# Perceptual Disfluency and Recognition Memory: A Response Time Distributional Analysis

Jason Geller1, Pablo Gomez2, Erin Buchanon3, and Dominique Makowski4

1Department of Psychology and Neuroscience, Boston College

2Department of Psychology, Skidmore College

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

4School 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

Experiment 1A and 2 were preregisted: https://osf.io/q3fjn; https://osf.io/kjq3t. Data, code, and materials for this manuscript can be found at https://osf.io/6sy7k/. The authors have no conflicts of interest to disclose. This work was supported by research start-up funds to JG.

Author roles were classified using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/) as follows: *Jason Geller***:** conceptualization, writing – original draft, data curation, writing – review & editing, software, and formal analysis. *Pablo Gomez***:** conceptualization, writing – review & editing, and formal analysis. *Erin Buchanon***:** writing – 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. Recent research suggests that this effect arises from interactions between early encoding and later decision-making processes. 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 these parameters, 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

## Introduction

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, 2021a; Halamish, 2018; Rosner et al., 2015). The resulting memory benefit has been called the perceptual disfluency effect (Geller et al., 2018). Many studies have begun to examine how changes in physical characteristics of to be learned materials influence encoding and memory.

## 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*, Narine 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, 2021a; 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 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 mixed. 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 it enhances memory retention due to the backward-slanting letters and gaps within each letter which requires 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 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 (2021a) 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).

Though perceptual disfluency can occur in certain scenarios, its usefulness in educational settings, where students anticipate tests, could be limited. However, Geller and Peterson (2021a) contended that perceptual disfluency has practical implications, particularly due to our reliance on incidental memory in everyday decision-making. For this to be effective, though, predicting when and where disfluency will occur is crucial.

## Theoretical accounts of the disfluency effect

To achieve this aim, we require 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 feedbackward. Thus, when there is perceptual noise (such as 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) (see Mathot (2018) for a review). They observed larger pupil size (which has been taken as an index of cognitive control; see (Wel & Steenbergen, 2018), for a review) 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 (2021a). In addition, requiring participants to make judgments of learning (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, 2021a; 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 examine RT distributions using mathematical models that capture the nuances of the RT distribution and consider various statistical properties, such as the shape ($\mu$) , spread ( ) , and skewness () 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 + . Together the three parameters represent different aspects of the distribution’s location and shape.

Heathcote et al. (1991) looked at this Stroop issue with an ex-Gaussian model and found both facilitation (from congruent trials) and interference (from incongruent trials) on . For , the analysis showed interference, but no facilitation. For , 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 + , the failure to observe a facilitation effect in the standard mean analysis likely arose from facilitation on and interference on 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 , but not . 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 . An example of an effect that only impacts 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 and , 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 and parameters have been tied to early, non-analytic, processing. In the area of semantic priming, the selective effect on 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 () has been tied to later, more analytic, processing (Balota & Spieler, 1999). Specifically, increases in have been attributed to working memory and attentional processes (Fitousi, 2020; Kane & Engle, 2003). For instance, Johnson et al. (2012) tied 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)

Contrary to the ex-Gaussian distribution discussed above, 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 () and non-decision time (ndt; ) parameters. Drift rate () 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) indicates a steeper slope, meaning that evidence is accumulated more quickly, leading to faster decisions. Conversely, a lower 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 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 component, but not 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 to enhance memory in the literature—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 tutorial 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 component. Specifically, a larger 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 , but not skewing . 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 ), but also change the shape of the distribution (an effect on ), 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 and parameters. Specifically, high blurred words would produce lower and high compared to clear and low blurred words. Additionally, we predicted that low blurred words would only affect . [Table 1](#tbl-predictions) summarizes each account of the perceptual disfluency and the predicted outcomes according to each mathematical model.

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.

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: No difference τ: Increase | v: No difference Tₑᵣ: Increase | Late Difference | No difference |
|  |  |  | Low blur vs. Clear | μ: No difference τ: Increase | v: No difference Tₑᵣ: Increase | Late Difference | No difference |
| Compensatory-processing | Perceptual disfluency affects the word recognition process | Lexical/semantic | High blur vs. Low blur/Clear | μ: Increase τ: No difference | v: Increase Tₑᵣ: No difference | Complete Shift | High blurred > Low blurred/Clear |
|  |  |  | Low blur vs. Clear | μ: Increase τ: No difference | v: Increase Tₑᵣ: No difference | No Difference | Low blurred > Clear |
| 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: Increase τ: Increase | v: Increase Tₑᵣ: Increase | Complete Shift + Late Differences | High blurred > Low blurred/Clear |
|  |  |  | Low blur vs. Clear | μ: Increase τ: Increase | v: No Increase Tₑᵣ: Increase | Early Difference | Low blurred = Clear |

## 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, 2021a), 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 δ ≥ 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 α = 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. Because of this 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 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 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](#fig-blur) 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, 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, LB, 28, HB) 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, LB old, HB old, clear new, LB new, and HB 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.

### Analysis Plan

All models were fit using The Stan modeling language (Grant et al., 2017) with the brms (Bürkner, 2017a) package in in R (v.4.4.2; R Core Team (2025)). For all reported models, we fit maximal random-effect structures justified by the design (Barr et al., 2013) *[REF should be bibtex]*. This included random intercepts for participants and items, and a random slope by blurring level for each varying random intercept. Contrast codes (effects coding) for each variable were created using the hypr package in R (Schad et al., 2019). We fit the models twice: Once with contrast codes for high blur vs. clear and for high blur vs. Low blur and once with the low blur vs. clear contrast.

In all experiments reported here, the statistical model was run with four chains of 5,000 Markov chain Monte Carlo iterations (For the DDM we fit a model with 2,000 chains to reduce computational time), with 1,000-iteration warm-ups for 4 chains (16,000 samples in total). Convergence and stability of the Bayesian sampling is quantified by the diagnostics below 1.01 and Effective Sample Size (ESS) greater than 1000 (Bürkner, 2017a). For both the RT data and accuracy data, we report our models with with weakly-informative priors for the population-level parameters. Using a weakly-informative prior as opposed to a default (which is an uniform prior where all effects are equally likely) allows for the calculation of evidence ratio (Bayes Factor) for one-sides tests. For the ex-Gaussian analysis we used a weak prior (i.e., N ~ (0, 100)). For the population-level effects in the accuracy and signal detection analyses, we used a Cauchy distribution with the mean of 0 and scale of 0.35 (cauchy ~ 0, 0.35)) recommended by (Kinoshita et al., 2023) for logistic regression.

For the marginal means and differences, we report the expected values under the posterior distribution and their 90% credible intervals (Cr. I.). For marginal mean differences, we also report the posterior probability that a difference δ is not zero. If a hypothesis states that δ \> 0, then it would be considered strong evidence for this hypothesis would be if zero is not included in the 90% Cr. I. of δ and the posterior P(δ \> 0) is close to one (by a reasonably clear margin). To extract the estimated marginal means from the posterior distribution of the fitted models we used a combination of emmeans R package (Lenth, 2023) bayestestr(Makowski et al., 2019), and brms (Bürkner, 2017b).

Model quality was thoroughly assessed via predictive prior and posterior checks, and divergence diagnostics. In order to assess the evidence in favor or against our hypotheses, we used Evidence Ratio (ER, a generalization of Bayes factors allowing for directional hypotheses). An ER above 3 indicates moderate to substantial evidence for our hypothesis, below 0.3 indicates moderate to substantial evidence for the null hypothesis, and anything in between is inconclusive evidence (Morey & Rouder, 2022).

#### Ex-Gaussian distribution.

We used the ex-Gaussian distribution to model response times, with both the mean of the Gaussian component 𝜇 and the scale parameter of the exponential component 𝛽 (equaling the inverse of the rate parameter 𝜆) being allowed to vary between conditions. In addition, to better visualize the distributional features of the latency data, we computed the delta plots for all variables.

