in point is the typeface Sans Forgetica, developed through a collaboration among marketers, psychologists, and graphic designers (see Earp, 2018). Originally launched with the strong claim that typeface enhances memory retention due to the backward-slanting letters and gaps within each letter which forces individuals to 'generate' the missing parts of each word. However, several subsequent studies have failed to replicate these

claims, finding Sans Forgetica to be *just as* forgettable (Cushing & Bodner, 2022; Geller et al., 2020; Huff et al., 2022; Roberts et al., 2023; A. Taylor et al., 2020; Wetzler et al., 2021). Similar negative results have been found for various other perceptual disfluency interventions, such as small font sizes (Rhodes & Castel, 2008), difficult-to-hear stimuli (Rhodes & Castel, 2009), minor blurring (Yue et al., 2013), and alternative typefaces (Rummer et al., 2015).

Due to the mixed findings, researchers began exploring boundary conditions of the disfluency effect. The importance of testing the effects of disfluency in the presence of other variables is key to its usefulness as an educational intervention. Geller et al. (2018), for instance, found level of disfluency (more vs. less disfluent) mattered. Using easy-to-read and hard-to-read handwritten cursive words,Geller et al. (2018) showed there is a *Goldilocks* zone for perceptual disfluency effects. That is, stimuli cannot be too easy to read (i.e., print words) or too hard to read (i.e., hard-to-read cursive). Only when the stimuli were moderately difficult or just right (i.e., easy-to-read cursive), did memory improve. In another paper, Geller and Peterson (2021) demonstrated that memory benefits for disfluent stimuli are more robust when test expectancy is low. That is, a disfluency effect is only seen when participants are not told about an upcoming memory test. The authors reasoned that knowing about a memory test engaged a strategy where all stimuli get processed to a deeper level, regardless of how perceptually disfluent the stimulus is. This countervails any benefit disfluency has on memory. Additionally, a few studies have noted the importance of individual differences. For example, Eskenazi and Nix (2021) showed better spellers remembered more words and meanings than poor spellers when placed in a disfluent font (Sans Forgetica).

Although perceptual disfluency can arise in certain contexts, its usefulness in educational settings—where students are prepared for tests—may be limited. However, Geller and Peterson (2021) argued that perceptual disfluency holds practical value because people often rely on

incidental memory in everyday decision-making. For disfluency to be leveraged effectively,

though, it is essential to predict when and where it will occur.

Theoretical accounts of the disfluency effect

To achieve the aim of utilizing perceptual disflueny in applied settings, a better

understanding of the mechanisms involved in eliciting the disfluency effect. Several theories

have been proposed to explain this phenomenon. Geller et al. (2018) provided a review of two

theories put forth to explain the disfluency effect. The metacognition account of disfluency

(Alter, 2013; Pieger et al., 2016) posits that disfluency acts as a signal to engage in more control/

regulatory processing (Pieger et al., 2016). Within this account, disfluency arises *after* the

stimulus has been identified. As a result, the type of disfluency experienced does not matter, just

that the learner perceives something as disfluent and regulates their behavior properly.

The compensatory processing account (Mulligan, 1996) suggests that the disfluency

effect is a result of increased top-down processing from lexical and semantic levels of

representation. This framework is largely based on the popular interactive activation mode (IA

model) of word recognition (McClelland & Rumelhart, 1981). In the IA model, visual input

activates representations at three levels of representation: the feature, letter, and word levels.

Activation in the IA model is both feed-forward and feed-backward. Thus, when there is

perceptual noise (such as the one induced by a mask), there is increased top-down processing

from higher, lexical or semantic, levels to aid in word identification. It is this increased top-down

processing to lower levels that produces better memory.

More recently, Ptok et al. (2019) put forth a limited capacity and stage-specific model to

explain conflict-encoding effects like the perceptual disfluency effect. Within their model,

memory effects arising from encoding conflict rely on (1) the level of processing tapped by the

task and (2) metacognitive processes that include monitoring and control. Across six

experiments, Ptok et al. (2020) demonstrated better recognition memory for target words when

shown with incongruent versus congruent semantic distractor words (i.e., category labels of size, animacy, or gender; e.g., "Chair - Alive" vs. "Chair - Inanimate"), but no memory benefit for incongruent versus congruent response distractor words (e.g., Lisa -"left"/ Lisa - "right"). While both tasks resulted in conflict evinced by longer reaction times (RTs) to targets preceded by incongruent primes, only when the encoding task focused attention on target meaning (i.e., semantic categorization) rather than response selection did a memory benefit arise. In a follow-up set of experiments, Ptok et al. (2020) replicated this pattern of findings, and in addition, provided physiological evidence from cognitive pupillometry (i.e., the study of eye pupil size and how it relates to cognitive processing), which has been shown to be an indicator of mental effort and congitive control Wel & Steenbergen (2018). In their study they observed larger pupil size for semantic incongruent and response incongruent primes, but only observed a memory benefit for semantic incongruent conditions. Interestingly, they also showed that these memory benefits can be mitigated by manipulating endogenous attention. Ptok et al. (2020) (Experiment 3) were able to eradicate the conflict encoding benefit by having participants sit in a chinrest and focus on the task. This is similar to what has been found in the perceptual disfluency literature. For example, having participants study words in anticipation for a test can eradicate the benefit of perceptual disfluency Geller and Peterson (2021). In addition, requiring participants to make judgments of learning (i.e., a metamemory judgment on a scale of 0-100 indicating how likely it is they will recall the word on a later memory test) after each studied word also eradicates the disfluency effect (Besken & Mulligan, 2013; Rosner et al., 2015). Taken together, this highlights the critical role of both the kind of processing done on the to-be-remembered stimuli, and control processes in eliciting a disfluency effect.

All three of these theories propose potential loci for the perceptual disfluency effect. In the metacognitive account, the disfluency effect occurs at a later post-lexical stage, after word recognition has taken place. The compensatory processing account (Mulligan, 1996) links the

perceptual disfluency effect directly to the word recognition process. That is, disfluent words receive more top-down processing from lexical or semantic levels during encoding. Lastly, the stage-specific model proposed by Ptok et al. (2019) associates perceptual disfluency effects with a specific stage of processing, namely the semantic level, but it also considers general attentional and cognitive control processes that are not solely tied to the word recognition process.

Moving beyond the mean: modeling RT distributions

Ex-Gaussian distribution

To test the different stages or loci involved in the perceptual disfluency effect, it is necessary to use methods that allow for a more fine-grained analysis of processes during encoding. In the perceptual disfluency literature (and learning and memory more broadly), it is common to use measures such as mean accuracy and RTs to assess differences between a disfluent and fluent condition (Geller et al., 2018; Geller & Peterson, 2021; Rosner et al., 2015). While this approach is often deemed as acceptable practice, there has been a call to go beyond traditional RT methods when making inferences (see Balota & Yap, 2011).

There are a several reasons for making the shift from traditional *mean*-RT analyses *(which is what linear models do)* to *approaches* that utilize the whole RT distribution. One reason is traditional approaches fail to capture the nuances inherent in RT distributions. Namely, RT distributions are unimodal and positively skewed. A standard analysis based on means can conceal effects that change only the shape of the tail of the distribution, only the location, or both the location and the shape of the distribution.

Another reason to transition away from traditional analyses is that RTs provide only a coarse measure of processing during encoding. RTs capture the total sum of various task-related factors, ranging from non-decisional components like stimulus encoding and motor responses to decisional components. This amalgamation does not allow one to parse out specific effects that might occur early or later in processing.

Lastly, from a statistical standpoint, RTs present significant challenges. Specifically, they often violate two crucial assumptions: they are not normally distributed and their variance is frequently heterogeneous. Such violations can lead to biased results when making statistical inferences, as pointed out by (Wilcox, 1998).

A perfect example of this comes from the Stroop task (Stroop, 1935) . The classic Stroop finding shows words presented in an incongruent color font (the word "red" printed in "blue" font) increases RTs compared to words in a baseline condition (e.g., XXXXX presented in a color font). The Stroop interference effect is something you can bet your house on. A more inconsistent finding is seeing facilitation (shortened RTs) when the word and color are congruent (i.e., "Green" presented in "Green") compared to a baseline condition.

One alternative for this conundrum is to examine RT distributions using mathematical models that capture the nuances of the RT distribution and consider various statistical properties, such as the location (μ) , spread (σ) , and skewness (τ) of the distribution. One popular choice is the ex-Gaussian distribution (Balota & Yap, 2011; Ratcliff, 1978). As the name suggests, the ex-Gaussian distribution decomposes RTs into three parameters: mu (μ) representing the mean of the Gaussian component, sigma (σ) representing the standard deviation of that Gaussian component, and tau (τ) representing the mean and standard deviation of an exponential component capturing the tail of the distribution. The algebraic mean of ex-Gaussian is a combination of $\mu + \tau$ and the sd is $\sqrt{\sigma^2 + \tau^2}$. Together the three parameters represent different aspects of the distribution's location and shape.

Heathcote et al. (1991) explored the facilitation-inhibition asymmetry in the Stroop task with an ex-Gaussian model and found both facilitation (from congruent trials) and interference (from incongruent trials) on μ. For σ, the analysis showed interference, but no facilitation. For τ, there was interference for both congruent and incongruent conditions. Comparing this to a mean RT analysis, they showed the standard interference Stroop effect, but no facilitation. Given that

the algerbraic mean of the ex-Gaussian is $\mu + \tau$, the failure to observe a facilitation effect in the standard mean analysis likely arose from facilitation on μ and interference on τ canceling each other out. A finding such as this would be impossible looking solely at mean RTs.

Exploring effects from a distributional perspective has provided a richer understanding of how different experimental manipulations affect word recognition. Experimental manipulations can produce several distinct patterns. One pattern involves a shift of the entire RT distribution to the right, without increasing the tail or skew. A pattern such as this would suggest a general effect and would manifest as an effect on μ, but not τ . As an example, semantic priming effects–where responses are faster to targets when preceded by a semantically related prime compared to an unrelated prime–can be nicely explained by a simple shift in the RT distribution (Balota et al., 2008). Alternatively, an experimental manipulation could produce a pattern where the RT distribution is skewed or stretched in the slower condition. This suggests that the manipulation only impacts a subset of trials, and is visible as an increase in τ. An example of an effect that only impacts τ is the transposed letter effect in visual word recognition (Johnson et al., 2012). The transposed letter (TL) effect involves misidentification of orthographically similar stimuli that with transposed internal like, like mistaking "JUGDE" for "JUDGE" (Perea & Lupker, 2003). Finally, you could observe a pattern wherein an experimental manipulation results in both changes in μ and τ, which would shift and stretch the RT distribution. Recognizing low frequency words have been shown to not only shift the RT distribution, but also stretch the RT distributions (Andrews & Heathcote, 2001; Balota & Spieler, 1999; Staub, 2010).

The ex-Gaussian model, while mostly descriptive in nature. has been used as a theoretical tool to map model parameters onto cognitive processes. For example, the μ and σ parameters have been tied to early, non-analytic, processing. In the area of semantic priming, the selective effect on μ has been taken as evidence for an automatic spreading activation process (or head-start), according to which the activation of a node in a semantic network spreads automatically to

interconnected nodes, preactivating a semantically related word (Balota et al., 2008; Wit & Kinoshita, 2015). The exponential component (τ) has been tied to later, more analytic, processing (Balota & Spieler, 1999). Specifically, increases in τ have been attributed to working memory and attentional processes (Fitousi, 2020; Kane & Engle, 2003). For instance, Johnson et al. (2012) tied τ differences for the TL effect to a post-lexical checking mechanism that arose from a failure to identify the stimulus on a select number of trials rather than a broader, lexical, effect occurring on every trial. When taken together, these findings suggest that ex-Gaussian parameters could map to early vs. late stages of cognitive processing. However, such mapping between distributional descriptives and and cognitive processes remains controversial and should be interpreted carefully (Heathcote et al., 1991; Matzke & Wagenmakers, 2009).

Drift-diffusion model (DDM)

Unlike the ex-Gaussian distribution discussed above, which makes little theortical commitments regarding process, the drift diffusion model - DDM (see Ratcliff et al., 2016, for a comprehensive introduction) is a process-model and it's parameters can be linked to latent cognitive constructs (Gomez et al., 2013). The DDM is a popular computational model commonly used in binary speeded decision tasks such as the lexical decision task (LDT). The DDM model assumes a decision is a cumulative process that begins at stimulus onset and ends once a noisy accumulation of evidence has reached a decision threshold.The DDM has led to important insights into cognition in a wide range of choice tasks, including perceptual-, memory-, and value-based decisions (Myers et al., 2022).

In the DDM, RTs are decomposed into several parameters that represent distinct cognitive processes. The most relevant to our purposes here are the drift rate (v) and non-decision time (ndt; T_{er}) parameters. Drift rate (v) represents the rate at which evidence is accumulated towards a decision boundary. In essence, it is a measure of how quickly information is processed to make a decision. A higher (more positive) v indicates a steeper slope, meaning that evidence is

accumulated more quickly, leading to faster decisions. Conversely, a lower v indicates a shallower slope, meaning that evidence is accumulated more slowly. Drift rate is closely linked to the decision-making process itself and serves as an index of global processing demands imposed by factors such as task difficulty, memory load, or other concurrent cognitive demands —particularly when these processes compete for the same cognitive resources (Boag, Strickland, Loft, et al., 2019). Additionally, drift rates have been implicated as a key mechanism of reactive inhibitory control (Braver, 2012), where critical events (e.g., working memory updates or task switches) trigger inhibition of prepotent response drift rates (Boag, Strickland, Loft, et al., 2019; Boag, Strickland, Heathcote, et al., 2019).

The T_{er} parameter represents the time taken for processes other than the decision-making itself. This includes early sensory processing (like visual or auditory processing of the stimulus) and late motor processes (like executing the response).

