The Role of Processing Fluency in Source Memory and Metamemory

Shih-Ting (Tina) Huang

University College London

Thesis submitted for the degree of Doctor of Philosophy

September 2019

Declaration

I declare that this thesis has been composed by myself, that the work contained herein is my own except where explicitly stated otherwise by reference or acknowledgement, and that it has not been submitted, in whole or in part, for any other degree or professional qualification.

Chapters 2-5 are based on the paper by Huang, T. S.-T. & Shanks, D. R., (*under review*), “Examining the relationship between processing fluency and memory for source information”.

Chapters 5-6 are based on the paper by Yang, C., Huang, T. S.-T., & Shanks, D. R., (2018), “Perceptual fluency affects judgments of learning: The font size effect” in *Journal of Memory and Language*, *99*, 99-110.

[redacted]

Shih-Ting (Tina) Huang
2nd July, 2019

Acknowledgements

I would like to express my sincerest gratitude to my primary supervisor, Professor Shanks, for his insightful knowledge, unwavering patience, and unfailing encouragement throughout this journey. Without his thoughtful guidance and persistent trust in me, this work would not have been possible. I am most fortunate and honoured to have him as my supervisor.

Besides my primary supervisor, I would like to thank Dr Yang for the fruitful exchange of ideas cumulating in the design of Experiments 4-6, and for his collaboration and assistance with computer coding and data analysis in those experiments. I am also grateful to my secondary supervisor, Dr Otten, and to other colleagues of the lab for their constructive feedback during the various stages of this project. Further thanks go to fellow occupants of room 204C for their companionship and interesting discussions on research. Additional thanks go to my examiners and to colleagues of the lab for their thought-provoking comments on my work.

I would like to extend my enduring gratitude to my mentors, particularly Professor Graf and Professor Hall, who have inspired my interest in research, and helped me embark on this journey. I offer my special thanks to my friends for their enthusiasm and motivation, and to all those who have played a part in this long endeavour.

Last but not least, I would like to express my heartfelt appreciation to my family. It is beyond words to describe my utmost gratitude for their love, and support in every way.

Abstract

Processing fluency influences various judgements in memory and cognition such as fluency-based familiarity in tests of item recognition memory. However, less is known about the interplay between fluency and source information in recognition memory and metamemory phenomena. The present thesis investigated the relationship between perceptual fluency and the accuracy of source memory decisions (Experiments 1-3b), as well as the contribution of perceptual fluency to the font size effect (i.e., the tendency to rate larger font words as easier to remember than smaller font words, despite font size having no effect on retention performance) in judgements of learning (Experiments 4-6). Fluency was indexed via identification response times (RTs) derived from adapted versions of the continuous identification (CID) task, in which stimuli gradually clarified through progressive demasking. Identification RTs were faster in trials with correct retrieval of source information compared to trials for which source could not be accurately retrieved, and JOLs were indirectly increased by the faster identification RTs associated with a larger font size. These findings suggest that fluency is related both to source memory and metamemory judgements.

Impact Statement

The present thesis explores the extent to which processing fluency relates to source memory and metamemory – in other words, how the speed, ease, and accuracy of perceptually processing a piece of information might contribute to how accurately people remember contextual details associated with that information, or to how likely they think they would be to remember that information at a future test.

Experiments 1-3b suggest that greater fluency is related to more accurate source retrieval. This finding provides evidence against some dual-process models of memory which propose a complete distinction between implicit (e.g., fluency) and explicit (e.g., source retrieval) processes in memory, and supports single-system models of memory as well as versions of dual-process models which allow for implicit-explicit correlations. These experiments might also help inform follow-up investigations on ways of improving people’s memory for complex information such as eyewitness information or everyday episodic events.

Experiments 4-6 demonstrate that greater fluency, associated with information being presented in a larger font size, contributes to a tendency to rate information as more likely to be recalled later on, despite font size itself having no impact on actual recall performance. This suggests that fluency can directly contribute to people’s predictions of their future memory performance, contrary to metamemory theories which propose that people’s own beliefs play the dominant role in informing their predictions. These findings might help people develop more effective study strategies and decisions when learning new material, for example, whether or not to spend more time to study information printed in a difficult-to-read font.