#### Diffusion model.

We used a hierarchical-Bayesian variant of the Wiener diffusion model (Vandekerckhove et al., 2011) with accuracy coding. This model accounts for the entire data (i.e., RT distributions of correct and error trials) with three latent parameters: (a) the drift rate, a measure of the efficiency of information processing in the decision process, (b) the boundary separation, a measure of response caution that controls the speed-accuracy trade-off (this was fixed to .5), and (c) the non-decision time on-coding parametrization – to each dataset.

#### Recognition memory.

For our recognition memory data, we fit a Bayesian generalized linear multilevel model with a Bernoulli distribution and a probit link function. In its simplest form, Signal Detection Theory (SDT) models can be viewed as regressions with a probit link. To estimate the key SDT parameter $d^{\prime}$, we modeled participant responses (i.e., whether they responded “old” or “new”) as a function of item status (actual “old” vs. “new”) and blurring level, using a Bayesian hierarchical generalized linear model with a binomial distribution and probit link.

Traditional SDT analyses have been informative and efficient for binary accuracy data. However, these approaches lack the precision and power of mixed-effects models. For this reason, we adopt a Bayesian generalized linear mixed-effects modeling approach to SDT [see @zloteanu2024 for an excellent tutorial on Bayesian SDT models]

When fitting a GLMM with a probit link, model estimates map directly onto SDT parameters. Specifically, the interaction between item status (old vs. new) and the variable of interest (e.g., blurring level) on participants’ responses corresponds to $d^{\prime}$. The criterion ($c$) can also be derived from the model by examining the main effect of the variable of interest on response behavior.

### Results and Discussion

All models presented no divergences, and all chains mixed well and produced comparable estimates ( < 1.01).

#### 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 fixed effect estimates for accuracy hypotheses in Experiment 1a

| Hypothesis | Mean | SE | 90% CrI | ER | Post.Prob |
| --- | --- | --- | --- | --- | --- |
| High Blur - Clear < 0 | -1.00 | 0.18 | [-1.304, -0.701] | Inf | 1.00 |
| High Blur - Low Blur < 0 | -1.04 | 0.19 | [-1.358, -0.724] | Inf | 1.00 |
| Low Blur - Clear = 0 | 0.02 | 0.10 | [-0.186, 0.22] | 1.23 | 0.55 |
| Note. A 95% CrI is used for the equivalence test against 0 | | | | | |

Model estimates can be found in [Table 2](#tbl-acc1a). Clear words were better identified ( = 0.985) compared to high blur words ( = 0.962), = -0.997, 90% Cr.I[-1.304, -0.701], ER = Inf. Low blurred words were better identified ( = 0.986) than high blurred words, = -1.036, 90% Cr.I[-1.358, -0.724], ER = Inf. However, the evidence was weak for there being no significant difference in the identification accuracy between clear and low blurred words, = 0.017, 90% Cr.I[-0.186, 0.22], ER = 1.225.

#### 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 fixed effect estimates for Ex-Gaussian parameter hypotheses in Experiment 1a

| Hypothesis | Parameter | Mean | SE | 90% CrI | ER | Post.Prob |
| --- | --- | --- | --- | --- | --- | --- |
| High Blur - Clear > 0 | mu | 0.21 | 0.01 | [0.198, 0.225] | Inf | 1.00 |
| High Blur - Low Blur > 0 | mu | 0.20 | 0.01 | [0.182, 0.209] | Inf | 1.00 |
| Low Blur - Clear > 0 | mu | 0.02 | 0.00 | [0.01, 0.022] | Inf | 1.00 |
| High Blur - Clear > 0 | sigma | 0.15 | 0.08 | [0.023, 0.275] | 33.93 | 0.97 |
| High Blur - Low Blur > 0 | sigma | 0.12 | 0.08 | [-0.016, 0.242] | 12.15 | 0.92 |
| Low Blur - Clear = 0 | sigma | 0.04 | 0.06 | [-0.072, 0.144] | 202.42 | 0.99 |
| High Blur - Clear > 0 | beta | 0.42 | 0.03 | [0.365, 0.47] | Inf | 1.00 |
| High Blur - Low Blur > 0 | beta | 0.42 | 0.03 | [0.367, 0.472] | Inf | 1.00 |
| Low Blur - Clear = 0 | beta | -0.00 | 0.02 | [-0.049, 0.044] | 578.07 | 1.00 |
| Note. A 95% CrI is used for the equivalence test against 0 | | | | | | |

A visualization of how blurring affected processing during word recognition can be seen in the quantile and delta plots in [Figure 2](#fig-deltaquant). Beginning with the μ parameter, there was greater shifting for high blurred words compared to clear words, *b* = 0.212, 90% Cr.I [0.198, 0.225], ER = Inf, and low blur words, *b* = 0.196, 90% Cr.I [0.182, 0.209], ER = Inf.

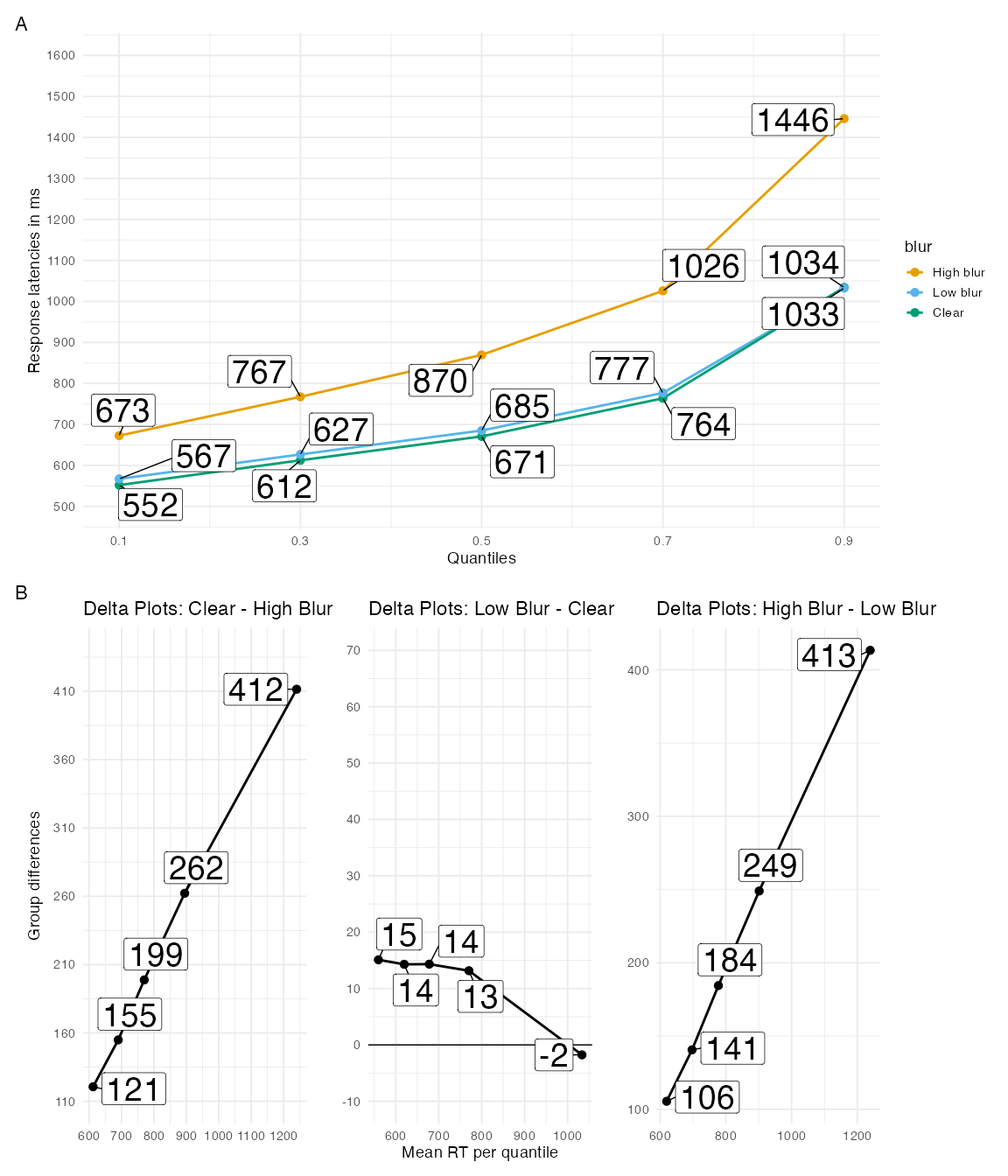
Analyses of the σ and τ parameters yielded a similar pattern. Variance was higher for high blurred words compared to clear words, *b* = 0.151, 90% Cr.I [0.023, 0.275], ER = 33.934, and low blurred words, *b* = 0.115, 90% Cr.I [-0.016, 0.242], ER = 12.147.