The DDM has been shown to be a valuable tool for studying the effects of different experimental manipulations on cognitive processes in visual word recognition. For example, Gomez and Perea (2014) demonstrated certain manipulations can deferentially affect specific parameters of the model. For instance manipulating the orientation of words (rotating them by 0, 90, or 180 degrees) affected the T_{er} component, but not v component. In contrast, word frequency (high-frequency words vs. low-frequency words) primarily influenced both the drift rate and non-decision time. These findings highlight the sensitivity of the DDM in identifying and differentiating the impact of various stimulus manipulations on different cognitive processes involved in decision-making.

Goals of the present experiments

In the present experiments, we pursued two aims related to perceptual disfluency. The first aim was to examine the replicability of the perceptual disfluency effect. To optimize our chances for observing this effect, we utilized a disfluency manipulation known in the literature to

enhance memory: blurring (Rosner et al., 2015). We manipulated perceptual disfluency by blurring the words at three levels. Participants were presented with clear words (no blur), low blurred words (5% Gaussian blur) and high blurred words (15% Gaussian blur). High level blurring has been shown to enhance memory (Rosner et al., 2015). As Geller et al. (2018) noted, not all manipulations are created equal. Perceptual manipulations affect processing in different ways. It is important to show just how these manipulations affect different stages of processing and what type of manipulations do and do not produce a disfluency effect. By examining different levels of perceptual disfluency, we provide a more nuanced account of encoding processes and how this affects memory.

The second, more pivotal aim was to enrich the methodological toolkit available to researchers investigating conflict encoding, such as perceptual disfluency. By applying distributional techniques—specifically ex-Gaussian analysis and the diffusion decision model (DDM)—our goal was to demonstrate how these approaches can provide deeper insight into how encoding difficulty influences memory. It is important to note that other distributional techniques for analyzing response times exist (e.g., the linear ballistic accumulator model, gamma distributions, etc.). However, we chose the ex-Gaussian and DDM approaches due to their popularity and extensive use in the word recognition literature, where they have received the most empirical attention (Balota et al., 2008). Ultimately, these efforts aim to clarify the conditions under which perceptual disfluency enhances memory—and when it does not.

To promote transparency and reproducibility, this paper was written in R (R Core Team, 2025) using Quarto (Allaire et al., 2024), an open-source publishing system that allows for dynamic and static documents. This allows figures, tables, and text to be programmatically included directly in the manuscript, ensuring that all results are seamlessly integrated into the document.To increase computational reproducibility we use the rix (Rodrigues & Baumann, 2025) package which harnesses the power of the nix (Dolstra & contributors, 2023) ecosystem to

help with computational reproducibility. Not only does this give us a snapshot of the packages used to create the current manuscript, but it also takes a snapshot of system dependencies used at run-time. This way reproducers can easily re-use the exact same environment by installing the nix package manager and using the included default.nix file to set up the right environment. The README file in the GitHub repository contains detailed information on how to set this up to reproduce the contents of the current manuscript. We have also included a video tutorial. We hope this will make it easier for researchers to apply this code to their own research.

Predictions

Using the ex-Gaussian distribution and the DDM will provide us with a descriptive account of how disfluency manipulations affect encoding. Each theoretical account makes specific predictions about the loci of the perceptual disfluency affect. We can use these predictions to forecast how the components in each mathematical model might be influenced.

If the metacognitive account is correct (Alter, 2013), which assumes a post-lexical locus for the perceptual disfluency effect, we might predict a lengthening of the distribution tail on some of the trials for blurred words. This would manifest itself on the τ component. Specifically, a larger τ parameter estimate for high blurred and low blurred words compared to no blurred words. As it relates to memory performance, there should be no difference between high and low blurred words.

Another scenario entails a general slow down of processing—causing distributional shifting μ, but not skewing τ. According to the the compensatory processing account (Mulligan, 1996), we would expect increased shifting for high blurred words compared to low and no blurred words. This should result in a mnenmonic benefit for high blurred words, but not low blurred words.

Lastly, we might observe not only a shift of the entire distribution (an effect on μ), but also change the shape of the distribution (an effect on τ), indicating a combination of early and

late processes. A similar pattern has been found with hard-to-read handwritten words (e.g., Perea et al., 2016; Vergara-Martínez et al., 2021). This extra post-lexical processing received by high blurred words is assumed to facilitate better recognition memory. This finding would be in line with the stage-specific account (Ptok et al., 2019). The finding of better memory for high blurred words, but not low blurred words would be in line with the stage-specific account of conflict encoding Ptok et al. (2020). Having a better sense of when and where disfluency effects arise is critical in determining its usefulness in the educational milieu.

In terms of the DDM parameters, we predicted high blurred words would affect both v and T_{er} parameters. Specifically, high blurred words would produce lower v and high T_{er} compared to clear and low blurred words. Additionally, we predicted that low blurred words would only affect T_{er}. Table 1 summarizes each account of the perceptual disfluency and the predicted outcomes according to each mathematical model. Importantly, some of the principles espoused by the three accounts are postulated verbally, and they can be realized in models in many different ways. We have made a good-faith effort to translate these verbal descriptions into formal models, but we recognize that reasonable researchers might make different modeling choices—a reality that reflects the unavoidable subjectivity in scientific inference [see McElreath (2020).

In addition to ex-Gaussian and DDM analyses, we will provide a graphical description of changes to the RT distribution using quantile and delta plots (see Angele et al., 2023). The process of visualization through quantile analysis can be broken down into four distinct steps:

1. Sorting and plotting: For correct trials, RTs are arranged in ascending order within each condition. We then plot the average of the specified quantiles (e.g., .1, .2, .3, .4, .5, .9).

2. Quantile averaging across participants: The individual quantiles for each participant are averaged, a concept reminiscent of Vincentiles.

3. Between-condition quantile averaging: The average for each quantile is computed between the conditions.

4. Difference calculation: We determine the difference between the conditions, ensuring the sign of the difference remains unchanged.

Typically, there are four observable patterns in the graphical depiction. No observable difference occurs when the conditions do not show any noticeable distinction. Late differences emerge when increasing differences appear later in the sequence, suggesting that the conditions diverge over time. A complete shift indicates a consistent difference across all quantiles, signaling an overall shift in the distribution. Finally, early differences reveal distinctions early in the reaction time distribution, suggesting an initial divergence between conditions. maps these patterns onto existing theoretical models of disfluency.

```
----- end of section:
```

Table 1

Mapping model predictions to theoretical constructs

Account	Description	Loci	Contrast	Ex-Gaussian Predictions	Drift Diffusion Predictions	Quantile Plots	Recognition Memory Predictions
Meta-cognitive	Perceptual disfluency affects meta-cognitive processes via increased system 2 processing	Post-lexical	High blur vs. Low blur/Clear	μ: × τ: ↑	T_{er}: × v: ↑	Late Difference	High > Low/Clear
			Low blur vs. Clear	μ: × τ: ↑	T_{er}: × v: ↑	Late Difference	Low > Clear
Compensatory-processing	Perceptual disfluency affects the word recognition process	Lexical/semantic	High blur vs. Low blur/Clear	μ: ↑ τ: ×	T_{er}: ↑ v: ×	Complete Shift	High > Low/Clear
			Low blur vs. Clear	μ: × τ: ×	T_{er}: × v: ×	No Difference	×
Stage-specific	Disfluency effects rely on (1) the stage or level of processing tapped by the task and (2) monitoring and control processes	Lexical/semantic and Post-lexical	High blur vs. Low blur/Clear	μ: ↑ τ: ↑	T_{er}: ↑ v: ↑	Complete Shift + Late Differences	High > Low/Clear
			Low blur vs. Clear	μ: × τ: ×	T_{er}: × v: ×	No Difference	×

```
----- end of section:
```

Experiment 1A: Context Reinstatement

In Experiment 1A, we collected RTs from a lexical decision task (LDT) during encoding followed by a surprise recognition memory test. Using a two-choice task like the LDT allowed

us to examine how perceptual disfluency affects encoding processes using mathematical models. Based on previous research (Geller & Peterson, 2021), there was no mention of the recognition test when participants signed up for the study to give us the best chance of observing a disfluency effect.

Method

Transparency and Openness

This study complies with transparency and openness guidelines. The preregistered analysis plan for this experiment can be found here: https://osf.io/q3fjn. All raw and summary data, materials, and R scripts for pre-processing, analysis, and plotting for Experiments 1 can be found at https://osf.io/6sy7k/. All deviations and changes from the preregistration are noted herein.

Participants

We preregistered a sample size of 216. All participants were recruited through the university subject pool (SONA). A design with a sample size of 216 can detect effect sizes of $\delta \geq 0.2$ with a probability of at least 0.90, assuming a one-sided criterion for detection that allows for a maximum Type I error rate of $\alpha = 0.05$. Per our exclusion criteria, we retained participants that were native English speakers, were over the age of 17, had overall accuracy on the LDT greater than 80%, and did not complete the experiment more than once. Due to our exclusion criteria, we oversampled participants. To ensure equal number of participants in each list we randomly chose 36 participants from each list to reach our target sample size.

Apparatus and stimuli

The experiment was run using PsychoPy software and hosted on Pavlovia (www.pavlovia.org). You can see an example of the experiment by navigating to this website: https://run.pavlovia.org/Jgeller112/ldt_dd_l1_jol_context.

We used 84 words and 84 nonwords for the LDT. Words were obtained from the LexOPS package in R (J. E. Taylor et al., 2020). All of our words were matched on a number of different lexical dimensions. All words were nouns, 4-6 letters in length, had a known proportion of between 90%-100%, had a low neighborhood density (OLD20 score between 1-2), high concreteness, imageability, and word frequency. Our nonwords were created using the English Lexicon Project (Balota et al., 2007). Stimuli can be found at our OSF project page cited above.

Blurring. Blurred stimuli were processed through the imager package in R (Barthelme, 2023) and a personal script (https://osf.io/gr5qv). Each image was processed through a high blur filter (Gaussian blur of 15) and low blur filter (Gaussian blur of 10). These pictures were then imported into PsychoPy as picture files. See Figure 1 for examples of what a clear, low blurred, and high blurred word would look like to the participant.

Figure 1

Clear (left), low blur (10% blur) (right), and high blur (15% blur) (center) examples

Design

       We created two lists: One list (84 words; 28 clear, 28 low blur, and 28 high blur) served as a study (old) list for the LDT task while the other list served as a test (new) list (84 words; 28 clear, 28 low blur, and 28 high blur) for our recognition memory test that occurred after the LDT. We counterbalanced each list so each word served as an old word and a new world and were presented in clear, low blurred, and high blurred across participants. This resulted in six counterbalanced lists. Lists were assigned to participants so that across participants each word occurs equally often in the six possible conditions: clear old, low blur old, high blur old, clear new, low blur new, and high blur new. For the LDT task, we generated a set of 84 legal nonwords that we obtained from the English Lexicon Project. These 84 nonwords were used across all 6 lists.

Procedure

       The experiment consisted of two phases: an encoding phase (LDT) and a test phase. During the encoding phase, a fixation cross appeared at the center of the screen for 500 ms. The fixation cross was immediately replaced by a letter string in the same location. To continue to the next trial, participants had to decide if the letter string presented on screen was a word or not by either pressing designated keys on the keyboard ("m" or "z") or by tapping on designated areas on the screen (word vs. nonword) if they were using a cell phone/tablet. After the encoding phase, participants were given a surprise old/new recognition memory test. During the test phase, a word appeared in the center of the screen that either had been presented during study ("old") or had not been presented during study ("new"). Old words occurred in their original typeface, and following the counterbalancing procedure, each of the new words was presented as clear, low blurred, or high blurred. All words were individually randomized for each participant during both the study and test phases and progress was self-paced. After the experiment, participants were debriefed. The entire experiment lasted approximately 15 minutes.

Data Analysis Plan

All models were fit in R [v.4.4.3; R Core Team (2025)] using the Stan modeling language (Grant et al., 2017) via the `brms` package (Bürkner, 2017).[1] We used maximal random-effects structures justified by the design (Barr et al., 2013).

We ran four chains of 5,000 MCMC iterations (1,000 warm-up), totaling 16,000 post-warm-up samples (except for the diffusion model, which used 2,000 iterations to reduce computation time). Model quality was checked via prior/posterior predictive checks, $\hat{R}$, and ESS (Vehtari et al., 2021). Convergence was assessed using $\hat{R}$ (target $\leq$ 1.01) and effective sample size (ESS $\geq$ 1000) (Bürkner, 2017). Default (non-informative) priors were used for most parameters. Weakly informative priors were used for population-level parameters to enable Bayes factor (Evidence Ratio(ER)) calculations for two sided-hypotheses against a point null. Full prior specifications are available in the Quarto source file on OSF:

We report posterior means and 90% credible intervals (CrIs) for one-sided hypotheses (preregistered differences), and 95% CrIs for two-sided hypotheses (against zero). Estimated marginal means were extracted using a combination of `emmeans` (Lenth, 2023), `bayestestR` (Makowski et al., 2019), and `brms` (Bürkner, 2017).

Additionally, we report the posterior probability (`Post. Probability`) that an effect lies in a particular direction and ER, which is a generalization of the Bayes factor for directional hypotheses. An ER > 3 indicates moderate to strong evidence for the hypothesis; ER < 0.3

---

[1] To promote transparency and reproducibility, this paper was written using Quarto (Allaire et al., 2024). To increase computational reproducibility we use the rix (Rodrigues & Baumann, 2025) package which harnesses the power of the nix (Dolstra & contributors, 2023) ecosystem to help with computational reproducibility. We used the following R packages: brms v. 2.21.0 (Bürkner 2017, 2018, 2021), cmdstanr v. 0.8.1 (Gabry et al. 2024), colorspace v. 2.1.1 (Zeileis, Hornik, and Murrell 2009; Stauffer et al. 2009; Zeileis et al. 2020), cowplot v. 1.1.3 (Wilke 2024), data.table v. 1.17.0 (T. Barrett et al. 2025), easystats v. 0.7.4 (Lüdecke et al. 2022), emmeans v. 1.10.4 (Lenth 2024), flextable v. 0.9.6 (Gohel and Skintzos 2024), ggdist v. 3.3.2 (Kay 2024b, 2024a), ggeffects v. 1.7.0 (Lüdecke 2018), ggokabeito v. 0.1.0 (M. Barrett 2021), ggrepel v. 0.9.6 (Slowikowski 2024), ggtext v. 0.1.2 (Wilke and Wiernik 2022), here v. 1.0.1 (Müller 2020), hypr v. 0.2.8 (Schad et al. 2019; Rabe et al. 2020), knitr v. 1.50 (Xie 2014, 2015, 2025), modelbased v. 0.10.0 (Makowski et al. 2020), parameters v. 0.24.2 (Lüdecke et al. 2020), rmarkdown v. 2.29 (Xie, Allaire, and Grolemund 2018; Xie, Dervieux, and Riederer 2020; Allaire et al. 2024), tidybayes v. 3.0.7 (Kay 2024c), tidylog v. 1.1.0 (Elbers 2024), tidyverse v. 2.0.0 (Wickham et al. 2019).

indicates support for the alternative; and ER values between 0.3 and 3 are considered

inconclusive. ERs were also used to assess point-null hypotheses ($\delta = 0$). Hypotheses were

considered supported if zero was excluded from the CrI, The posterior probability approached 1,

and ER was > 3.