Table of Contents

Index of Figures and Tables

Chapter 1: Introduction

Processing fluency – the ease, speed, and accuracy with which information can be processed – influences various judgements and behaviours in our day-to-day lives. In order to respond adaptively to the vast quantity of stimuli from the environment, processing fluency is often employed as a heuristic to evaluate and/or interact with a stimulus, as it is fast (Hertwig, Herzog, Schooler, & Reimer, 2008) and cognitively nondemanding (Jacoby & Brooks, 1984). The cognitive and social dimensions that are shown to be affected by processing fluency are wide-ranging: from judgements of truth (Begg, Anas, & Farinacci, 1992; Schwarz & Reber, 1999), to face-to-trait inferences (Zebrowitz & Montepare, 2008), aesthetic judgements (Reber, Schwarz, & Winkielman, 2004), the grading of handwritten exam answers (Greifeneder, Alt, Bottenberg, Seele, Zelt, & Wagner, 2010), and so on. Some of the earliest researched, yet currently most active, domains of fluency effects are in memory and learning (Alter & Oppenheimer, 2009; Jacoby & Dallas, 1981).

1.1 Fluency and Priming in Recognition memory

Repetition priming, a phenomenon of implicit memory in which the fluency of processing a stimulus is enhanced as a result of previous exposure, has been extensively studied in relation to explicit item recognition memory. Experiments on recognition memory often require participants to identify ‘old’ items from a study list and ‘new’ items which were not presented previously. When a test item is recognized as ‘old’, participants might simply *know* (K) that the item was previously encountered based on a sense of *familiarity* elicited by the item, or they might *remember* (R) experiencing the event of the item being presented based on their *recollection* of the episodic details associated with the item (Tulving, 1985). For instance, participants in Jacoby and Whitehouse's (1989) experiment showed a greater tendency to rate primed items as 'old' regardless of whether they had previously encountered them during study.

Numerous models have been proposed to explain the roles of familiarity and recollection in recognition memory. In dual-process models, familiarity and recollection are assumed to be two fundamentally distinct processes which contribute to recognition decisions and correspond respectively to the subjective experiences of knowing and remembering (e.g., Atkinson & Juola, 1973; Jacoby & Dallas, 1981; Mandler, Pearlstone, & Koopmans, 1969; Yonelinas, 1994). Despite some points of disagreement on areas such as the time course and neural substrates of familiarity and recollection, the general consensus amongst dual-process theorists is that familiarity is a faster process than recollection, and that the two processes operate independently at the time of retrieval (Yonelinas, 2002). Most dual-process accounts assume that familiarity is driven by a continuum of memory strength, whereas recollection arises from an all-or-none threshold process in which memory for recollective information occurs only in high-confidence instances of item retrieval, but not at all in other retrieval instances.

On the other hand, single-system models of recognition memory interpret R and K response decisions as being driven by a unidimensional continuum of memory strength rather than by two discrete memory systems (e.g., Hirshman & Henzler, 1998; Inoue & Bellezza, 1998; Wixted, 2007;

Wixted & Stretch, 2004). A signal detection process is assumed by single-system models to underlie the evaluation of memory strength. Since the judgement criterion is expected to be higher for R than for K, single-system models predict that an item receives an R response if its memory strength signal value is above the R criterion, whereas it receives a K response if its strength falls between the K and R criteria. Although dual-process models predict that any degree of recollection is only associated with the highest level of confidence, the single-system view is that some degree of recollection can be involved regardless of memory strength. Hence, familiarity can be interpreted as a weaker memory strength signal compared to recollection, but can nonetheless contribute to the recollective process.

Using a similar repetition priming paradigm as Jacoby and Whitehouse (1989), Rajaram (1993) investigated the impact of processing fluency on remember/know judgements and found an increase in the proportion of K responses for primed versus unprimed trials, with no significant effects on the proportion of R responses. This demonstrated that processing fluency may be attributed to past experience and contribute to the subjective feeling of familiarity with a stimulus. Although Rajaram’s finding appears to support the dual-process prediction that fluency manipulations selectively affect familiarity, its generalisability has been criticised on a range of methodological grounds (Brown & Bodner, 2011; Higham & Vokey, 2004; Tunney & Fernie, 2007).

1.2 Recollection and Source memory

Since definitions of recollection involve the remembering of specific qualitative details associated with an encoded item, source memory is commonly also tested in recognition memory experiments as a proxy for recollection. Broadly speaking, source information entails the origins and/or contextual attributes of a piece of information or of a focal stimulus item. For example, we might remember the name of a particular dish that we enjoyed, but we would also need to remember which restaurant we tried the dish in, in order to be able to have it again in the future. Source memory refers to memory for where, when, or how a piece of information was acquired. The Source Monitoring Framework (Johnson, Hashtroudi, & Lindsay, 1993) conceptualises source memory as involving a series of decision and evaluation processes, resulting in varying degrees of precision and accuracy. As an individual continuously processes information from his or her environment, memory traces are tacitly formed which can later be reactivated to guide source attributions. In contrast to source memory, item memory refers to memory for a focal stimulus or event, without encompassing its spatiotemporal context or associated features.