Finally, there was greater skewing for high blurred words compared to clear words, *b* = 0.417, 90% Cr.I [0.365, 0.47], ER = , and low blurred words, *b* = 0.42, 90% Cr.I [0.367, 0.472], ER = .

Low blurred words compared to clear words only differed on the μ parameter, *b* = 0.016, 90% Cr.I [0.01, 0.022], ER = Inf, with greater shifting for low blurred words. For and , the 95% Cr.I crossed zero and the ER was greater than 100.

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: Diffusion Model.

Table 4

Summary of posterior fixed effect estimates for DDM parameter hypotheses in Experiment 1a

| Hypothesis | Parameter | Mean | SE | 90% CrI | ER | Post.Prob |
| --- | --- | --- | --- | --- | --- | --- |
| High Blur - Clear < 0 | v | -0.87 | 0.07 | [-0.976, -0.758] | Inf | 1.00 |
| High Blur - Low Blur < 0 | v | -0.87 | 0.07 | [-0.986, -0.755] | Inf | 1.00 |
| Low Blur - Clear = 0 | v | 0.00 | 0.05 | [-0.091, 0.096] | 30.15 | 0.97 |
| High Blur - Clear > 0 | T\_er | 0.11 | 0.00 | [0.099, 0.113] | Inf | 1.00 |
| High Blur - Low Blur > 0 | T\_er | 0.09 | 0.01 | [0.085, 0.101] | Inf | 1.00 |
| Low Blur - Clear = 0 | T\_er | 0.01 | 0.00 | [0.008, 0.018] | Inf | 1.00 |
| Note. A 95% CrI is used for the equivalence test against 0 | | | | | | |

A summary of the diffusion model results can be found in [Table 4](#tbl-diffexpt1a). High-blurred words had a lower drift rate than clear words, *b* = -0.866, 90% Cr.I [-0.976, -0.758], ER = Inf, and low-blurred words, *b* = -0.869, 90% Cr.I [-0.986, -0.755], ER = Inf. There was no difference in drift rate between low-blurred words and clear words, *b* = 0.003, 90% Cr.I [-0.091, 0.096], ER = 30.146.

Non-decision time was higher for high-blurred words compared to clear words, *b* = 0.106, 90% Cr.I [0.099, 0.113], ER = Inf, and low-blurred words, *b* = 0.093, 90% Cr.I [0.085, 0.101], ER = Inf. Low-blurred words had a higher non-decision time than clear words, *b* = 0.013, 90% Cr.I [0.008, 0.018], ER = Inf.

#### Recognition Memory.

##### Sensitivity.

[Figure 3](#fig-dprimeexp1a) highlights High-blurred words were better remembered than clear words, = 0.128, 90% Cr.I [0.059, 0.196], ER = 694.652, and low-blurred words, = 0.124, 90% Cr.I [0.057, 0.192], ER = 665.667. There was no difference in sensitivity between clear words and low-blurred words, = 0.018, 90% Cr.I [-0.061, 0.099], ER = 2.017.

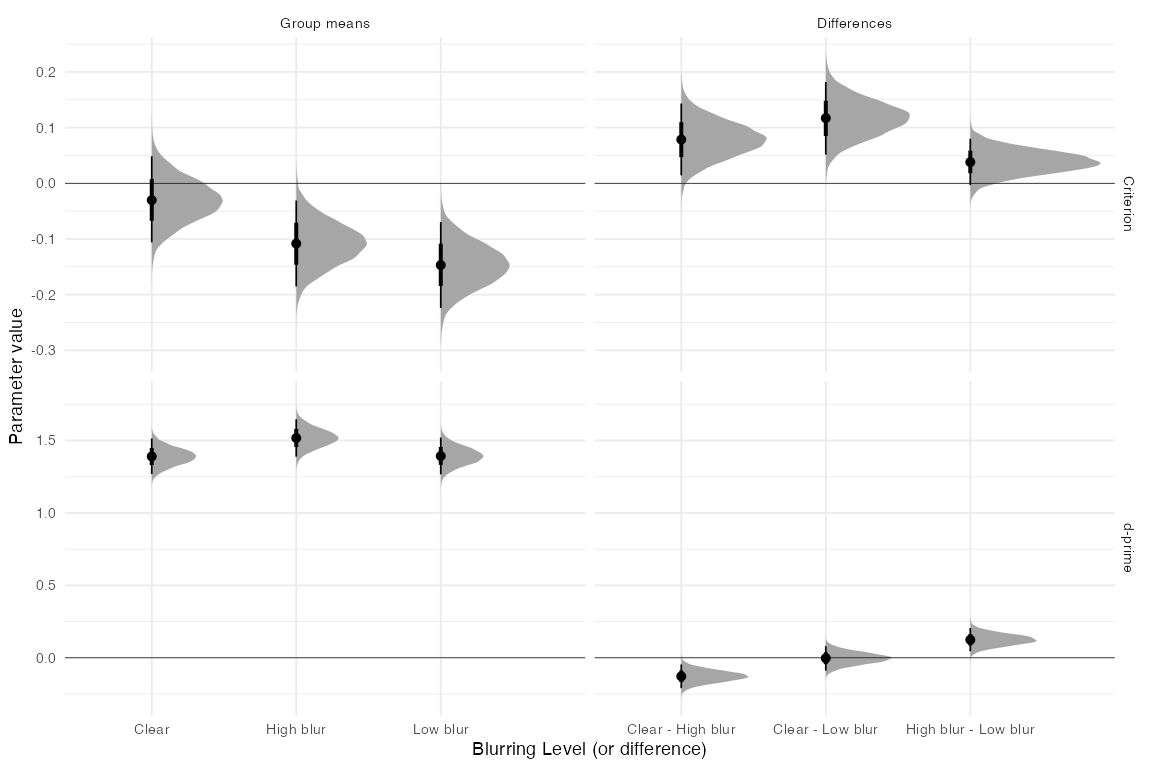
##### Exploratory.

###### Bias.

Low blurred words had a bias towards more “old” responses compared to clear words, = 0.119, 90% Cr.I[0.061, 0.177], ER = 0, and high blurred words, = -0.038, 90% Cr.I[-0.073, -0.003], ER = 26.444. There was no difference in bias between high blurred words and clear words, = 0.079, 95% Cr.I[0.014, 0.143], ER = 0.282.

Figure 3

Estimated posterior distributions for d-prime and criterion, and differences, with 66% (thick) and 95% CIs (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 (τ) compared to clear and low blurred words. However, low blurred words did not differ compared to clear words on or . 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 and a later more analytic, component evinced by a lower than clear or low blurred words. On the other hand, low blurred words only affected .

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.

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 as Experiment 1a (*N* = 216). All participants were recruited through the university subject pool at Rutgers University (SONA). We used a similar exclusion criteria to Experiment 1a. Because of this, we oversampled we randomly chose 36 participants from each list to reach our target sample size.

#### 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.

We fit the same statistical models as Experiment 1a.