For all models (excluding the diffusion decision models), we applied ANOVA-style

(effects) coding using contrast variables. For the Blur factor in Experiments 1A, 1B, and 2, we

defined two orthogonal contrasts to capture the primary comparisons of interest. Contrast 1

compared High Blur against the average of Clear and Low Blur, coding High Blur as 0.5 and

both Clear and Low Blur as –0.5. Contrast 2 isolated the difference between Low Blur and Clear,

with Low Blur coded as 0.5, Clear as –0.5, and High Blur as 0. In Experiment 2, we also

included a Frequency factor, with High Frequency coded as 0.5 and Low Frequency as –0.5.

Although these contrasts deviate from our preregistered comparisons, we believe they

offer a more targeted test of our hypotheses. For transparency, we provide all pairwise

comparisons in the accompanying visualizations.

Accuracy. Accuracy (coded as correct (1) vs. incorrect (0)) was modeled using a

Bayesian logistic regression with a Bernoulli distribution.

Ex-Gaussian model. We modeled response times with an ex-Gaussian distribution[2],

allowing the Gaussian mean/location (μ), the Guassian standard deviation (σ) and the

exponential scale ($\beta = 1/\lambda$) to vary by condition. Please note that when we refer to β we are

referring to τ. To visualize distributional shifts, we generated quantile and delta plots (Balota et

al., 2008; De Jong et al., 1994).

Diffusion model. We fit a hierarchical Bayesian Wiener diffusion model

(Vandekerckhove et al., 2011) with accuracy coding, estimating drift rate (v), boundary

separation (response caution; fixed at 0.5), and non-decision time (T_{er}).

---

[2]The parameterization used by brms does not match the standard formulation of the ex-Gaussian distribution. A custom script was implemented to recover the traditional parameters.

Recognition memory. Following recent trends (Zloteanu & Vuorre, 2024) recognition memory data were analyzed using a Bayesian generalized linear multilevel model (Bernoulli distribution with a probit link). Here the response of the participant ("say old" vs. "say new") are modeled as function of item status (" is old" vs. " is new") and condition.

Bayesian GLMMs provide a more precise and flexible approach than traditional SDT analyses. Following Signal Detection Theory [SDT; Green and Swets (1966)], participant responses can be classified as hits, correct rejections, misses, or false alarms, depending on the item status ("old" vs. "new"). In the probit regression framework, the interaction between item status and a predictor of interest corresponds directly tp d', while the main effects reflect response criterion (DeCarlo, 1998; for a detailed discussion of Bayesian SDT modeling see Zloteanu & Vuorre, 2024). Note that the model parameterization reflects $-c$ (i.e., reversed sign) and this is what is reported in the paper. For visualization purposes we use the conventional parameterization, where positive values indicate more conservative responding and negative values indicate a more liberal bias.

Results

All models presented no divergences, and all chains mixed well and produced comparable estimates ($\hat{R} < 1.01$ and ESS > 1000).

Accuracy

The analysis of accuracy is is based on 18144 data points. After removing fast and slow RTs we were left with 17873 data point points (0.015).

Table 2

Summary of posterior distributions for fixed effects in the Bayesian accuracy model (Experiment 1A).

Hypothesis	Mean	SE	90% CrI	ER	Posterior Prob
High Blur < (Low Blur + Clear)	-1.03	0.16	[-1.293, -0.77]	Inf	1.00
Low Blur < Clear	0.04	0.13	[-0.216, 0.297]	1.26	0.56

Note. 95% CrI for equivalency tests. Posterior probability indicates the proportion of the posterior distribution that falls on one side of zero (either positive or negative), representing the probability that the effect is greater than or less than zero.

Model estimates can be found in Table 2. High blur words had lower accuracy compared to clear and low blurred words, , $b = -1.031$, 90% CrI$[-1.293, -0.77]$, ER = Inf. However, the evidence was weak for there being no significant difference in the identification accuracy between clear and low blurred words, $b = 0.041$, 90% CrI$[-0.216, 0.297]$, ER = 1.257.

RTs: Ex-Gaussian

The analysis of RTs (correct trials and words) is based on 16980 data points, after removing fast and slow RTs (0.013 .

Table 3

Summary of posterior distributions for fixed effects in the ex-Gaussian model (Experiment 1A).

Hypothesis	Parameter	Mean	SE	90% CrI	ER	Posterior Prob
High Blur > (Low Blur + Clear)	Mu	0.11	0.00	[0.1, 0.114]	Inf	1.00
Low Blur > Clear	Mu	0.02	0.00	[0.012, 0.02]	Inf	1.00
High Blur > (Low Blur + Clear)	Sigma	0.16	0.06	[0.057, 0.253]	163.95	0.99
Low Blur = Clear	Sigma	0.03	0.05	[-0.066, 0.136]	16.22	0.94
High Blur > (Low Blur + Clear)	Beta	0.43	0.04	[0.367, 0.487]	Inf	1.00
Low Blur = Clear	Beta	-0.00	0.03	[-0.062, 0.061]	7.77	0.89

Note. 95% CrI for equivalency tests. Posterior probability indicates the proportion of the posterior distribution that falls on one side of zero (either positive or negative), representing the probability that the effect is greater than or less than zero. Sigma and Beta parameters are on the log (logit) scale.

A visualization of how blurring affected processing during word recognition can be seen in the quantile and delta plots in Figure 2. Beginning with the μ parameter, there was greater shifting for high blurred words compared to clear and low blurred words, $b = 0.107$, 90% CrI

[0.1, 0.114], ER = Inf. Low blurred compared to clear words showed greater shifting, $b = 0.016$,

90% CrI [0.012, 0.02], ER = Inf.

Analyses of the σ and τ parameters yielded a similar pattern. Variance was higher for

high blurred words compared to clear and low blurred words, $b = 0.157$, 90% CrI [0.057, 0.253],

ER = 163.948. There was strong evidence for no difference between low blurred and clear

words, $b = 0.034$, 90% CrI [−0.066, 0.136], ER = 16.22.

Finally, there was greater skewing for high blurred words compared to clear and low

blurred words, $b = 0.427$, 90% CrI [0.367, 0.487], ER = . There was strong evidence for no

difference between low blurred and clear words, $b = 0$, 90% CrI [−0.062, 0.061], ER = 7.769.

Figure 2

Quantile plots for each condition (A) and Delta plots (B) depicting the magnitude of the effect for

hypotheses of interest over time in Experiment 1A. Each dot represents the mean RT at

the .1, .3, .5, .7 and .9 quantiles.
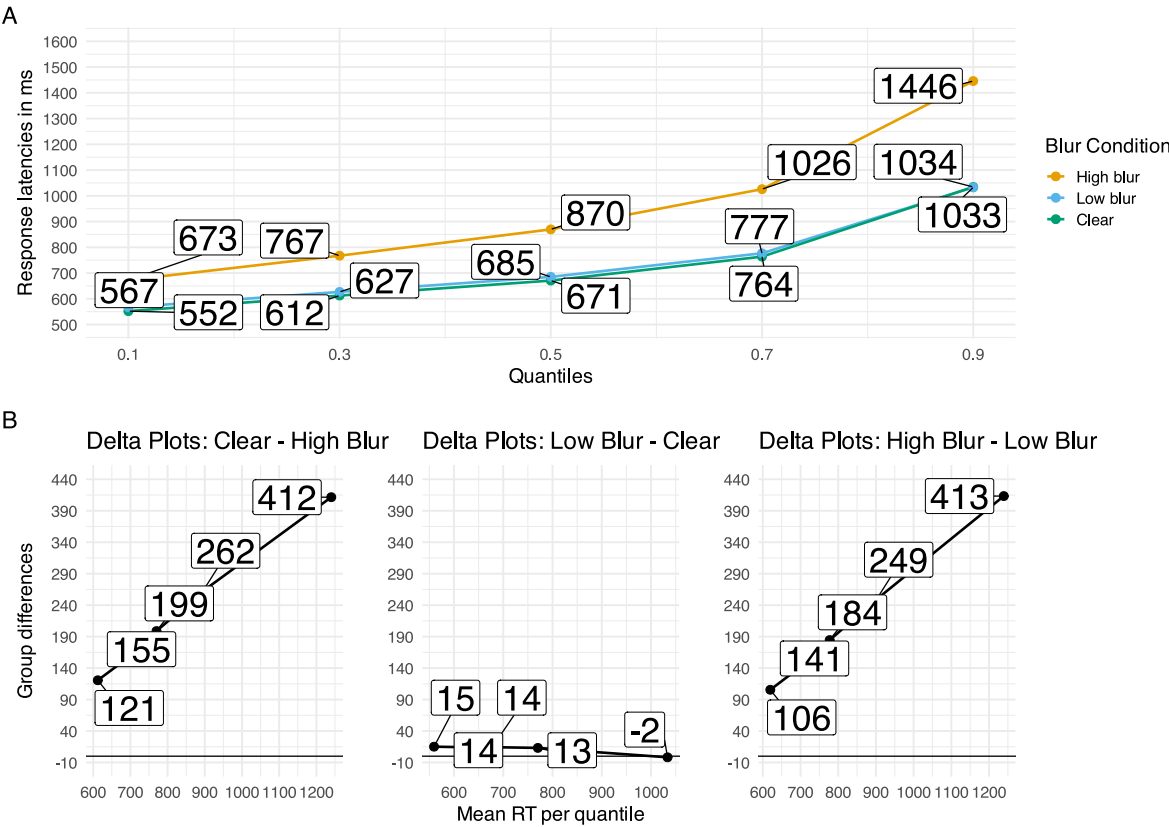
RTs: DDM

We suppressed intercepts when modeling DDM parameters [see (*Diffusion/Wiener Model Analysis with Brms Part i*, 2017) and calculated all comparisons manually.

Table 4

Summary of posterior distributions for fixed effects in the DDM (Experiment 1A)

Hypothesis	Parameter	Mean	SE	90% CrI	ER	Posterior Prob
High Blur > (Low Blur + Clear)	v	-2.22	0.09	[-2.369, -2.073]	Inf	1.00
Low Blur = Clear	v	0.00	0.05	[-0.091, 0.096]	28.16	0.97
High Blur > (Low Blur + Clear)	T_er	0.10	0.00	[0.092, 0.106]	Inf	1.00
Low Blur = Clear	T_er	0.01	0.00	[0.008, 0.018]	Inf	1.00

Note. 95% CrI for equivalency tests

A summary of the diffusion model results can be found in Table 4. There was strong evidence that high-blurred words were associated with a lower drift rate than both clear and low-blurred words, $b = -2.219$, 90% CrI $[-2.369, -2.073]$, ER = Inf. In contrast, the evidence supported the absence of a drift rate difference between low-blurred and clear words, $b = 0.003$, 90% CrI $[-0.091, 0.096]$, ER = 28.157.

There was also substantial evidence that non-decision time was greater for high-blurred words compared to both clear and low-blurred words, $b = 0.099$, 90% CrI $[0.092, 0.106]$, ER = Inf. Additionally, the posterior distribution indicated that low-blurred words had a longer non-decision time than clear words, $b = 0.013$, 90% CrI $[0.008, 0.018]$, ER = Inf.

Recognition Memory

Sensitivity. Figure 3 highlights D' and c means and comparisons across all groups. High-blurred words were better remembered than clear and low blurred words words, $\beta = 0.097$, 90% CrI

[0.019, 0.175], ER = 47.632. The evidence for no difference in sensitivity between clear words and low-blurred words was strong, $\beta = 0.089$, 90% CrI [0.011, 0.168], ER = 32.473.

Exploratory. Bias.

Low blurred words had a bias towards more "old" responses compared to clear words, $\beta = -0.102$, 90% CrI[−0.151, −0.053], ER = 3199, and high blurred words, $\beta = 0.131$, 90% CrI[0.063, 0.199], ER = 0.001. High blurred words showed a more liberal bias compared to clear and low blurred words, $\beta = -0.102$, 90% CrI[−0.151, −0.053], ER = 3199.