It is widely accepted that source memory tasks are more difficult than item memory tasks (e.g., Cabeza, Mangels, Nyberg, Habib, Houle, McIntosh, & Tulving, 1997; Davidson, McFarland, & Glisky, 2006). Several factors are known to impact source memory accuracy. Poor attention during encoding can lead to incomplete revival of source information and impair source memory to a greater extent than item recognition (Troyer et al., 1999). Source memory accuracy can also be compromised when sources are highly similar in perceptual, semantic, or conceptual qualities (Lindsay, Johnson, & Kwon, 1991), and has been shown to be more vulnerable to ageing than item memory accuracy (e.g., Cansino, 2009). Although source memory is concerned with aspects that are often beyond our focal centre of attention at a given point in time, the ability to remember source information enables us to evaluate our knowledge and beliefs, and is therefore essential for everyday social interactions and for our subjective experience of autobiographical recollection. Failures of source memory underlie phenomena such as cryptomnesia (e.g., Macrae, Bodenhasen, & Calvini, 1999) and confabulation (e.g., Johnson, 1991). Additionally, it has even been argued that most tests of recognition memory are essentially source memory tests (Anderson & Bower, 1972). Since all participants have encountered countless words and objects outside of the experimental session, recognition tests in effect require participants to discriminate study list items from ones encountered extra-experimentally.

There are limitations to using source memory as an index of recollection. For instance, closed-ended source memory questions may not always offer a comprehensive measure of recollection (e.g., it is possible to recollect source information that has been neglected by the question, such as remembering that one had to sneeze when studying the word ‘pocket’). However, it is a more objective measure compared to self-reports of familiarity and recollection such as the R/K procedure. Despite the close ties between source memory and recollection, Hicks, Marsh, and Ritschel (2002) found that correct perceptual source memory judgements are not necessarily accompanied by a greater number of ‘remember’ than ‘know’ judgements.

This implies that there may be some qualitative differences between source memory and generic recollection. Specifically, the high proportion of correct source responses that corresponded to K responses in Hicks et al.’s (2002) experiment suggest that source judgements can successfully use partial memorial information which lack clarity or detail.

1.3 Fluency, Priming, and Source Memory

Despite the abundance of studies examining their roles in item memory, few experiments to date have investigated the role of processing fluency and priming in source memory. One example is Kelley, Jacoby, and Hollingshead (1989) in which participants studied a list consisting of auditorily and visually presented words, and during each trial in the test phase, they were asked to first identify rapidly presented old or new words, and then make a seen/heard/new judgement. Using participants’ probability of identification as an indirect index of perceptual fluency, Kelly et al. demonstrated that correctly identified items were more likely than incorrectly identified items to be judged as previously seen, irrespective of whether the items were previously seen, heard, or new.

Another example is Kurilla (2011), which reported a series of experiments using a masked repetition priming paradigm. In Experiments 1A and 1B, participants were presented with primed and unprimed target and lure words, and were instructed to make either an old/new decision followed by a seen/heard source decision (1A) or a combined seen/heard/new decision (1B). Experiment 2 tested whether masked repetition priming may also affect source memory for different perceptual features within the same modality (i.e., two different font types). Results indicated that for primed items, participants were more likely to report that they had studied the words in the same font style that matched the font presented in the test phase, irrespective of whether the words were actually presented in the same font at study and test. Similar to Kelly et al.'s (1989) experiments, Kurilla's two experiments showed that priming increased the proportion of "seen" responses but had no effect on the proportion of "heard" responses, and Kurilla concluded that primed items have a greater tendency to be endorsed as being presented in the same sensory or perceptual form during study and test.

Yonelinas (1999) proposed that familiarity cannot be used for source judgements when two source attributes are of approximately equal familiarity (e.g., words spoken by either a male or a female experimenter), as is the case in most source monitoring experiments. Consistent with dual-process assumptions, Perfect, Mayes, Downes, and Van Eijk (1996) showed that participants were usually able to make accurate source identifications for items with "remember" (R) responses but not for items with "know" (K) responses. It has also been observed that both recollection and source memory performance can show impairments despite item recognition performance remaining intact in amnesic patients (e.g., Shimamura & Squire, 1991), patients with frontal lobe damage (Janowsky, Shimamura, & Squire, 1989), and older participants (e.g., Schacter, Kaszniak, Kihlstrom, & Valderri, 1991).