### 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 fixed effect estimates for accuracy hypotheses in Experiment 1b

| Hypothesis | Mean | SE | 90% CrI | ER | Post.Prob |
| --- | --- | --- | --- | --- | --- |
| High Blur - Clear < 0 | -1.11 | 0.21 | [-1.454, -0.783] | Inf | 1.00 |
| High Blur - Low Blur < 0 | -1.07 | 0.19 | [-1.384, -0.769] | Inf | 1.00 |
| Low Blur - Clear = 0 | 0.07 | 0.11 | [-0.149, 0.285] | 0.91 | 0.47 |
| Note. A 95% CrI is used for the equivalence test against 0 | | | | | |

A summary of posterior estimates are located in [Table 5](#tbl-acc1b). Clear words were better identified ( = .987) compared to high blur words ( = .962), = -1.111, 95% Cr.I[-1.454, -0.783], ER = . Low blurred words were better identified ( = .987) than high blurred words, = -1.072, 95% Cr.I[-1.384, -0.769], ER = . However, the evidence was weak for there being no significant difference in the identification accuracy between clear and low blurred words, b = 0.069, 95% Cr.I[-0.149, 0.285], ER = 0.905.

#### 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 fixed effect estimates for Ex-Guassian parmater hypotheses in Experiment 1b

| Hypothesis | Parameter | Mean | SE | 90% CrI | ER | Post.Prob |
| --- | --- | --- | --- | --- | --- | --- |
| High Blur - Clear > 0 | mu | 0.21 | 0.01 | [0.194, 0.222] | Inf | 1.00 |
| High Blur - Low Blur > 0 | mu | 0.19 | 0.01 | [0.18, 0.209] | Inf | 1.00 |
| Low Blur - Clear > 0 | mu | 0.01 | 0.00 | [0.008, 0.019] | Inf | 1.00 |
| High Blur - Clear > 0 | sigma | 0.27 | 0.07 | [0.149, 0.393] | 3,199.00 | 1.00 |
| High Blur - Low Blur > 0 | sigma | 0.40 | 0.08 | [0.263, 0.535] | Inf | 1.00 |
| Low Blur - Clear = 0 | sigma | -0.13 | 0.06 | [-0.25, -0.009] | 27.87 | 0.96 |
| High Blur - Clear > 0 | beta | 0.37 | 0.03 | [0.317, 0.419] | Inf | 1.00 |
| High Blur - Low Blur > 0 | beta | 0.35 | 0.03 | [0.296, 0.401] | Inf | 1.00 |
| Low Blur - Clear = 0 | beta | 0.02 | 0.02 | [-0.028, 0.066] | 437.49 | 1.00 |
| Note. A 95% CrI is used for the equivalence test against 0 | | | | | | |

A visualization of how blurring affected processing during word recognition can be seen in the quantile and delta plots in [Figure 4](#fig-deltaquant1b). Beginning with the μ parameter, there was greater shifting for high blurred words compared to clear words, *b* = 0.208, 90% Cr.I [0.194, 0.222], ER = Inf, and low blur words, *b* = 0.194, 90% Cr.I [0.18, 0.209], ER = Inf.

Analyses of the σ and τ parameters yielded a similar pattern. Variance was higher for high blurred words compared to clear words, *b* = 0.273, 90% Cr.I [0.149, 0.393], ER = 3199, and low blurred words, *b* = 0.399, 90% Cr.I [0.263, 0.535], ER = Inf.

Finally, there was greater skewing for high blurred words compared to clear words, *b* = 0.368, 90% Cr.I [0.317, 0.419], ER = , and low blurred words, *b* = 0.349, 90% Cr.I [0.296, 0.401], ER = .

Low blurred words compared to clear words only differed on the μ parameter, *b* = 0.014, 90% Cr.I [0.008, 0.019], ER = Inf, with greater shifting for low blurred words. For and , the 95% Cr.I crossed zero and the ER was greater than 100.

#### DDM.

Table 7

Summary of posterior fixed effect estimates for DDM parameter hypotheses in Experiment 1b

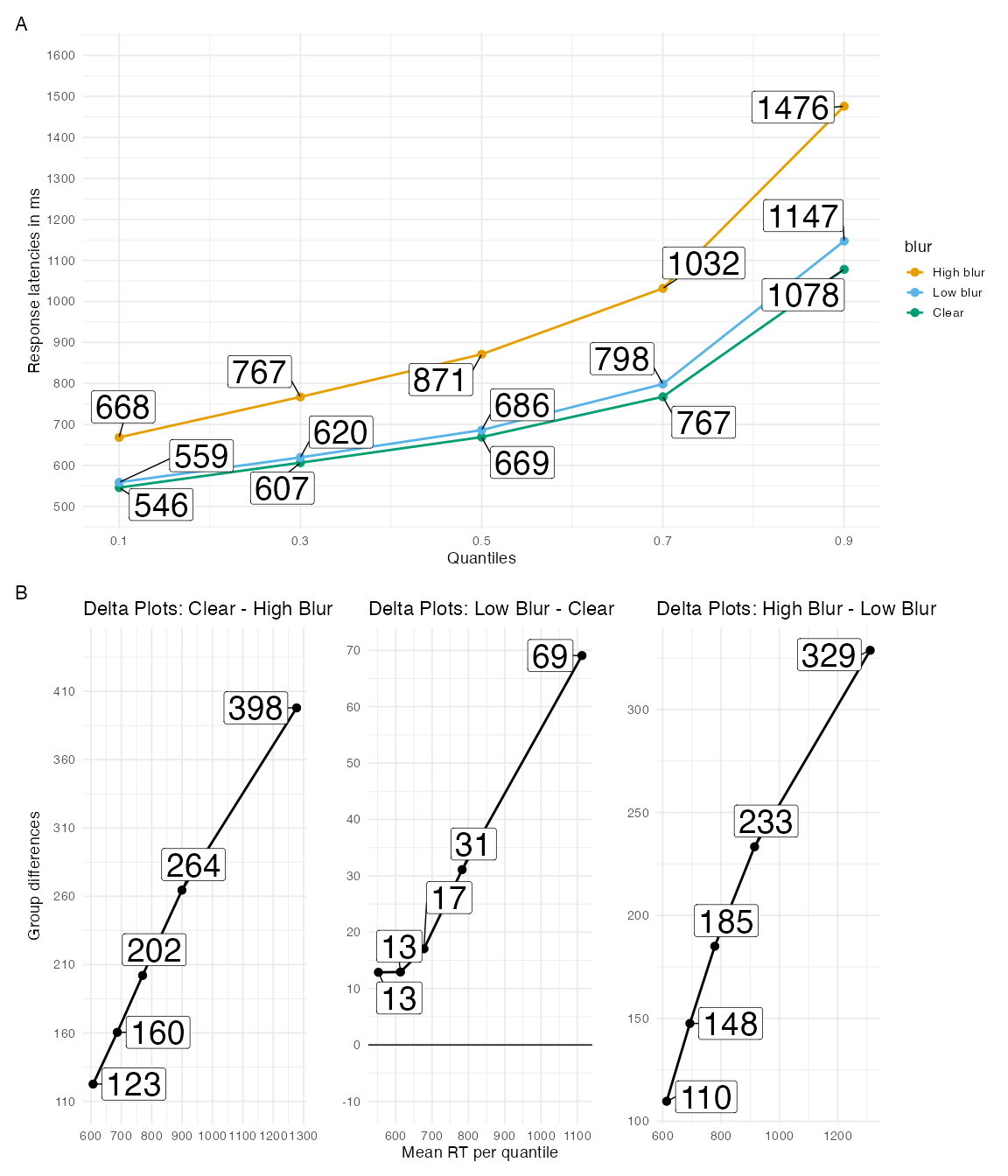
| Hypothesis | Parameter | Mean | SE | 90% CrI | ER | Post.Prob |
| --- | --- | --- | --- | --- | --- | --- |
| High Blur - Clear > 0 | v | -0.91 | 0.07 | [-1.016, -0.799] | Inf | 1.00 |
| High Blur - Low Blur > 0 | v | -0.88 | 0.07 | [-1.005, -0.766] | Inf | 1.00 |
| Low Blur - Clear = 0 | v | -0.02 | 0.06 | [-0.13, 0.086] | 23.42 | 0.96 |
| High Blur - Clear > 0 | T\_er | 0.11 | 0.01 | [0.1, 0.115] | Inf | 1.00 |
| High Blur - Low Blur > 0 | T\_er | 0.09 | 0.01 | [0.085, 0.102] | Inf | 1.00 |
| Low Blur - Clear = 0 | T\_er | 0.01 | 0.00 | [0.009, 0.02] | Inf | 1.00 |
| Note. A 95% CrI is used for the equivalence test against 0 | | | | | | |

A summary of the diffusion model results can be found in [Table 7](#tbl-diffexpt1b). High-blurred words had a lower drift rate than clear words, *b* = -0.908, 90% Cr.I [-1.016, -0.799], ER = Inf, and low-blurred words, *b* = -0.884, 90% Cr.I [-1.005, -0.766], ER = Inf. There was no difference in drift rate between low-blurred words and clear words, *b* = -0.023, 90% Cr.I [-0.13, 0.086], ER = 23.417.