Figure 3

Estimated posterior distributions for d-prime and criterion, and differences, with 95% CrIs (thin lines)
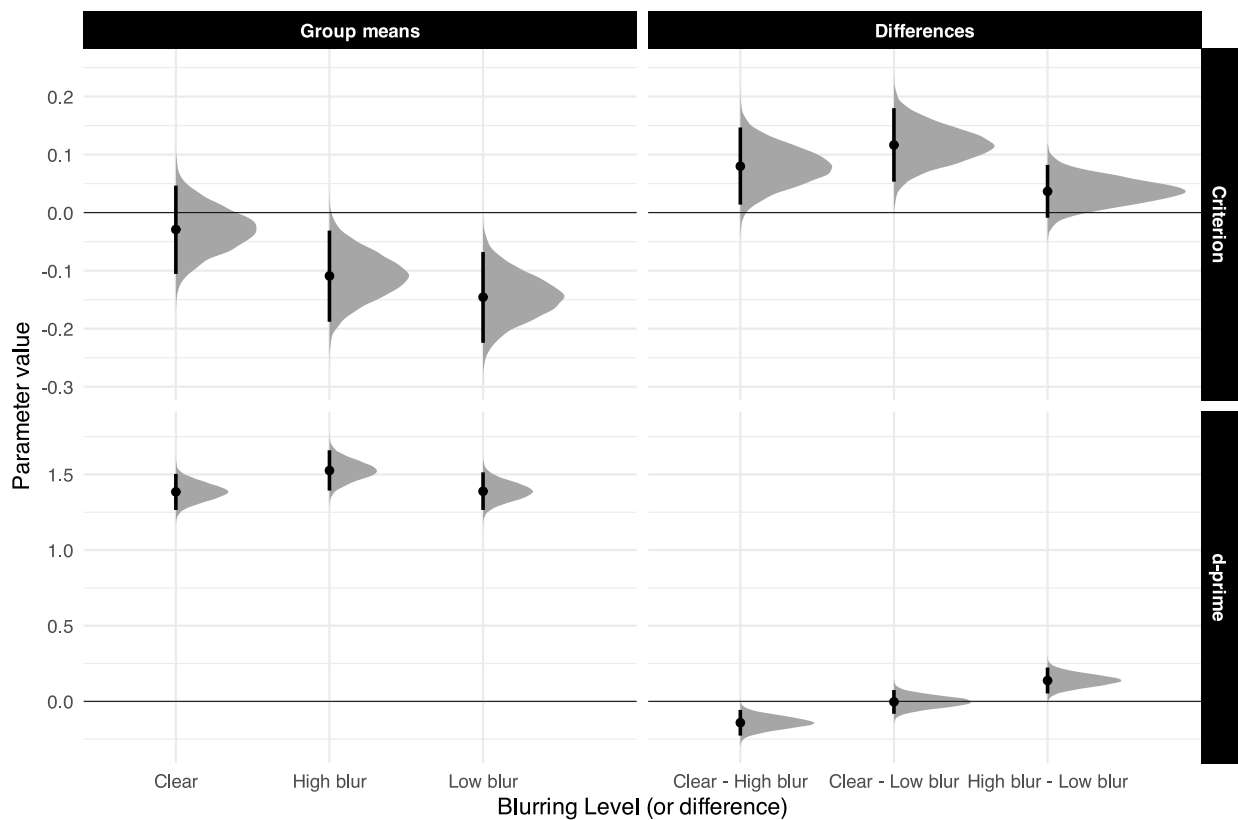
Discussion

Experiment 1A successfully replicated the pattern of results found in Rosner et al. (2015). Specifically, we found high blurred words had lower accuracy than clear and low blurred words, but had better memory.

Adding to this, we utilized cognitive and mathematical modeling to gain further insights into the mechanisms underlying the perceptual disfluency effect. Descriptively, high blurred words induced a more pronounced shift in the RT distribution (μ) and exhibited a higher degree of skew (β) compared to clear and low blurred words. However, low blurred words did not differ compared to clear words on μ or β. These patterns can be clearly seen in our quantile delta plots in Fig. 3.

We also fit the RTs and accuracy data to a diffusion model, which allowed us to make stronger inferences as it relates to stages of processing. High blurred words impacted both an early, non-decision, component evinced by higher T_{er} and a later more analytic, component evinced by a lower v than clear or low blurred words. On the other hand, low blurred words only affected T_{er}.

We present evidence that different levels of disfluency can influence distinct stages of encoding, potentially contributing to the presence or absence of a mnemonic effect for perceptually blurred stimuli. Unlike most studies that commonly employ a single level of disfluency, our study incorporated two levels of disfluency. The results indicate that a subtle manipulation such as low blur primarily affects early processing stages, whereas a more pronounced perceptual manipulation (i.e., high blur) impacts both early and late processing stages. Regarding recognition memory, high blurred stimuli were better recognized compared to low blurred and clear words. This suggests that in order to observe a perceptual disfluency effect, the perceptual manipulation must be sufficiently disfluent to do so and tap later stages of encoding.

Given the important theoretical implications of these findings, Experiment 1B served as a conceptual replication. Due to the bias observed in the recognition memory test (i.e., low blurred words were responded to more liberally), we will not present old and new items as blurred, instead all of the words will be presented in a clear, different, font at test.

Experiment 1B

Method

Transparency and Openness

This study was not pregregistered. All raw and summary data, materials, and R scripts for pre-processing, analysis, and plotting for Experiment 1B can be found at https://osf.io/6sy7k/.

Participants

We used the same sample size procedure as in Experiment 1A ($N = 216$). All participants were recruited through the Rutgers University subject pool (SONA system). The exclusion criteria were the same as in Experiment 1A.

Apparatus, Stimuli, Design, Procedure, and Analysis

Similar to Experiment 1A, the experiment was run using PsychoPy (Peirce et al., 2019) and hosted on Pavlovia (www.pavlovia.org). You can see an example of the experiment by navigating to this website: https://run.pavlovia.org/Jgeller112/ldt_dd_l1_jol.

We used the same stimuli from Experiment 1A. The main difference between Experiment 1A and 1B was all items were presented in a clear, Arial font. To make it more similar to Experiment 1A each set of words presented as clear, low blur, and high blur at study were yoked to a set of new words that were counterbalanced across lists. Therefore, instead of there being one false alarm rate there were 3, one for each blurring level. This ensured each word was compared to studied clear, studied high blurred, and studied low blurred words.

The same model specifications and analyses used in Experiment 1A were used in Experiment 1B.

Results

Accuracy

The analysis of accuracy is is based on 18144 data points. After removing fast and slow RTs we were left with 17809 data point (0.018)

Table 5

Summary of posterior distributions for fixed effects in the Bayesian accuracy model (Experiment 1B)

Hypothesis	Mean	SE	90% CrI	ER	Posterior Prob
High Blur < (Low Blur + Clear)	-1.10	0.17	[-1.376, -0.829]	Inf	1.00
Low Blur = Clear	0.03	0.15	[-0.278, 0.322]	0.90	0.47

Note. 95% CrI for equivalency tests. Posterior probability indicates the proportion of the posterior distribution that falls on one side of zero (either positive or negative), representing the probability that the effect is greater than or less than zero.

A summary of posterior estimates are located in Table 5. High blur words had lower accuracy compared to clear and low blurred words, $b = -1.101$, 90% CrI$[-1.376, -0.829]$, ER $=$ Inf. However, the evidence was weak for there being no significant difference in the identification accuracy between clear and low blurred words, $b = 0.028$, 90% CrI$[-0.278, 0.322]$, ER $= 0.9$.

RTs: Ex-Gaussian

The analysis of RTs (correct trials and word stimuli) is based on 16939 data points, after removing fast and slow RTs (0.016)

Table 6

Summary of posterior distributions for fixed effects in the ex-Gaussian model (Experiment 1B).

Hypothesis	Parameter	Mean	SE	90% CrI	ER	Posterior Prob
High Blur > (Low Blur + Clear)	Mu	0.12	0.01	[0.11, 0.127]	Inf	1.00
Low Blur > Clear	Mu	0.01	0.00	[0.006, 0.015]	Inf	1.00
High Blur > (Low Blur + Clear)	Sigma	0.32	0.07	[0.214, 0.43]	Inf	1.00
Low Blur = Clear	Sigma	-0.09	0.06	[-0.212, 0.035]	5.92	0.85
High Blur > (Low Blur + Clear)	Beta	0.38	0.03	[0.318, 0.43]	Inf	1.00
Low Blur = Clear	Beta	0.03	0.03	[-0.026, 0.084]	5.05	0.83

Note. 95% CrI for equivalency tests. Posterior probability indicates the proportion of the posterior distribution that falls on one side of zero (either positive or negative), representing the probability that the effect is greater than or less than zero. Sigma and Beta parameters are on the log (logit) scale.

A visualization of how blurring affected processing during word recognition can be seen in the quantile and delta plots in Figure 4. Beginning with the μ parameter, there was greater shifting for high blurred words compared to clear and low blurred words, $b = 0.118$, 90% CrI $[0.11, 0.127]$, ER = Inf. Low blurred words had greater shifting compared to clear words, $b = 0.01$, 90% CrI $[0.006, 0.015]$, ER = Inf.

Analyses of the σ parameter yielded a similar pattern. Variance was higher for high blurred words compared to clear and low blurred words, $b = 0.323$, 90% CrI $[0.214, 0.43]$, ER = Inf. There was no evidence of a difference in variance between low blurred and clear words, $b = -0.087$, 90% CrI $[-0.212, 0.035]$, ER = 5.918

Finally, there was greater skewing for high blurred words compared to clear and low blurred words, $b = 0.375$, 90% CrI $[0.318, 0.43]$, ER = . There was no evidence of a difference in variance between low blurred and clear words, $b = 0.029$, 90% CrI $[-0.026, 0.084]$, ER = 5.052.

RTs: DDM

Table 7

Summary of posterior distributions for fixed effects in the DDM (Experiment 1B)

| Hypothesis	Parameter	Mean	SE	90% CrI	ER	Posterior Prob
High Blur < (Low Blur + Clear)	v	-0.90	0.06	[-1.001, -0.794]	Inf	1.00
Low Blur = Clear	v	-0.02	0.06	[-0.13, 0.086]	23.45	0.96
High Blur > (Low Blur + Clear)	T_er	0.10	0.01	[0.093, 0.108]	Inf	1.00
Low Blur = Clear	T_er	0.01	0.00	[0.009, 0.02]	Inf	1.00

Note. 95% CrI for equivalency tests. Posterior probability indicates the proportion of the posterior distribution that falls on one side of zero (either positive or negative), representing the probability that the effect is greater than or less than zero.

A summary of the diffusion model results is presented in Table 7. There was strong evidence that high-blurred words were associated with a lower drift rate than both clear and low-blurred words, $b = -0.896$, 90% CrI $[-1.001, -0.794]$, ER = Inf. In contrast, the evidence supported the absence of a drift rate difference between low-blurred and clear words, $b = -0.023$, 90% CrI $[-0.13, 0.086]$, ER = 23.451.

There was also substantial evidence that non-decision time was greater for high-blurred words compared to both clear and low-blurred words, $b = 0.1$, 90% CrI $[0.093, 0.108]$, ER = Inf. Additionally, the posterior distribution indicated that low-blurred words had a longer non-decision time than clear words, $b = 0.014$, 90% CrI $[0.009, 0.02]$, ER = Inf.

Figure 4

Quantile plots for each condition (A) and Delta plots (B) depicting the magnitude of the effect for hypotheses of interest over time in Experiment 1B. Each dot represents the mean RT at the .1, .3, .5, .7 and .9 quantiles.
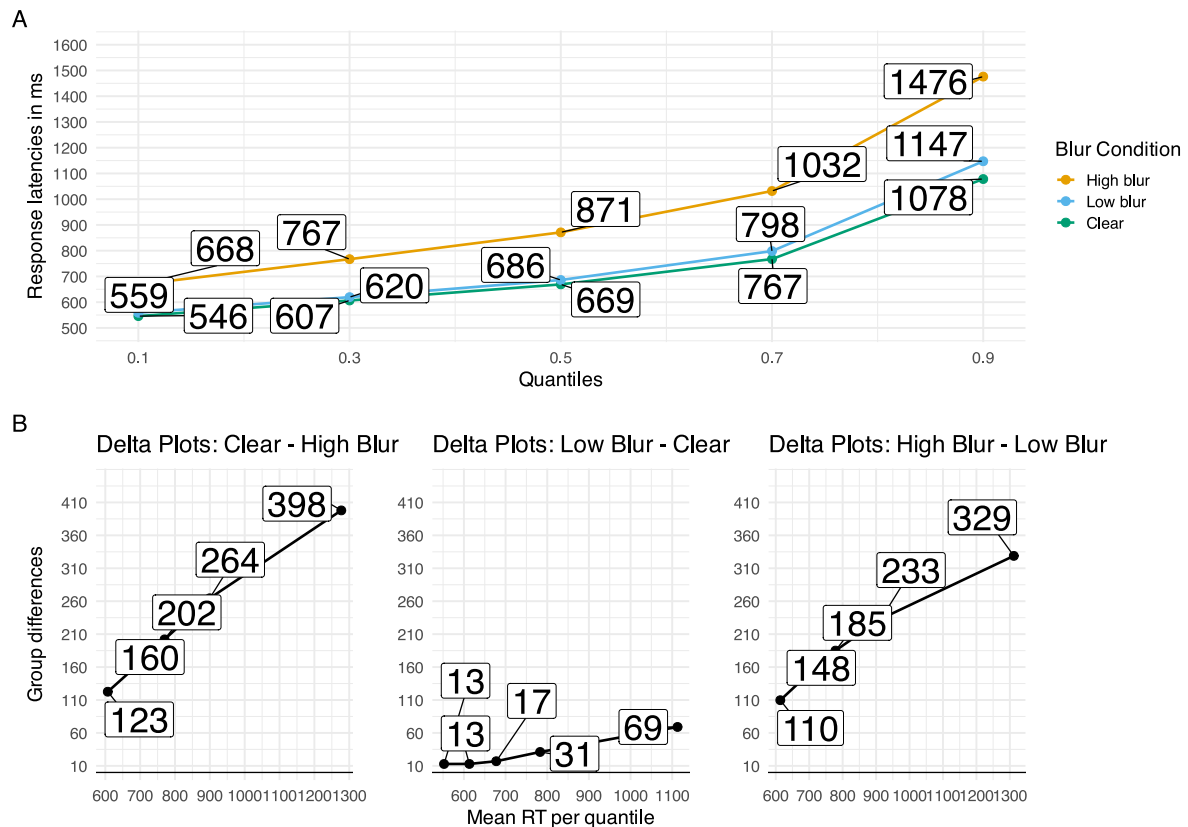

Recognition Memory

Sensitivity. Figure 5 highlights D' and c means and comparisons across all groups. High-blurred words were better remembered than clear and low blurred words words, $\beta = 0.1$, 90% CrI [0.04, 0.161], ER = 319. The evidence for no difference in sensitivity between clear words and low-blurred words was strong, $\beta = -0.04$, 90% CrI [–0.115, 0.036], ER = 1.455.

Figure 5

Estimated posterior distributions (mean) for d-prime and criterion, and differences, with 95% CrIs (thin lines)
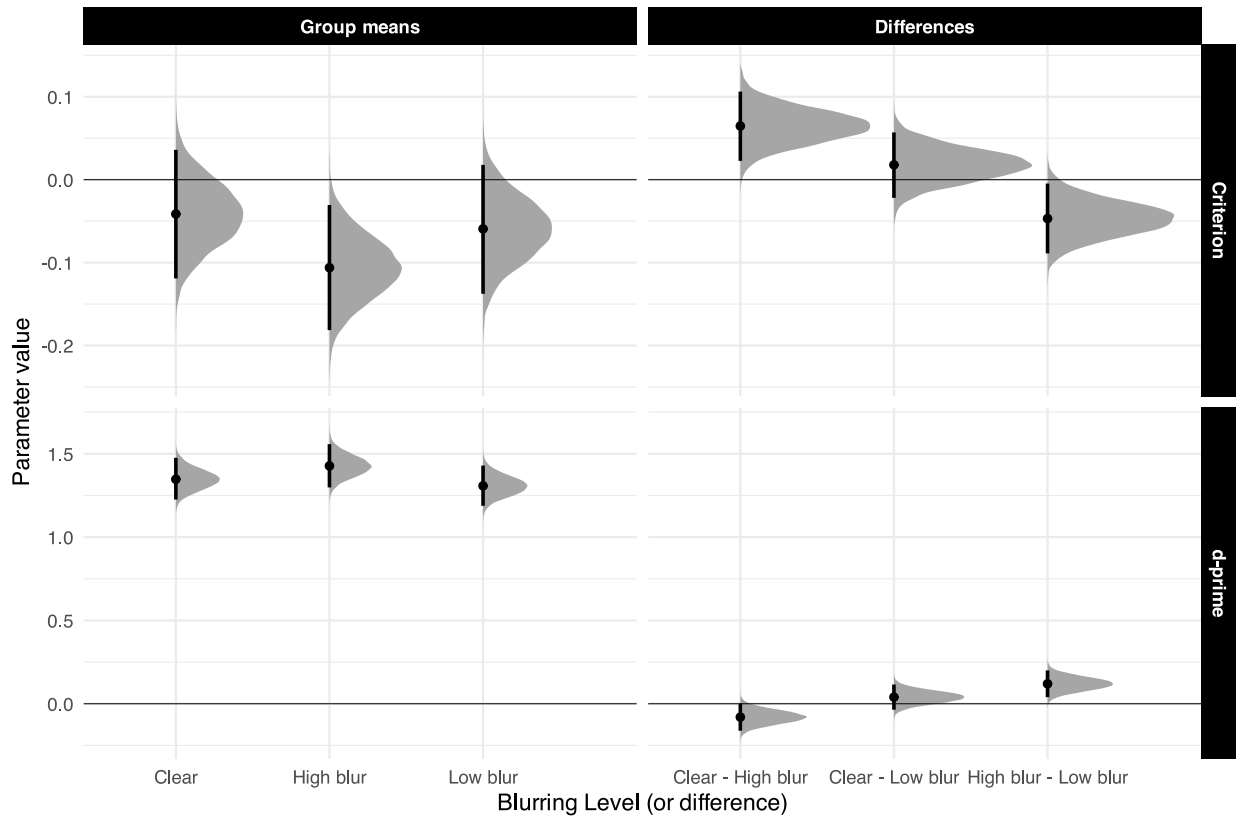
Discussion

We replicated all the findings of Experiment 1A, with no context reinstatement.

Experiment 2

In Experiments 1A and 1B, we applied mathematical modeling and distributional analyses to investigate how visual blurring influences encoding and recognition memory. High-blurred words affected both early and late processing stages, as evidenced by reduced drift rates (v), elevated non-decision times (T_{er}), and changes to the shape of the response time distribution (i.e., both shifts and increased skew). In contrast, low-blurred words—when compared to clear words—appeared to influence only early-stage processing, primarily through increased T_{er} and distributional shifts, without affecting the drift rate.

Recognition memory also reflected this dissociation: sensitivity was greater for high-blurred words than for either clear or low-blurred words. Taken together, these findings support

the stage-specific account, wherein the disfluency effect comprises an early, automatic or non-analytic process, and a later, analytic component. The precise locus of the late-stage benefit remains unclear, however. To explore whether this effect arises from increased attentional engagement or from control processes that operate in parallel with word recognition mechanisms, we further tested the conditions under which this memory enhancement emerges.

One way to more directly test the stage-specific account is to identify a higher-level linguistic variable known to influence word recognition and assess its interaction with blurring. Word frequency is an ideal candidate. Numerous models of word recognition propose that the speed and ease of lexical access varies systematically with word frequency (Coltheart et al., 2001; McClelland & Rumelhart, 1981). Analyses of response time distributions have shown that word frequency affects both early and late stages of processing, with low-frequency words producing larger shifts and greater skew in the RT distribution compared to high-frequency words (Andrews & Heathcote, 2001; Balota & Spieler, 1999; Plourde & Besner, 1997; Staub, 2010; Yap & Balota, 2007; but see Gomez & Perea, 2014 for a DDM perspective).

In recognition memory, low-frequency words are generally remembered better than high-frequency words (Glanzer & Adams, 1985). This advantage has been attributed to the increased cognitive effort or attentional resources required to encode low-frequency items (Diana & Reder, 2006), a view referred to as the elevated attention hypothesis (Malmberg & Nelson, 2003; but see Pazzaglia et al., 2014 for alternative perspectives).

Critically, in tasks such as semantic categorization and pronunciation, word frequency has been shown to interact with stimulus degradation, yielding over-additive effects (Yap & Balota, 2007). According to the logic of additive factors (Sternberg, 1969), such interactions suggest that the manipulated variables impact a shared processing stage. This interaction likely arises because perceptual disfluency disrupts early visual processing and lexical identification, thereby amplifying the frequency effect. Indeed, prior work has documented magnified word

frequency effects under perceptual disfluency, including with handwritten cursive text (Barnhart & Goldinger, 2010; Perea et al., 2016) and rotated letterforms (Fernández-López et al., 2022).

In Experiment 2, we manipulated word frequency (high vs. low) and visual blur (clear, low blur, high blur) within a semantic categorization task, followed by a surprise recognition memory test.

The stage-specific account allows for the possibility that perceptual disfluency and lexical difficulty may interact to engage greater extra- or post-lexical processing, particularly when both factors are present—such as in the case of blurred, low-frequency words. While each manipulation alone may be resolved through more effortful lexical access (reflected in changes to μ), their combination might exceed the limits of lexical-level compensation, thereby recruiting additional control or decision-related processes.

Accordingly, this view predicts a potential interaction between blurring and word frequency on β, reflecting increased engagement of late-stage, resource-limited mechanisms. If blurred, low-frequency stimuli are associated with especially long response time tails, this may indicate that processing demands extend beyond initial encoding and into later stages of cognitive control or response selection. However, this increased engagement may not always lead to a memory advantage. When both perceptual and lexical demands are high, limited cognitive resources could become overtaxed, potentially attenuating or eliminating any downstream mnemonic effects.

With respect to memory, the stage-specific account assumes that not all items benefit equally from perceptual disfluency. Low-frequency words—like highly blurred words—already tend to attract increased attention during encoding [see pupillometry evidence; Kuchinke et al. (2007)]. As a result, the attentional boost provided by disfluency may be redundant for these items. In contrast, high-frequency words—which are typically processed more automatically— may benefit more from disfluency manipulations that increase attention and depth of encoding.

Consistent with this view, Ptok et al. (2019) argued that memory benefits from disfluency or conflict arise primarily under conditions that are otherwise fluent and encoding-poor. When tasks already require sustained attention or effort, these benefits are likely to diminish. For instance, Ptok et al. (2020) found that stabilizing participants' gaze with a chinrest—thus increasing endogenous attention—disrupted memory gains from semantic interference. Similarly, Geller and Peterson (2021) showed that informing participants of an upcoming memory test, thereby prompting deliberate encoding, reduced the disfluency effect. In a related study, Westerman and Greene (1997) (Experiment 3) found that when encoding effort was uniformly high—such as when participants had to spell rather than simply read words—the memory advantage for disfluent stimuli disappeared.

Taken together, the stage-specific account suggests that disfluency effects on memory may depend on the degree to which processing demands extend beyond lexical access and engage post-lexical control mechanisms. If blurring and lexical difficulty interact to increase these demands—as reflected in longer response time tails (i.e., elevated β)—we may observe downstream consequences for memory. However, any mnemonic benefit is expected to emerge primarily for stimuli that are otherwise processed fluently, such as high-frequency words, and only when cognitive resources remain available to support deeper encoding. The present experiment examines these possibilities by testing whether blurring and word frequency interact to influence early and late components of processing (as indexed by ex-Gaussian parameters) and whether these effects are associated with differences in subsequent memory performance.

In Experiment 2, we opted not to use the DDM due to its substantial computational demands, particularly given the size and complexity of our dataset. Instead, we focused on the ex-Gaussian decomposition of RTs during encoding. Findings from Experiments 1A and 1B revealed systematic associations between ex-Gaussian and DDM parameters: specifically, a positive relationship between μ and non-decision time (T_{er}), and a negative relationship between

β and drift rate (v). Based on these observed correspondences, we interpret the ex-Gaussian parameters as indirect markers of early (encoding-related) and late (decision-related) processing stages.

Method

Transparency and Openness

This study was preregistered https://osf.io/kjq3t. All raw and summary data, materials, and R scripts for pre-processing, analysis, and plotting for Experiment 2 can be found at our OSF page: https://osf.io/6sy7k/.

Participants

We preregistered a sample size of 432, which is twice the size of Experiments 1A and 1B. This sample size was chosen based on the interaction effect we aimed to test. All participants were recruited through SONA and Prolific. Participants recruited via Prolific were compensated $12 per hour. On Prolific, we applied built-in filters to include only monolingual, native English-speaking Americans currently residing in the United States, with normal or corrected-to-normal vision. Participants recruited via SONA were given course credit for their participation.

Materials

One hundred and eighty words (half low-frequency and half high-frequency) were adapted from Fernández-López et al. (2022). We further selected an additional 45 animal words from their stimuli. To make the experiment more feasible for online participants and to balance our conditions, we split the remaining non-animal words and presented 90 (half high-frequency, half low-frequency) non-animal words along with the 45 animal words during the study phase. This maintained the 2:1 ratio of non-animal to animal words used in previous experiments (e.g., Fernández-López et al. (2022); Perea et al. (2018)).

At test, an additional 90 non-animal words not shown during the study phase served as "new" items in the recognition task. We created six counterbalanced lists to ensure that each

word appeared as both "old" and "new," and under each of the three blurring conditions (clear, high blur, low blur) across participants. Similar to nonwords in Experiments 1A and 1B, animal words were excluded from the final analysis.

The animal word stimuli used by Fernández-López et al. (2022) varied in length (M = 5.3 letters; range: 3–9), but their average length closely matched that of the non-animal words (high-frequency: M = 5.3, range: 3–8; low-frequency: M = 5.3, range: 3–9). Animal words also covered a wide range of frequencies in the SUBTLEX database (M = 11.84 per million; range: 0.61–192.84).

Procedure

We used the same procedure as Experiments 1B. The main difference is that instead of making a word/non-word decision, participants made a semantic categorization judgement (i.e., "is an animal" or "is not an animal"). You can view the task here: https://run.pavlovia.org/Jgeller 112/hf_lf_sem_1.

Results

Accuracy

We started with 38880. After we removed RTs below .2 and above 2.5 (0.009)cwe were left with 38526 data points.

Table 8

Summary of posterior distributions for fixed effects in the Bayesian accuracy model (Experiment 2)

Hypothesis	Mean	SE	90% CrI	ER	Posterior Prob
High Blur < (Low Blur + Clear)	-0.91	0.21	[-1.263, -0.56]	Inf	1.00
Low Blur = Clear	-0.19	0.21	[-0.628, 0.203]	0.57	0.36
High Freq = Low Freq	0.15	0.20	[-0.215, 0.548]	0.67	0.40
Blur × Frequency (High vs. Low/Clear) = 0	-0.01	0.23	[-0.47, 0.438]	0.73	0.42
Blur × Frequency (Low vs. Clear) = 0	0.05	0.24	[-0.406, 0.554]	0.72	0.42

Note. 95% CrI for equivalency tests

The full model summary for accuracy is presented in Table 8. There was strong evidence that high-blurred words were associated with lower accuracy compared to low-blurred and clear words, $b = -0.914$, 90% CrI $[-1.263, -0.56]$, ER = Inf.

The evidence for an accuracy difference between low-blurred and clear words was weak, $b = -0.186$, 90% CrI $[-0.628, 0.203]$, ER = 0.571. The credible interval spanned zero, and the evidence ratio suggested only weak support for either hypothesis.

The evidence for a frequency effect was similar weak , $b = 0.146$, 95% CrI [–0.215, 0.548], ER $= 0.665$.

Finally, interaction terms between blurring and word frequency yielded posterior distributions centered near zero, with 95% CrIs that included zero. Evidence ratios for these terms were close to 1, indicating substantial uncertainty and no clear preference for either the null or alternative hypothesis.

RTs: Ex-Gaussian

Given the complexity of the model, we employed stronger priors to facilitate convergence. For the μ parameter, we specified the same prior used in Experiments 1A and 1B. For the β parameter, we applied a more constrained prior: $Normal(0, 0.25)$. Default priors were retained for all remaining model parameters.

Table 9

Frequency effect by blurring level

Blur	High	Low	Frequency Effect
Clear	705.30	723.09	18.00
High Blur	960.98	1,016.44	55.00
Low Blur	717.93	730.44	13.00

We had 38152 correct RT trials for non-animal responses. After removing RTs below .2 and above 2.5 we were left with 37823 trials.