Non-decision time was higher for high-blurred words compared to clear words, *b* = 0.107, 90% Cr.I [0.1, 0.115], ER = Inf, and low-blurred words, *b* = 0.093, 90% Cr.I [0.085, 0.102], ER = Inf. Low-blurred words had a higher non-decision time than clear words, *b* = 0.014, 90% Cr.I [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.



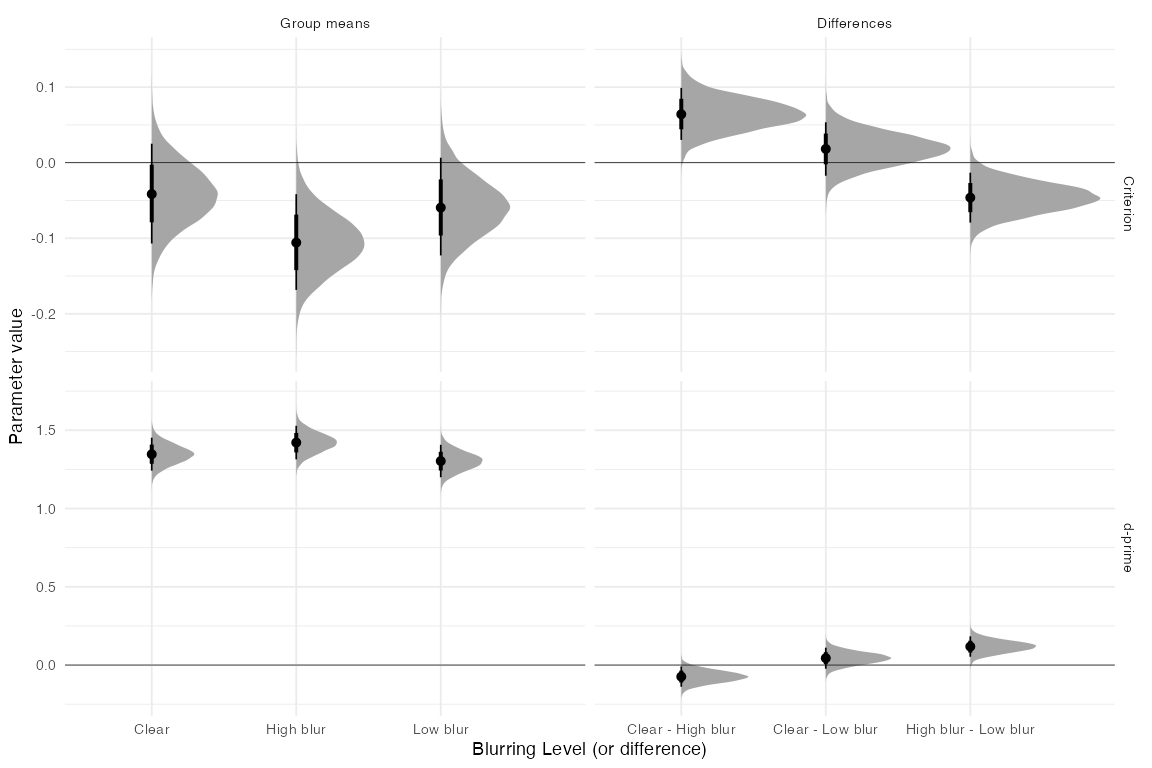
#### Recognition Memory.

##### Sensitivity.

[Figure 5](#fig-dprimeexp1b) highlights High-blurred words were better remembered than clear words, = 0.074, 90% Cr.I [0.009, 0.139], ER = 32.684, and low-blurred words, = 0.118, 90% Cr.I [0.053, 0.183], ER = 550.724. There was no difference in sensitivity between clear words and low-blurred words, = -0.04, 90% Cr.I [-0.116, 0.038], ER = 1.793.

Figure 5

Estimated posterior distributions for d-prime and criterion, and differences, with 66% (thick) and 95% CIs (thin lines)



## Experiment 2

In Experiments 1a and 1b, we employed mathematical and computational techniques to study the impact of blurring on encoding and recognition memory. High blurred words influenced both early and late stages evident by increased distributional shifting and skewing, lower , and higher . Low blurred words (compared to clear words), on the other hand, only impacted an early phase, indicated by increased distributional shifting and higher . In terms of recognition memory, sensitivity was higher for high blurred words than clear and low blurred words. This implies two facets to the disfluency effect: an early, automatic/non-analytic component, and a subsequent, analytic component. The locus of this later component remains ambiguous. The mnemonic benefit for recognizing high blurred words might arise from enhanced top-down (lexical or semantic processing) processing which offsets the challenge of reading blurred text. Alternatively, the benefit might stem from increased attention or control processes operating along side processes needed to recognize the word.

One way to test more directly the accounts of perceptual disfluency would be to identify an aspect of higher level information that plays a role in word perception and to examine its impact on the disfluency effect. Several models of word recognition assume the speed and ease of word identification varies as a function of word frequency (Coltheart et al., 2001; McClelland & Rumelhart, 1981). Looking at RT distributions for word frequency effects has shown both an early and late locus, showing larger distribution shifts and more skewing for low frequency words (Andrews & Heathcote, 2001; Balota & Spieler, 1999; Plourde & Besner, 1997; Staub, 2010; Yap & Balota, 2007; but see Gomez & Perea, 2014 for a DDM account). In regards to memory, low frequency words are generally better recognized than high frequency word in recognition memory(Glanzer & Adams, 1985). The recognition advantage for less frequent words has been ascribed to the additional cognitive effort or attenton required to process them (Diana & Reder, 2006; but see Pazzaglia et al., 2014). This has been called the elevated attention hypothesis (Malmberg & Nelson, 2003).

In tasks like semantic categorization and pronunciation, the interaction between word frequency and stimulus degradation (in this case, perceptual disfluency) leads to an over-additive effect (Yap & Balota, 2007). By the logic of additive factors (Sternberg, 1969), if factors do interact, they are believed to be associated with similar processing stages. The interplay between perceptual disfluency and word frequency originates from perceptual disfluency hindering initial processing and word identification. This leads to a magnification of the word frequency effect. Perceptual disfluencies like handwritten cursive (Barnhart & Goldinger, 2010; Perea et al., 2016) and research on letter rotation in words (Fernández-López et al., 2022) have shown a magnification of the word frequency effect.

In Experiment 2, we manipulated word frequency (high vs. low frequency words) and perceptual blur (i.e., clear, low, and high) within a semantic categorization task. Mirroring Experiments 1a and 1b, the categorization task preceded a surprise memory recognition test. Our goal here was to evaluate the compensatory processing and stage-specific theories as both theories offer predictions about memory performance.

In Experiment 2, we opted to forgo using the DDM. Instead we focus on ex-Gaussian parameters during encoding. Both the compensatory processing and stage specific accounts predict an interaction of word frequency and blurring on and where the word frequency effect is largest for high blurred words. Where both accounts differ is in terms of memory performance.

The compensatory processing account predicts items receiving the most top-down processing during encoding should be better remembered. On a recognition test, this account would predict items low in frequency should show a disfluency effect due to low frequency items receiving more top-down processing during encoding.

The stage specific account, on the other hand, is influenced by extra attentional or control processes taking place during and after word recognition. Here, memory performance relies not only on the type of processing during encoding but also on limited-capacity resources such as cognitive control.

Low frequency words, like high blurred words, routinely attract attention during encoding [see evidence from pupillometry; Kuchinke et al. (2007)]. Thus the benefits of perceptual disfluency may be less effective for these items. High frequency items on the other hand should be more likely to benefit from a manipulation that enhances attention to, and encoding of, the item. The presence of low frequency words could make perceptual disfluency redundant in this regard. In this regard, we might not see a perceptual disfluency effect for low frequency items. Ptok et al. (2019) argued that the memory benefits for conflict encoding phenomena are limited to tasks that are relatively fluent, automatic, and encoding poor. Any additional demands placed on the participants imposed by the task could reduce the disfluency effect. As evidence for this, Ptok et al. (2020) showed manipulating endogenous attention (by using a chinrest) eliminated the memory benefit from semantic interference during encoding. Similarly, Geller and Peterson (2021a) manipulated attention through a test expectancy manipulation (i.e., being told about an upcoming memory test or not) which presumably oriented participants to study all words for the upcoming memory test, regardless of disfluency. This resulted in the elimination of the disfluency effect. Lastly, Westerman and Greene (1997) (Experiment 3), showed, with a masking manipulation, that changing encoding instructions from reading the target word (more automatic) to spelling the target word eliminated the perceptual disfluency effect. In Experiment 2, we examine how word frequency interacts with perceptual disfluency and how this affects memory.

### Method

### Transparenency 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

Non-animal and animal words were adapted from Fernández-López et al. (2022). To make the experiment more feasible for online participants and to evenly split our conditions, we winnowed their non-animal words and presented 90 (1/2 HF and 1/2 LF) non-animal words and 45 animal words during study. This kept the 2:1 ratio used in previous experiments (e.g. (Fernández-López et al., 2022; Perea et al., 2018). At test, 90 non-animal words we did not use during the semantic categorization task were used as new words for the recognition test. We created six counterbalanced lists to ensure that each non-animal word was presented as both old and new and as clear, high blurred, and low blurred across participants. Similar to non-words from Experiments 1a and 1b, we excluded animal words from analysis.

The number of letters of the animal words (M = 5.3; range: 3-9) was similar to that of the non-animal words (high-frequency words: M = 5.3, range: 3-8; low-frequency words: M = 5.3, range: 3-9). The animal words had an ample range of word-frequency 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., animal/not animal). You can view the task here: https://run.pavlovia.org/Jgeller112/hf\_lf\_sem\_1.

### Results

#### Accuracy.

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

Table 8

Summary of posterior fixed effect for accuracy hypothses in Experiment 2

| Hypothesis | Mean | SE | 90% CrI | ER | Post.Prob |
| --- | --- | --- | --- | --- | --- |
| High Blur - Clear < 0 | -1.13 | 0.27 | [-1.59, -0.689] | Inf | 1.00 |
| High Blur-Low Blur < 0 | -0.76 | 0.24 | [-1.154, -0.374] | 1,999.00 | 1.00 |
| Low Blur - Clear = 0 | -0.07 | 0.12 | [-0.273, 0.13] | 2.64 | 0.72 |
| Low Frequency - High Frequency | 0.13 | 0.20 | [-0.235, 0.528] | 1.17 | 0.54 |
| (High Blur-Clear) - (Low Frequency-High Frequency) < 0 | 0.02 | 0.24 | [-0.462, 0.514] | 1.16 | 0.54 |
| (High Blur-Low Blur) - (Low Frequency-High Frequency) < 0 | -0.00 | 0.23 | [-0.466, 0.463] | 1.24 | 0.55 |
| (Low Blur-Clear) - (Low Frequency-High Frequency) = 0 | 0.09 | 0.21 | [-0.337, 0.508] | 0.60 | 0.37 |
| Note. A 95% CrI is used for the equivalence test against 0 | | | | | |

The model summary can be found in [Table 8](#tbl-acc2). Clear words were better identified than high-blur words ( = .963), = -1.128, 90% Cr.I [-1.59, -0.689], ER = Inf. Low-blur words were also better identified than high-blur words, = -0.764, 90% Cr.I [-1.154, -0.374], ER = 1999. There was weak evidence for a difference between clear and low-blur words, = -0.072, 95% Cr.I [-0.273, 0.13], ER = 2.638. There was no frequency effect, = 0.132, 95% Cr.I [-0.235, 0.528], ER = 1.166, although the evidence for the absence of a difference was weak.

There were no interactions between blurring and word frequency—all Cr.I included 0; however, evidence for this lack of difference was ambiguous (ER ≈ 1).

#### RTs: Ex-Gaussian.

Table 9

Frequency effect by blurring level

| blur | High | Low | Frequency Effect |
| --- | --- | --- | --- |
| C | 705.30 | 723.09 | 18.00 |
| HB | 960.98 | 1,016.44 | 55.00 |
| LB | 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 fixed effect estimates for Posterior Ex-Guassian parameter hypothses in Experiment 2

| Hypothesis | Parameter | Mean | SE | 90% CrI | ER | Post.Prob |
| --- | --- | --- | --- | --- | --- | --- |
| High Blur - Clear > 0 | mu | 0.27 | 0.01 | [0.262, 0.286] | Inf | 1.00 |
| High Blur-Low Blur > 0 | mu | 0.27 | 0.01 | [0.253, 0.277] | Inf | 1.00 |
| Low Blur - Clear = 0 | mu | 0.00 | 0.01 | [-0.01, 0.015] | 1,500.04 | 1.00 |
| Low Frequency - High Frequency | mu | -0.03 | 0.01 | [-0.035, -0.016] | 0.00 | 0.00 |
| (High Blur-Clear) - (Low Frequency-High Frequency) < 0 | mu | -0.03 | 0.01 | [-0.05, -0.009] | 95.97 | 0.99 |
| (High Blur-Low Blur) - (Low Frequency-High Frequency) < 0 | mu | -0.03 | 0.01 | [-0.052, -0.012] | 180.82 | 0.99 |
| (Low Blur-Clear) - (Low Frequency-High Frequency) = 0 | mu | -0.02 | 0.01 | [-0.036, 0] | 170.38 | 0.99 |
| High Blur - Clear > 0 | sigma | 0.69 | 0.05 | [0.608, 0.769] | Inf | 1.00 |
| High Blur-Low Blur > 0 | sigma | 0.71 | 0.05 | [0.623, 0.791] | Inf | 1.00 |
| Low Blur - Clear = 0 | sigma | 0.01 | 0.04 | [-0.056, 0.081] |  |  |
| Low Frequency - High Frequency | sigma | -0.08 | 0.04 | [-0.136, -0.016] | 50.28 | 0.98 |
| (High Blur-Clear) - (Low Frequency-High Frequency) < 0 | sigma | 0.07 | 0.08 | [-0.06, 0.21] | 0.22 | 0.18 |
| (High Blur-Low Blur) - (Low Frequency-High Frequency) < 0 | sigma | 0.12 | 0.08 | [-0.018, 0.256] | 12.05 | 0.92 |
| (Low Blur-Clear) - (Low Frequency-High Frequency) = 0 | sigma | 0.01 | 0.06 | [-0.089, 0.121] | 1.46 | 0.59 |
| High Blur - Clear > 0 | beta | 0.53 | 0.03 | [0.487, 0.575] | Inf | 1.00 |
| High Blur-Low Blur > 0 | beta | 0.55 | 0.03 | [0.504, 0.593] | Inf | 1.00 |
| Low Blur - Clear = 0 | beta | -0.04 | 0.04 | [-0.11, 0.034] |  |  |
| Low Frequency - High Frequency | beta | -0.03 | 0.02 | [-0.062, 0.005] | 0.09 | 0.08 |
| (High Blur-Clear) - (Low Frequency-High Frequency) < 0 | beta | -0.08 | 0.04 | [-0.146, -0.008] | 30.81 | 0.97 |
| (High Blur-Low Blur) - (Low Frequency-High Frequency) < 0 | beta | -0.14 | 0.04 | [-0.213, -0.073] | 2,665.67 | 1.00 |
| (Low Blur-Clear) - (Low Frequency-High Frequency) < 0 | beta | -0.10 | 0.05 | [-0.184, -0.022] | 57.82 | 0.98 |
| Note. A 95% CrI is used for the equivalence test against 0 | | | | | | |

[Table 10](#tbl-expt2summary) provides a model summary. [Figure 6](#fig-quantiledeltaexp2) visualizes RTs as quantile and delta plots highlighting how blurring and word frequency affected processing during word recognition.