Table 10

Summary of posterior distributions for fixed effects in the ex-Gaussian model (Experiment 2).

Hypothesis	Parameter	Mean	SE	90% CrI	ER	Posterior Prob
High Blur > (vs. Clear/Low Blur)	Mu	0.17	0.01	[0.164, 0.18]	Inf	1.00
Low Blur > Clear	Mu	0.01	0.00	[0.008, 0.013]	Inf	1.00
High Frequency < Low Frequency	Mu	-0.02	0.01	[-0.026, -0.011]	Inf	1.00
High Blur (vs. Low Blur/Clear) × Frequency	Mu	-0.01	0.01	[-0.02, 0.009]	1,024.59	1.00
Low Blur (vs. Clear) × Frequency	Mu	-0.00	0.00	[-0.009, 0.002]	1,453.96	1.00
High Blur > (vs. Clear/Low Blur)	Sigma	0.64	0.04	[0.562, 0.706]	Inf	1.00
Low Blur < Clear	Sigma	-0.01	0.03	[-0.061, 0.045]	1.41	0.58
High Frequency < Low Frequency	Sigma	-0.05	0.04	[-0.108, 0.01]	11.05	0.92
High Blur (vs. Low Blur/Clear) x Frequency	Sigma	0.08	0.07	[-0.031, 0.197]	7.67	0.89
Low Blur (vs. Clear) × Frequency	Sigma	-0.03	0.06	[-0.129, 0.065]	2.31	0.70
High Blur > (vs. Clear/Low Blur)	beta	0.55	0.03	[0.499, 0.603]	Inf	1.00
Low Blur > (vs. Clear)	Beta	-0.01	0.02	[-0.055, 0.027]	9.72	0.91
High Frequency < Low Frequency	Beta	-0.07	0.03	[-0.114, -0.016]	67.18	0.98
High Blur (vs. Low Blur/Clear) × Frequency	Beta	-0.14	0.05	[-0.222, -0.056]	332.33	1.00
Low Blur (vs. Clear) × Frequency	Beta	0.06	0.04	[0, 0.129]	19.52	0.95

Note. 95% CrI for equivalency tests.Posterior probability indicates the proportion of the posterior distribution that falls on one side of zero (either positive or negative), representing the probability that the effect is greater than or less than zero. Sigma and Beta parameters are on the log (logit) scale.

Table 10 provides a model summary. Figure 6 visualizes RTs as quantile and delta plots highlighting how blurring and word frequency affected processing during word recognition. High-blurred words showed greater central tendency shifts (μ) compared to the average of clear and low-blurred words,

$b = 0.172, 90\%$ CrI $[0.164, 0.18]$, ER $=$. Low-blurred words also showed greater shifting compared to clear words,

$b = 0.01, 90\%$ CrI $[0.008, 0.013]$, ER $=$.

There was a consistent effect of word frequency: high-frequency words exhibited less shifting than low-frequency words, $b = -0.018, 90\%$ CrI $[-0.026, -0.011]$, ER $=$.

The posterior distributions for blur $\times$ frequency interactions were centered near zero, providing little support for interaction effects. Specifically, the High Blur (vs. Clear/Low Blur) $\times$ Frequency interaction was, $b = -0.006, 90\%$ CrI $[-0.02, 0.009]$, ER $= 1024.587$, and the Low Blur (vs. Clear) $\times$ Frequency interaction was $b = -0.004, 90\%$ CrI $[-0.009, 0.002]$, ER $= 1453.958$.

There was evidence for greater variability ($\boxed{?}$) in response times for high-blurred words relative to the average of clear and low-blurred words, $b = 0.635, 90\%$ CrI $[0.562, 0.706]$, ER $=$. The contrast between low-blurred and clear words showed no reliable difference in variability, $b = -0.007, 90\%$ CrI $[-0.061, 0.045]$, ER $= 1.409$.

There was modest support for a frequency effect on variability: high-frequency words showed less variability than low-frequency words, $b = -0.05, 90\%$ CrI $[-0.108, 0.01]$, ER $= 11.048$.

Blur × Frequency interactions on variability were mixed. For the High Blur (vs. Clear/ Low Blur) × Frequency interaction: $b = 0.083$, 90% CrI $[-0.031, 0.197]$, ER $= 7.675$, and for the low blur (vs. clear) × grequency interaction: $b = -0.031$, 90% CrI $[-0.129, 0.065]$, ER $= 2.307$.

Posterior estimates indicated greater skew ($\boxed{?}$) for high-blurred words compared to the average of clear and low-blurred words, $b = 0.551$, 90% CrI $[0.499, 0.603]$, ER $=$. The difference in skew between low-blurred and clear words was negligible, $b = -0.014$, 90% CrI $[-0.055, 0.027]$, ER $= 9.724$.

There was strong evidence that high-frequency words were associated with less skew than low-frequency words, $b = -0.065$, 90% CrI $[-0.114, -0.016]$, ER $= 67.182$.

The interaction between high blur (vs. clear/low blur) and frequency on skew was supported: $b = -0.139$, 90% CrI $[-0.222, -0.056]$, ER $= 332.333$.

In contrast, the low blur (vs. clear) × fequency interaction showed strong evidence for greater skewing for high frequency words (even though CrI included 0), $b = 0.064$, 90% CrI $[0, 0.129]$, ER $= 19.525$.

Table 11

Word frequency effect across .1, .3, .5, .7, .9 quantiles as a function of blurring

Blur	0.1	0.3	0.5	0.7	0.9
Clear	12.10	17.17	18.59	19.24	25.06
High Blur	31.46	41.84	48.68	67.18	98.36
Low Blur	9.93	12.14	11.99	13.62	17.73

Figure 6

Group RT distributions in the blurring and word frequency manipulations in word stimuli. A. Quantile plots with each point represents the average RT quantiles (.1, .3, .5, .7,and .9) in each

condition. B. Delta plots obatined by computing the quantiles for each participant and subsequently averaging the obtained values for each quantile over the participants and subtracting the values from each condition.
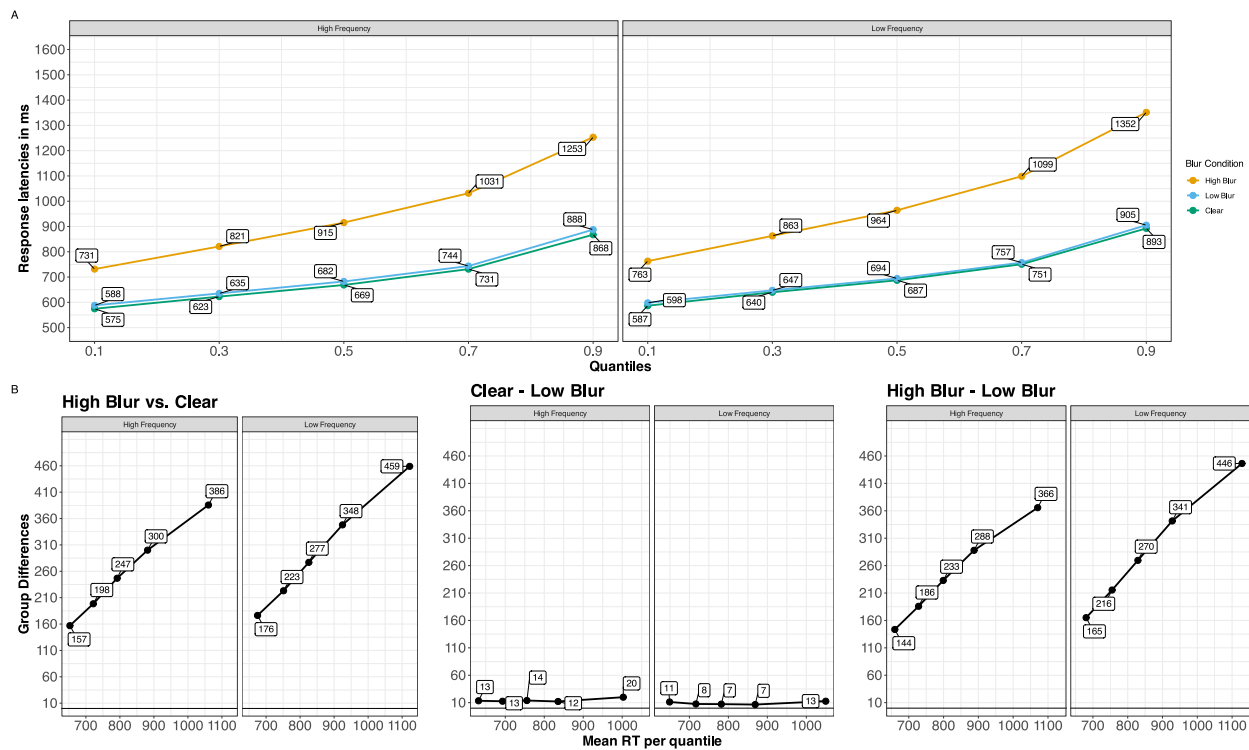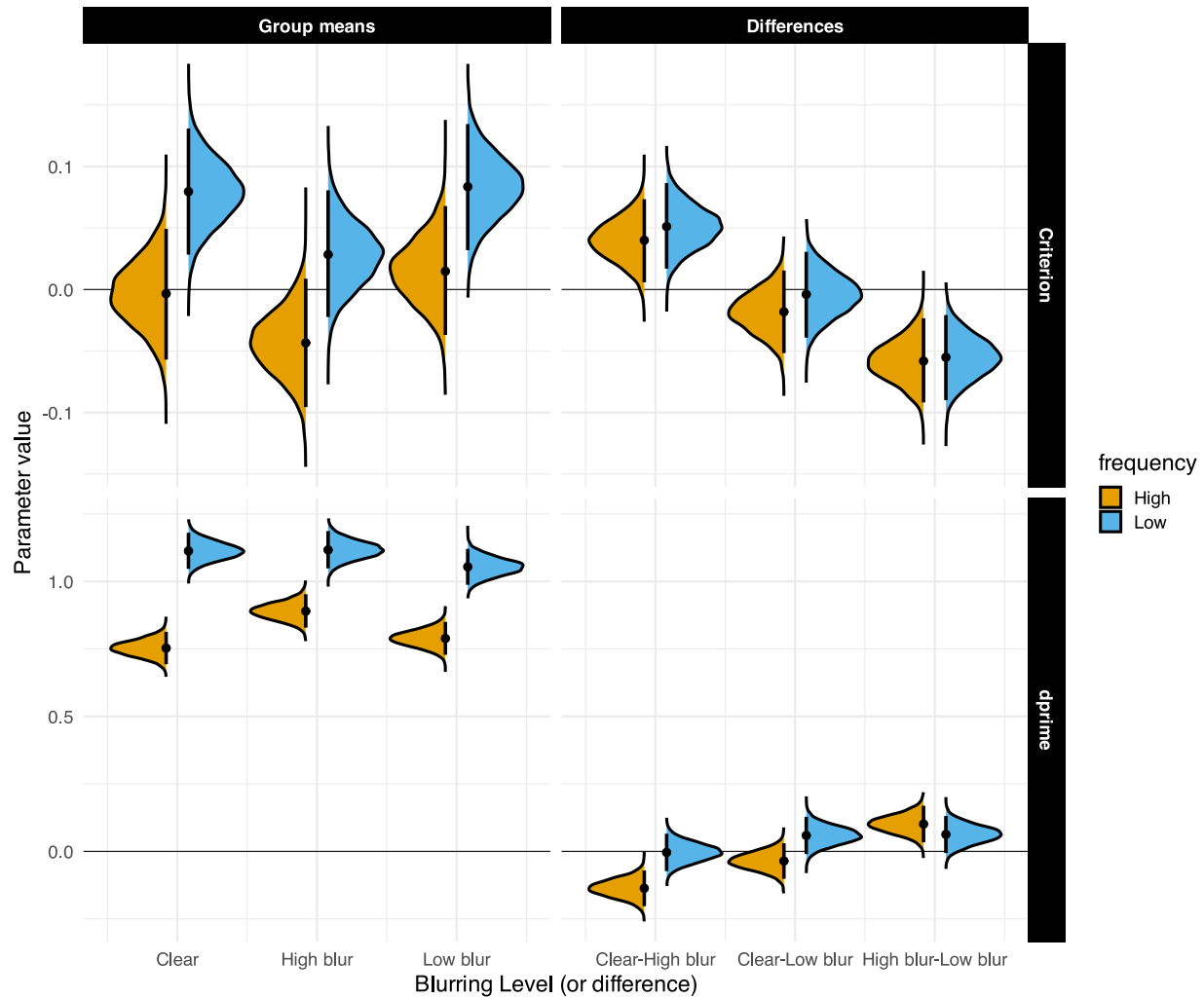

Recognition Memory

To aid in model convergence, we used a binomial (probit) family rather than a Bernoulli family when modeling our results. While both approaches are appropriate for binary outcomes, the binomial model allowed us to aggregate responses within each condition, reducing the number of observations and easing computational demands. This aggregation improved sampling efficiency and stability, particularly given the complexity of the full factorial structure. However, as a trade-off, we were unable to model random intercepts and slopes for individual items, since aggregation collapses trial-level variability.

Figure 7

Estimated posterior distributions for d-prime and criterion, and differences between all conditions, 95% CrIs (thin lines)

Sensitivity. Consistent with Experiments 1A and 1B, sensitivity in recognition memory was higher for high-blurred words compared to both clear and low-blurred words (see Figure 7), $\beta = 0.076$, 90% CrI $[0.041, 0.112]$, ER = 4999. The difference in recognition between clear and low-blurred words was negligible, with strong evidence in favor of the null hypothesis: $\beta = -0.012$, 95% CrI $[-0.06, 0.036]$, ER = 4.911.

Low-frequency words were more likely to be recognized than high-frequency words, $\beta = -0.27$, 90% CrI $[-0.306, -0.234]$, ER = 0. Crucially, there was strong evidence for an interaction between blur and frequency, such that high-frequency words showed a selective memory benefit under high-blur conditions: $\beta = 0.086$, 95% CrI $[0.017, 0.155]$, ER = 47.544.