##### Mu.

Looking at first, high blurred words had greater shifting than clear words, = 0.274, 90% Cr.I [0.262, 0.286], ER = , and low blurred words, = 0.265, 90% Cr.I [0.253, 0.277], ER = .

There was no difference in shifting between low blurred words and clear words, = 0.002, 90% Cr.I [-0.01, 0.015], ER = 1500.037.

For word frequency, there was no evidence for greater shifting for low frequency compared to high frequency words, = -0.026, 90% Cr.I [-0.035, -0.016], ER = 0.

Regarding the interaction between frequency and blurring, there was an amplified word frequency effect for high blurred words compared to clear words, = -0.029, 90% Cr.I [-0.05, -0.009], ER = 95.97, and low blurred words, = -0.032, 95% Cr.I [-0.052, -0.012], ER = 180.818, with greater shifting for low frequency words.

There was strong evidence that there was no amplification of the word frequency effect for the low blurred vs. clear comparison, = -0.017, 95% Cr.I [-0.036, 0], ER = 170.385.

##### Sigma.

High blurred words had higher compared to clear words, = 0.688, 90% Cr.I [0.608, 0.769], ER = , and low blurred words, = 0.707, 95% Cr.I [0.623, 0.791], ER = .

There was weak evidence that low blurred words had greater variance than clear words, = 0.012, 90% Cr.I [-0.056, 0.081], ER = NA.

Low frequency words showed greater variance than high frequency words, = -0.076, 90% Cr.I [-0.136, -0.016], ER = 50.282.

There were no significant interactions—all Cr.I included 0.

##### Beta.

High blurred words showed greater skewing than clear words, = 0.531, 90% Cr.I [0.487, 0.575], ER = , and low blurred words, = 0.549, 90% Cr.I [0.504, 0.593], ER = . There was strong evidence for no skewing difference between low blurred words and clear words, = -0.038, 95% Cr.I [-0.11, 0.034], ER = NA.

Low frequency words did not show greater skewing than high frequency words, = -0.029, 95% Cr.I [-0.062, 0.005], ER = 0.091.

However, the word frequency effect was magnified for high blurred words compared to clear, = -0.077, 95% Cr.I [-0.146, -0.008], ER = 30.809, and low blurred words, = -0.143, 90% Cr.I [-0.213, -0.073], ER = 2665.667, with greater skewing for low frequency words than high frequency words.

There was also an interaction for the low blurred vs. clear words comparison, = -0.098, 95% Cr.I [-0.184, -0.022], ER = 57.824. However, the word frequency effect was reversed here, with low blurred-high frequency words having greater skewing than low blurred-low frequency words.

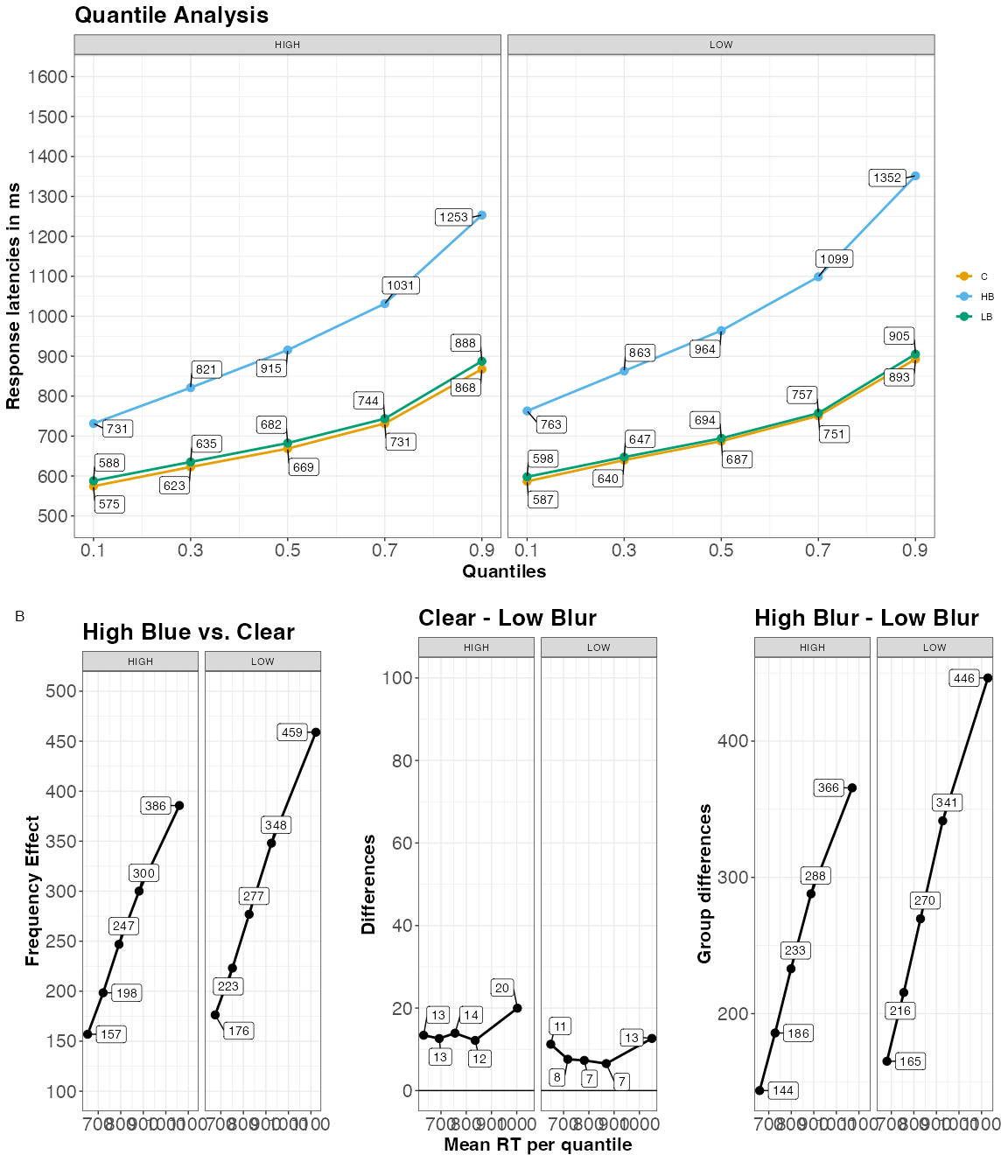
Table 11

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

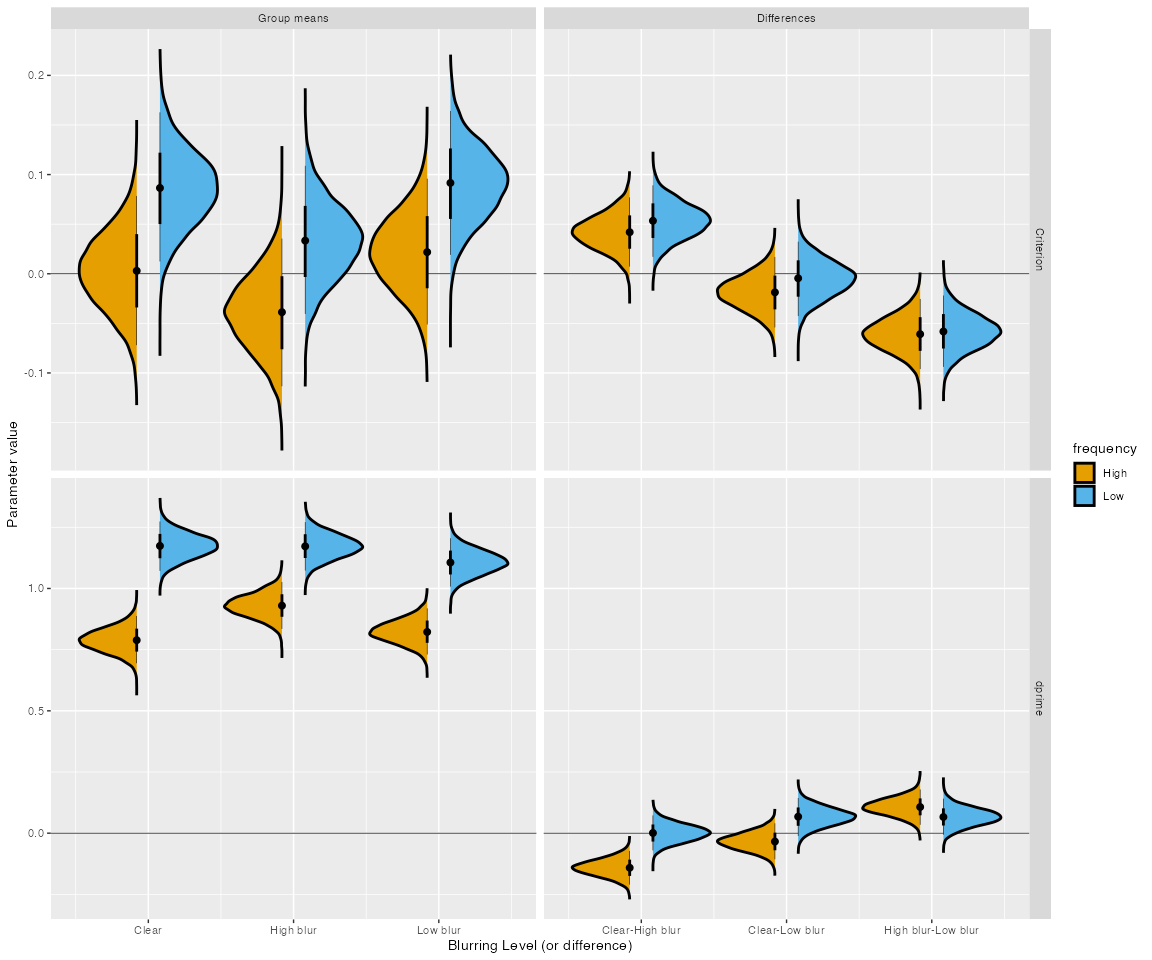
Group RT distributions in the blurring and word frequency manipulations in word stimuli. A. Quantile plots with each point represents the average RT quantiles (.1, .3, .5, .7,and .9) in each condition. B. Delta plots obatined by computing the quantiles for each participant and subsequently averaging the obtained values for each quantile over the participants and subtracting the values from each condition.



#### Recognition Memory.

Figure 7

Estimated posterior distributions for d-prime and criterion, and differences, with 66% (thick) and 95% CIs (thin lines)



#### Sensitivity.

Similar to Experiments 1a and 1a, there was better memory recognition for high blurred words compared to clear words (see [Figure 7](#fig-m2-emmeans)), $\beta$ = 0.0702149, 90% Cr.I[0.0282945, 0.1117977], ER = 409.2564103, and low blurred words, $b$ = 0.0868257, 90% Cr.I[0.0435005, 0.1309682], ER = 1999. There was no recognition memory difference between clear and low blur words, $\beta$ = -0.016903, 95% Cr.I[-0.0678498, 0.0340915], ER = 4.433816. Low frequency words were better recognized than high frequency words, = -0.3031522, 90% Cr.I[-0.392499, -0.2108739], ER = . There was strong evidence for an interaction between high blurred words (vs. clear) and frequency, $\beta$ = 0.1422165, 95% Cr.I[0.0608359, 0.2262012], ER = 483.8484848, with better memory for high frequency-high blurred words, $\beta$ = -0.14, 90% Cr.I[-0.21, -0.07]. There was some evidence of an interaction between blurring and frequency for high blurred words vs. low blurred words, $\beta$ = 0.040769, 90% Cr.I[-0.0438705, 0.1264736], ER = 3.6511628. Looking at this interaction post hoc, high frequency-high blurred words were better recognized than low-frequency-blurred words. There was also an interaction between frequency and low blurred vs. clear words, $\beta$= 0.0987905, 95% Cr.I[0.0142211, 0.1847873], ER = 4.433816. This interaction arose from low-frequency clear words being better remembered than low-frequency low blurred words.

### Discussion

Experiment 2 explored the source of the late stage processing underlying the disfluency effect. Using a word frequency manipulation coupled with a semantic categorization task, we discovered non-additive effects of frequency and blurring on response time distributions. Specifically, the word frequency effect was magnified for high blurred words (compared to clear and low blurred words). We observed this on and parameters, indicating that when stimuli are degraded, word frequency influences early and late stages of processing during word recognition. This pattern has also been found with other disfluent stimuli, such as hard-to-read handwritten cursive words (Barnhart & Goldinger, 2010; Perea et al., 2016; Vergara-Martínez et al., 2021).

Looking at quantile and delta plots, we see there was a robust word-frequency advantage that increased in the higher quantiles for clear and low-blurred words at the 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles, respectively). Critically, for the high blurred condition, the word-frequency effect not only changed across quantiles with a steeper slope, but it was also larger in the first quantiles (see [Table 11](#tbl-freq)). This finding suggests that word-frequency already taps onto an encoding stage of processing when the stimuli appear in hard-to-read format like high blurred words. This replicates earlier research with easy-to-read and hard-to-read handwriting (Perea et al., 2016; Vergara-Martínez et al., 2021).

Critical here is how the observed non-additivity impacts memory. Replicating Experiments 1a and 1b, we report an overall memory benefit for high blurred words. Additionally, we observed better recognition memory for low frequency words compared to high frequency words. Examining the interaction between blurring and frequency revealed a distinct pattern. We observed a disfluency effect for high frequency-high blurred words. However, low frequency words only benefited memory when they were unaltered (clear words showed a low frequency benefit). When words were blurred, the low ferquency memory benefit did not emerge. This pattern of findings helps us shed some light on the potential source of late stage processing in the disfluency effect.

## 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 disfluendcy 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 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 and lower . Conversely, low blurred words compared to clear words showed greater distributional shifting, but there was no difference in skewing. DDM fits showed higher , but had no effect on .

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. In fact, the effect seemed to be reversed for clear words and low blurred words. However, high frequency words did 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 perceptual 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. When there was minimal control demands (i.e., clear words), we did see better memory for 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.

More significantly, our current experiments underscore the benefits of using mathematical and computational models—such as the ex-Gaussian model and the drift diffusion model (DDM)—to examine perceptual disfluency during encoding. In Experiments 1a and 1b, both models converged on similar findings. Specifically, response time distributions were differentially 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 $T\_{\text{er}}$ and $\mu$), whereas highly blurred words impacted both early and later stages ($T\_{\text{er}}$, $\tau$, 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, 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 (i.e., Sans Forgetica typeface; Geller et al. (2020)) influenced DDM parameters during recognition and how they relate to confidence judgments. 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 , but not drift rate. They also looked at how parameters of the DDM were related to confidence judgments. Higher 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 Experiments 1a and 1b–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 (2021a) 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 (2021b), 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 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>

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. (2017a). *Brms: An r package for bayesian multilevel models using stan*. *80*. <https://doi.org/10.18637/jss.v080.i01>

Bürkner, P.-C. (2017b). *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>

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>

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

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. (2021a). 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., & Peterson, D. (2021b). 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>

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>

Kinoshita, S., Amos, A., & Norris, D. (2023). Diacritic priming in novice readers of diacritics. *Journal of Experimental Psychology: Human Perception and Performance*, *49*(3), 370–383. <https://doi.org/10.1037/xhp0001084>

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>

Morey, R. D., & Rouder, J. N. (2022). *BayesFactor: Computation of bayes factors for common designs*. <https://CRAN.R-project.org/package=BayesFactor>

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>

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2019). *How to capitalize on a priori contrasts in linear (mixed) models: A tutorial*. *110*. <https://doi.org/10.1016/j.jml.2019.104038>

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>

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>

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>