A second interaction emerged for the comparison between low-blur and clear words: $\beta = 0.095$, 90% CrI $[0.016, 0.174]$, ER $= 37.911$. In this case, recognition was better for low-frequency than high-frequency words when stimuli were presented clearly.

Discussion

Experiment 2 probed how late-stage processing underlies the mnemonic boost that disfluent stimuli sometimes confer—and how that boost interacts with word frequency. Looking at accuracy, we did not see any differences between the levels of blurring, or any interactions with frequency. We did see an accuracy difference between high frequency and low frequency words. This largely corresponds with what Fernández-López et al. (2022) found.

In line with Experiments 1A and 1B, we observed systematic effects of visual blurring on the response time distribution. High blur led to greater shifting and skewing of the distribution compared to both low blur and clear conditions, while low blur produced more shifting than clear text. Additionally, word frequency significantly affected both the location (μ) and skewness (β) parameters: low-frequency words were associated with slower responses and increased positive skew.

Critically, we observed a significant interaction between blurring and word frequency on the β parameter. Under high blur, the frequency effect was magnified, with low-frequency words exhibiting a substantially longer tail compared to both low and no blur conditions—indicating more pronounced delays in late-stage processing.

Turning to memory, both blurred conditions elicited a perceptual disfluency effect relative to their respective baselines. High-frequency words showed a clear memory benefit under both high and low blur, whereas low-frequency words did not. This pattern is particularly striking when considered alongside the response time data. High blur produced a significant interaction on the β parameter, with low-frequency words showing a disproportionately longer tail—indicative of greater late-stage processing difficulty. Yet, despite this increased effort, low-

frequency words did not exhibit a corresponding memory advantage. In contrast, high-frequency words, which showed relatively smaller β increases, did benefit from the disfluency.

The present findings suggest that late-stage processing may be necessary but not sufficient for disfluency-based memory benefits. In our study, high blurring magnified the word frequency effect on β, indicating more prolonged or effortful late-stage processing—particularly for low-frequency words. However, memory benefits were observed only for high-frequency words, implying that this added processing must be paired with robust lexical access to translate into improved encoding.

Interestingly, low blur words also interacted with frequency, producing a memory advantage for high-frequency words. While this may initially seem counterintuitive—given that low blur did not significantly increase β—a closer look reveals a more nuanced picture. Although the 95% credible interval for the blur–frequency interaction on β included zero, the evidence ratio (ER = 19.10) provided substantial evidence for a positive effect. Intriguingly, the direction of the effect was reversed from the typical pattern: under low blur, high-frequency words exhibited greater late-stage processing than low-frequency words.

This suggests that high-frequency words may have been more impaired by low blurring, possibly because the disfluency disrupted the otherwise automatic processing of these familiar items, prompting additional extra-lexical processing. Yet, rather than hindering performance, this modest disruption may have facilitated deeper encoding, yielding a selective memory benefit. In contrast, low-frequency words may not have been processed deeply enough to benefit from the added perceptual difficulty. This finding highlights that memory benefits from disfluency are not solely a function of increased processing effort, but rather emerge when perceptual challenge intersects with lexical accessibility in a way that supports enriched encoding.

This asymmetry suggests that modest perceptual disfluency can amplify processing of high-frequency words, possibly by prompting additional elaboration or semantic integration,

while not overburdening cognitive resources. In contrast, high blur induces excessive late-stage processing, particularly for low-frequency words, as reflected in increased β values—without producing corresponding memory benefits. Together, these findings support the idea that additional processing is beneficial only when the perceptual challenge is manageable and the lexical stimulus is sufficiently accessible. When disfluency becomes too demanding, as under high blur, it may exceed processing capacity—especially for low-frequency words—ultimately hindering memory encoding rather than helping it.

General Discussion

Interfering with stimulus perception during encoding can sometimes improve later explicit memory. The mixed data on perceptual disfluency has called into question the utility of such manipulations in the learning domain. One of the main aims of the current set of experiments was to examine the underlying mechanisms of the perceptual disfluency effect to better understand when perceptual disfluency aids memory and when it does not. To this end, our study delved into the impact of one type of perceptual disfluency–blurring (i.e., low blurring and high blurring)–on the process of encoding, as assessed through a LDT (Experiments 1A and 1B), and a semantic categorization task (Experiment 2). RT distributions were analyzed with an ex-Gaussian model (Experiments 1A, 1B, 2) and DDM (Experiments 1A and 1B) to better understand how perceptual disfluency affects encoding. These models offered a comprehensive descriptive and theoretical framework through which to examine the perceptual disfluency effect.

To recapitulate our findings, during encoding, high blurred words showed greater distributional shifting and skewing compared to clear and low blurred words. In addition, DDM fits indicated high blurred words had a higher T_{er} and lower v. Conversely, low blurred words compared to clear words showed greater distributional shifting, but there was no difference in skewing. DDM fits showed higher T_{er}, but had no effect on v.

Turning to recognition memory, high blurred words were more likely to be recognized at test compared to clear words and low blurred words. This pattern arose regardless if context was reinstated at test (Experiment 1B). This pattern replicates the results from Rosner et al. (2015). In addition, we showed word frequency (Experiment 2) also modulates the disfluency effect. Namely, low frequency words did not show a disfluency effect.

These findings have several implications. At a theoretical level, the current data suggests that in order for perceptual disfluency to benefit memory it has to be disfluent enough to affect both early and late stages of processing. A manipulation that only produces a general slowing of responses is not sufficient to enact an mnemonic effect. However, an important caveat to this is that processes during encoding of the word itself are not enough to produce an menominc benefit. In Experiment 2, we did not observe better memory for low frequency-high blurred words which are the hardest and presumably receive the most top-down processing. We only observed a disfluency effect for high frequency-high blurred words. This points to the importance of control processes and processing limitations in producing the disfluency effect.

We argue the current findings align more closely with the stage-specific account proposed by Ptok et al. (2019). While the account was proposed to explain memory effects that require conflict during encoding, like the semantic interference effect, we feel that it is a useful theoretical framework to explain the current findings. In fact, Ptok et al. (2019) and Ptok et al. (2020) suggested a connection between their framework and perceptual disfluency effects and desirable difficulties, more broadly.

Within the stage-specific account, memory performance depends on the nature of processing during encoding and the utilization of cognitive control mechanisms. In our experiments, participants were tasked with determining whether letter strings represented words or non-words (Experiments 1A and 1B), or whether a word belonged to an animal category (Experiment 2). For skilled readers, these tasks are executed automatically and smoothly.

Coupled with perceptual disfluency, this combination is believed to lead to memory advantages seen with perceptually disfluent stimuli.

When we manipulated word frequency, however, the process of recognizing low-frequency words demanded more effort and attentional resources in addition to the perceptual disfluency of blurring. Consequently, this lead to the task becoming more challenging and difficult. Thus, the increased processing demands from recognizing low frequency words may have countered the benefits from high-blurred words, as more attentional resources were allocated to recognizing low-frequency words.

There are other examples that support this capacity-limited view of perceptual disfluency. For instance, Geller et al. (2018) showed that easy-to-read cursive words and hard-to-read cursive words are better remembered than computer print words, but the memory effect is much larger for easier to read cursive words. As another notable example, participants with low working memory capacity do not seem to benefit from perecptaul disfluency as much as those with higher working memory capacity (Lehmann et al., 2015). At a broader level, Wenzel and Reinhard (2019) suggested that intelligence is an important factor for when desirable difficulties are desirable for learning.

At a methodological level, our experiments demonstrate that a straightforward blurring manipulation can benefit memory, which we observed whether or not we reinstated the context during testing. However, blurring has to be sufficiently difficult do so. If the secondary task requires too much attentional control the effect might not be observed or attenuated.

More significantly, our current experiments underscore the benefits of using mathematical and computational models—such as the ex-Gaussian model and DDM to examine perceptual disfluency during encoding. In Experiments 1A and 1B, both models converged on similar findings. Specifically, response time distributions were differential affected by the degree of visual blur. Words with low levels of blur primarily influenced early or non-decision stages of

processing (reflected in parameters such as T_{er} and μ), whereas highly blurred words impacted both early and later stages (T_{er}, τ, and v). These findings suggest that both the DDM and ex-Gaussian model are sensitive to perceptual disfluency and can be used to uncover underlying cognitive mechanisms during encoding (see also gomez2014). Although the models converged on similar patterns, it remains an open question whether one should be favored over the other. However, in terms of directly linking disfluency to cognitive processes, the DDM offers a clearer theoretical framework and may be the more informative model in this context.

Furthermore, our distrbution modeling of RTs appears to be a more sensitive method. Although we found weak evidence for differences between clean and low blurred conditions in measures like accuracy, we did notice variations in non-decision time and a shift in the response time distribution for low blurred words compared to clear words. We recommend that future studies employ distribution modeling and DDM to decompose response times and directly quantify the impact of perceptual disfluency on encoding.

While we applied the DDM to inform processes during encoding, there is one case of the DDM being applied to the study of perceptual disfluency during retrieval. In one recent study, Hu et al. (2022) examined RTs during retrieval. Hu et al. (2022) were primarily interested in how perceptual disfluency manipulated with Sans Forgetica typeface influenced DDM parameters and confidence judgments during recognition. At test, they found a non-significant difference in mean RTs between Sans Forgetica typeface and Arial typeface. However, looking at the DDM parameters, Sans Forgetica and Airal typefaces differed on T_{er}, but not drift rate. They also looked at how parameters of the DDM were related to confidence judgments. Higher T_{er} was related to lower confidence and higher drift rate was associated with higher confidence. While their focus was on retrieval and not encoding, this corresponds with what we observed for our low blurred words in our experiments–a weak manipulation affected non-decision time, but not

drift rate. Overall, this further highlights the utility of the the DDM in studying perceptual disfluency (and other encoding conflict effects) during encoding and at test.

Finally, at a practical level, we do show that blurring can benefit later memory. However, caution needs to be taken here. First, the current experiments were conducted online using simple materials (i.e., list learning). It is unclear how these effects would generalize to a classroom setting or more educational realistic materials (see Geller et al. (2020)). Second, participants were not told about the upcoming recognition test. Geller and Peterson (2021) has showed that low test expectancy is an important moderator for this effect. Third, while we did not establish a region of practical equivalence, the size of the disfluency effect appears to be be small in nature. Looking at the default region of proximal equivalence (ROPE) from the `bayestestr` package (here -0.10 -.10 in standardized units) many of the critical contrasts either were completely inside this region or overlapped this region, suggesting negligible differences. For more applied work, this might be well below the smallest effect size of interest. Now this is not to say all research on perceptual disfluency is superfluous. As noted in Geller and Peterson (2021), a fruitful avenue for future work might be to investigate how perceptual influenced effects processing in every day life where memory is largely incidental (Castel et al., 2015).

These results provide some context for the large number of replication failures. Many of the studies looking at disfluency do not take care in ensuring the disfluency manipulation is actually disfluent. Most times, studies use only two levels (disfluent/fluent) and perform analyses that may not be well suited for the type of data they have. As we have hopefully shown here, it is important to take into account the entire RT distribution. By examining the RT distributions of different levels of disfluency we obtained a richer better understanding of the stages or loci manipulations have during encoding. It is our hope that learning and memory researches will begin to use these tools to help understand encoding processes involved in perceptual disfluency effects but also encoding contexts where there is considerable conflict.

Conclusion

Our paper contributes nuanced insights to the intricate relationship between perceptual disfluency and memory encoding. We have shown that perceptual disfluency can aid in memory retention, but its efficacy is contingent upon the degree of disfluency and other contextual factors such as word frequency. Our findings endorse the stage-specific account, emphasizing the role of cognitive control mechanisms in the observed memory advantages with perceptual disfluency. Furthermore, our methodological contributions, employing an ex-Gaussian model and DDM, not only validate the benefits of examining RT distributions, but also open new avenues for future research in learning and memory studies. We caution, however, that the applicability of these findings in real-world educational settings remains an open question, and the effect sizes observed were relatively small, thus warranting further investigation.

Ultimately, this work stands as a call to action for a more comprehensive, nuanced approach to studying perceptual disfluency, incorporating both advanced statistical methods and a more exhaustive range of experimental conditions to better elucidate when and how disfluency can facilitate memory.

References

Allaire, J. J., Teague, C., Scheidegger, C., Xie, Y., Dervieux, C., & Woodhull, G. (2024). *Quarto* (Version 1.6) [Computer software]. https://doi.org/10.5281/zenodo.5960048

Alter, A. L. (2013). The Benefits of Cognitive Disfluency. *Current Directions in Psychological Science*, *22*(6), 437–442. https://doi.org/10.1177/0963721413498894

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 514–544. https://doi.org/10.1037/0278-7393.27.2.514

Angele, B., Baciero, A., Gómez, P., & Perea, M. (2023). Does online masked priming pass the test? The effects of prime exposure duration on masked identity priming. *Behavior Research Methods*, *55*(1), 151–167. https://doi.org/10.3758/s13428-021-01742-y

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, *128*(1), 32–55. https://doi.org/10.1037/0096-3445.128.1.32

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry the power of response time distributional analyses. *Current Directions in Psychological Science*, *20*(3), 160166. http://cdp.sagepub.com/content/20/3/160.short

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language*, *59*(4), 495–523. https://doi.org/10.1016/j.jml.2007.10.004

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/bf03193014

Barnhart, A. S., & Goldinger, S. D. (2010). Interpreting chicken-scratch: Lexical access for handwritten words. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(4), 906–923. https://doi.org/10.1037/a0019258

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barthelme, S. (2023). *Imager: Image processing library based on 'CImg'*. https://CRAN.R-project.org/package=imager

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory

performance. *Memory & Cognition*, *41*(6), 897903. http://link.springer.com/article/10.3758/s13421-013-0307-8

Boag, R. J., Strickland, L., Heathcote, A., Neal, A., & Loft, S. (2019). Cognitive control and capacity for prospective memory in complex dynamic environments. *Journal of Experimental Psychology: General*, *148*(12), 2181–2206. https://doi.org/10.1037/xge0000599

Boag, R. J., Strickland, L., Loft, S., & Heathcote, A. (2019). Strategic attention and decision control support prospective memory in a complex dual-task environment. *Cognition*, *191*, 103974. https://doi.org/10.1016/j.cognition.2019.05.011

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Bürkner, P.-C. (2017). *Brms: An r package for bayesian multilevel models using stan*. *80*. https://doi.org/10.18637/jss.v080.i01

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology*, *1*(9), 496–511. https://doi.org/10.1038/s44159-022-00089-1

Castel, A. D., Nazarian, M., & Blake, A. B. (2015). *Attention and incidental memory in everyday settings* (pp. 463–483). Boston Review.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204256. https://doi.org/10.1037/0033-295X.108.1.204

Cushing, C., & Bodner, G. E. (2022). Reading aloud improves proofreading (but using sans forgetica font does not). *Journal of Applied Research in Memory and Cognition*, *11*, 427–436. https://doi.org/10.1037/mac0000011

De Jong, R., Liang, C. C., & Lauber, E. (1994). Conditional and unconditional automaticity: a dual-process model of effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(4), 731–750. https://doi.org/10.1037//0096-1523.20.4.731

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, *3*(2), 186205. https://doi.org/10.1037/1082-989X.3.2.186

Diana, R. A., & Reder, L. M. (2006). The low-frequency encoding disadvantage: Word frequency affects processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 805–815. https://doi.org/10.1037/0278-7393.32.4.805

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the: Effects of disfluency on educational outcomes. *Cognition*, *118*(1), 111115. https://doi.org/10.1016/j.cognition.2010.09.012

Diffusion/wiener model analysis with brms part i: Introduction and estimation. (2017). http://singmann.org/wiener-model-analysis-with-brms-part-i/

Dolstra, E., & contributors, T. N. (2023). *Nix* (Version 2.15.3) [Computer software]. https://nixos.org/

Earp, J. (2018, October 8). *Q&a: Designing a font to help students remember key information*. https://www.teachermagazine.com/au_en/articles/qa-designing-a-font-to-help-students-remember-key-information

Eskenazi, M. A., & Nix, B. (2021). Individual differences in the desirable difficulty effect during lexical acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 45–52. https://doi.org/10.1037/xlm0000809

Fernández-López, M., Gómez, P., & Perea, M. (2022). Letter rotations: through the magnifying glass and What evidence found there. *Language, Cognition and Neuroscience*, *38*(2), 127–138. https://doi.org/10.1080/23273798.2022.2093390

Fitousi, D. (2020). Decomposing the composite face effect: Evidence for non-holistic processing

   based on the ex-Gaussian distribution. *Quarterly Journal of Experimental Psychology*,

   73(6), 819–840. https://doi.org/10.1177/1747021820904222

Geller, J., Davis, S. D., & Peterson, D. J. (2020). Sans forgetica is not desirable for learning.

   Memory, *28*(8), 957–967. https://doi.org/10.1080/09658211.2020.1797096

Geller, J., & Peterson, D. (2021). Is this going to be on the test? Test expectancy moderates the

   disfluency effect with sans forgetica. *Journal of Experimental Psychology: Learning,*

   Memory, and Cognition, *47*(12), 1924–1938. https://doi.org/10.1037/xlm0001042

Geller, J., Still, M. L., Dark, V. J., & Carpenter, S. K. (2018). Would disfluency by any other

   name still be disfluent? Examining the disfluency effect with cursive handwriting.

   Memory and Cognition, *46*(7), 11091126. https://doi.org/10.3758/s13421-018-0824-6

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory &*

   Cognition, *13*(1), 8–20. https://doi.org/10.3758/bf03198438

Gomez, P., & Perea, M. (2014). Decomposing encoding and decisional components in visual-

   word recognition: A diffusion model analysis. *Quarterly Journal of Experimental*

   Psychology, *67*(12), 2455–2466. https://doi.org/10.1080/17470218.2014.937447

Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked versus

   unmasked priming: Are they qualitatively different? *Journal of Experimental*

   Psychology: Human Perception and Performance, *39*(6), 1731–1740. https://doi.org/10.

   1037/a0032333

Grant, R. L., Carpenter, B., Furr, D. C., & Gelman, A. (2017). Introducing the StataStan Interface

   for Fast, Complex Bayesian Modeling Using Stan. *The Stata Journal: Promoting*

   Communications on Statistics and Stata, *17*(2), 330–342. https://doi.org/10.1177/

   1536867x1701700205

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley.

Halamish, V. (2018). Can very small font size enhance memory? *Memory & Cognition*, *46*(6), 979–993. https://doi.org/10.3758/s13421-018-0816-6

Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*(2), 340–347. https://doi.org/10.1037/0033-2909.109.2.340

Hu, X., Yang, C., & Luo, L. (2022). Retrospective confidence rating about memory performance is affected by both retrieval fluency and non-decision time. *Metacognition and Learning*, *17*(2), 651–681. https://doi.org/10.1007/s11409-022-09303-0

Huff, M. J., Maxwell, N. P., & Mitchell, A. (2022). Distinctive Sans Forgetica font does not benefit memory accuracy in the DRM paradigm. *Cognitive Research: Principles and Implications*, *7*(1). https://doi.org/10.1186/s41235-022-00448-9

James, W. (1890). *The principles of psychology, vol i*. Henry Holt; Co. https://doi.org/10.1037/10538-000

Johnson, R. L., Staub, A., & Fleri, A. M. (2012). Distributional analysis of the transposed-letter neighborhood effect on naming latency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1773–1779. https://doi.org/10.1037/a0028222

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70. https://doi.org/10.1037/0096-3445.132.1.47

Kuchinke, L., Vo, M., Hofmann, M., & Jacobs, A. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology*, *65*(2), 132–140. https://doi.org/10.1016/j.ijpsycho.2007.04.004

Lehmann, J., Goussios, C., & Seufert, T. (2015). Working memory capacity and disfluency effect: an aptitude-treatment-interaction study. *Metacognition and Learning*, *11*(1), 89– 105. https://doi.org/10.1007/s11409-015-9149-z

Lenth, R. V. (2023). *Emmeans: Estimated marginal means, aka least-squares means*. https:// CRAN.R-project.org/package=emmeans

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). *bayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework*. *4*, 1541. https://doi.org/10.21105/joss.01541

Malmberg, K. J., & Nelson, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition*, *31*(1), 35–43. https://doi.org/10. 3758/bf03196080

Mathot, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1). https://doi.org/10.5334/joc.18

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-gaussian and shifted wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*(5), 798817. http://link.springer.com/article/10.3758/PBR.16.5.798

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375– 407. https://doi.org/10.1037/0033-295X.88.5.375

McElreath, R. (2020). *Statistical rethinking: A bayesian course with examples in r and stan* (2nd ed.). CRC Press.

Mulligan, N. W. (1996). The effects of perceptual interference at encoding on implicit memory, explicit memory, and memory for source. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1067–1087. https://doi.org/10.1037/0278-7393.22.5.1067

Myers, C. E., Interian, A., & Moustafa, A. A. (2022). A practical introduction to using the drift diffusion model of decision-making in cognitive psychology, neuroscience, and health sciences. *Frontiers in Psychology*, *13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1039172

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition*, *16*(4), 343–352. https://doi.org/10.3758/BF03197045

Pazzaglia, A. M., Staub, A., & Rotello, C. M. (2014). Encoding time and the mirror effect in recognition memory: Evidence from eyetracking. *Journal of Memory and Language*, *75*, 77–92. https://doi.org/10.1016/j.jml.2014.05.009

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perea, M., Fernández-López, M., & Marcet, A. (2018). Does CaSe-MiXinG disrupt the access to lexico-semantic information? *Psychological Research*, *84*(4), 981–989. https://doi.org/10.1007/s00426-018-1111-7

Perea, M., Gil-López, C., Beléndez, V., & Carreiras, M. (2016). Do handwritten words magnify lexical effects in visual word recognition? *Quarterly Journal of Experimental Psychology*, *69*(8), 1631–1647. https://doi.org/10.1080/17470218.2015.1091016

Perea, M., & Lupker, S. J. (2003). Transposed-letter confusability effects in masked form priming. *Masked Priming: State of the Art*, 97120. http://defiant.ssc.uwo.ca/faculty/lupkerpdfs/Perea%20&%20Lupker,%202003,%20chapter.pdf

Pieger, E., Mengelkamp, C., & Bannert, M. (2016). Metacognitive judgments and disfluency does disfluency lead to more accurate judgments, better control, and better performance? *Learning and Instruction*, *44*, 31–40. https://doi.org/10.1016/j.learninstruc.2016.01.012

Plourde, C. E., & Besner, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *51*(3), 181–194. https://doi.org/10.1037/1196-1961.51.3.181

Ptok, M. J., Hannah, K. E., & Watter, S. (2020). Memory effects of conflict and cognitive control are processing stage-specific: evidence from pupillometry. *Psychological Research*, *85*(3), 1029–1046. https://doi.org/10.1007/s00426-020-01295-3

Ptok, M. J., Thomson, S. J., Humphreys, K. R., & Watter, S. (2019). Congruency encoding effects on recognition memory: A stage-specific account of desirable difficulty. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00858

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*(4), 615625. https://doi.org/10.1037/a0013684

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin and Review*, *16*(3), 550554. https://doi.org/10.3758/PBR.16.3.550

Roberts, B. R. T., Hu, Z. S., Curtis, E., Bodner, G. E., McLean, D., & MacLeod, C. M. (2023). Reading text aloud benefits memory but not comprehension. *Memory & Cognition*. https://doi.org/10.3758/s13421-023-01442-2

Rodrigues, B., & Baumann, P. (2025). *Rix: Reproducible data science environments with 'nix'*.

https://docs.ropensci.org/rix/

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning. *Psychological Science*,

17(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning.

Instructional Science, *35*(6), 481–498. https://doi.org/10.1007/s11251-007-9015-8

Rosner, T. M., Davis, H., & Milliken, B. (2015). Perceptual blurring and recognition memory: A

desirable difficulty effect revealed. *Acta Psychologica*, *160*, 11–22. https://doi.org/10.

1016/j.actpsy.2015.06.006

Rummer, R., Schweppe, J., & Schwede, A. (2015). Fortune is fickle: Null-effects of disfluency

on learning outcomes. *Metacognition and Learning*, 114. http://link.springer.com/article/

10.1007/s11409-015-9151-5

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal

of Experimental Psychology: Human Learning and Memory*, *4*(6), 592–604. https://doi.

org/10.1037/0278-7393.4.6.592

Staub, A. (2010). The effect of lexical predictability on distributions of eye fixation durations.

Psychonomic Bulletin & Review, *18*(2), 371–376. https://doi.org/10.3758/s13423-010-

0046-9

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta

Psychologica*, *30*, 276–315. https://doi.org/10.1016/0001-6918(69)90055-9

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental

Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for

inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin

and Review*, *18*(5), 973978. https://doi.org/10.3758/s13423-011-0114-9

Taylor, A., Sanson, M., Burnell, R., Wade, K. A., & Garry, M. (2020). Disfluent difficulties are not desirable difficulties: The (lack of) effect of sans forgetica on memory. *Memory*, *28*(7), 850–857. https://doi.org/10.1080/09658211.2020.1758726

Taylor, J. E., Beith, A., & Sereno, S. C. (2020). LexOPS: An R package and user interface for the controlled generation of word stimuli. *Behavior Research Methods*, *52*(6), 2372–2382. https://doi.org/10.3758/s13428-020-01389-1

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, *16*(1), 44–62. https://doi.org/10.1037/a0021765

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved R^ for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, *16*(2), 667–718. https://doi.org/10.1214/20-BA1221

Vergara-Martínez, M., Gutierrez-Sigut, E., Perea, M., Gil-López, C., & Carreiras, M. (2021). The time course of processing handwritten words: An ERP investigation. *Neuropsychologia*, *159*, 107924. https://doi.org/10.1016/j.neuropsychologia.2021.107924

Weissgerber, S. C., & Reinhard, M. A. (2017). Is disfluency desirable for learning? *Learning and Instruction*, *49*, 199–217. https://doi.org/10.1016/j.learninstruc.2017.02.004

Wel, P. van der, & Steenbergen, H. van. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Weltman, D., & Eakin, M. (2014). Incorporating unusual fonts and planned mistakes in study materials to increase business student focus and retention. *INFORMS Transactions on Education*, *15*(1), 156165. https://doi.org/10.1287/ited.2014.0130

Wenzel, K., & Reinhard, M.-A. (2019). Relatively unintelligent individuals do not benefit from intentionally hindered learning: The role of desirable difficulties. *Intelligence*, *77*, 101405. https://doi.org/10.1016/j.intell.2019.101405

Westerman, D. L., & Greene, R. L. (1997). The effects of visual masking on recognition: Similarities to the generation effect. *Journal of Memory and Language*, *37*(4), 584–596. https://doi.org/10.1006/jmla.1997.2531

Wetzler, E. L., Pyke, A. A., & Werner, A. (2021). Sans Forgetica is Not the "Font" of Knowledge: Disfluent Fonts are Not Always Desirable Difficulties. *SAGE Open*, *11*(4), 21582440211056624. https://doi.org/10.1177/21582440211056624

Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, *53*(3), 300–314. https://doi.org/10.1037/0003-066x.53.3.300

Wit, B. de, & Kinoshita, S. (2015). The masked semantic priming effect is task dependent: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1062–1075. https://doi.org/10.1037/xlm0000074

Yap, M. J., & Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 274–296. https://doi.org/10.1037/0278-7393.33.2.274

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency isand is nota desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, *41*(2), 229241. http://link.springer.com/article/10.3758/s13421-012-0255-8

Zloteanu, M., & Vuorre, M. (2024). A Tutorial for Deception Detection Analysis or: How I Learned to Stop Aggregating Veracity Judgments and Embraced Signal Detection Theory

Mixed Models. *Journal of Nonverbal Behavior*, *48*(1), 161–185. https://doi.org/10.1007/s

10919-024-00456-x