Supporting the single-system view, studies have demonstrated that the recollection of source information can be assisted by vague, incomplete information such as whether a to-be-remembered item was pleasant or unpleasant, and that an above-chance proportion of K responses can be accompanied by accurate source judgements (Duarte, Ranganath, Trujillo, & Knight, 2006; Hicks et al., 2002; Koriat, Levy-Sadot, Edry, & de Marcas, 2003). Despite a greater number of neuroimaging and neuropsychological studies seemingly supporting the dual-process account, many also suggest that recollection (e.g., Squire, Wixted, & Clark, 2007; Turk-Browne, Yi, & Chun, 2006), and even source memory (Kirwan, Wixted, & Squire, 2008), share similar or overlapping neural structures and processes with recollection.

Although studies such as those by Kelley et al. (1989) and Kurilla (2011) imply that source memory judgements can be affected when processing fluency is manipulated through priming, they did not address whether fluency is related to source memory accuracy, nor did either of those studies directly measure variations in fluency. The continuous-identification with recognition (CID-R) paradigm (Stark & McClelland, 2000) allows concurrent measures of priming, fluency, and recognition to be obtained for every test item. In the CID-R test procedure, an old (target) or new (lure) item from the study list initially appears at a fragmented level. The item then gradually clarifies via a progressive demasking paradigm, and participants press a button immediately when they can identify the item, and then make a recognition response (e.g., old/new or confidence rating). Their identification response times (RTs) form the basis of a measure for priming and fluency. Therefore, the CID-R task additionally allows for a finer-grained measurement of variations in fluency than the perceptual identification task used by Kelley et al. (1989).

Using a modified CID-R paradigm to examine the relationship between priming and R/K judgements, Berry, Shanks, Speekenbrink, and Henson (2012, Experiment 3) found that trials with R responses were associated with the fastest identification RTs, followed by K trials and ‘new’ trials. If fluency can be linked to source memory decisions, one implication of Berry et al.’s findings is that faster identification RTs on the CID-R task might occur for trials in which participants are able to correctly remember source information compared to trials on which source judgements are incorrect. The present experiments in Chapters 2-4 employed a test procedure combining CID-R and source memory confidence ratings in order to examine this expectation. The chapters will explore and discuss several potential factors that may moderate the relationship between fluency and source memory accuracy.

1.4 Fluency and source information in metamemory

Memory in naturalistic settings consists of rich and complex interplays between item and source information, such as between to-be-remembered objects and their contextual features. For example, source information retrieval is usually dependent on successful item retrieval (Cooks, Marsh, & Hicks, 2006). However, there are also numerous situations in which source and contextual information or characteristics can substantially affect item memory performance. In the perceptual match effect (also known as the *modality effect*; Murdock & Walker, 1969), item information is remembered more successfully under greater perceptual similarity between the encoding stimulus and test probe. In the context reinstatement effect, memory performance is enhanced when encoding and retrieval are both conducted in the same (or highly similar) context (e.g., Rutherford, 2004; Smith & Vela, 2001; c.f. Hockley, 2008). Regardless of whether or not a source attribute could really affect memory performance in a given scenario, people often have their own metacognitive preconceptions of how a source would influence their ability to remember an item, for better or for worse (e.g., Hanczakowski, Zawadzka, Collie, & Macken, 2016).

Another example underscoring the impact of source information on metamemory is the font size effect on judgments of learning (JOLs; i.e., subjective estimates of the likelihood that a given item will be remembered at a future memory test). This effect refers to the fact that people assign higher JOLs to words with larger font sizes than to words with smaller font sizes, despite font size having no actual effect on retention performance (Rhodes & Castel, 2008). The font size effect on JOLs is robust and well replicated (e.g., Ball, Klein, & Brewer, 2014; Besken, 2016; Hu et al., 2015; Hu, Liu, Li, & Luo, 2016; Kornell, Rhodes, Castel, & Tauber, 2011; Li, Xie, Li, & Li, 2015; Miele, Finn, & Molden, 2011; Mueller et al., 2014; Price & Harrison, 2017; Price, McElroy, & Martin, 2016; Susser, Mulligan, & Besken, 2013). Importantly, JOLs can determine individuals’ study strategies (Metcalfe & Finn, 2008; Yang, Potts, & Shanks, 2017b). Any process dissociation between JOLs and actual memory performance can potentially produce inefficient learning and remembering (e.g., Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013; Yang et al., 2017b; Yang, Sun, & Shanks, 2017), and hence, understanding such process dissociations is essential to developing interventions to improve individuals’ study strategies.

Two main theories have been proposed to account for the font size effect on JOLs. The first is a belief-based theory, which postulates that: firstly, people hold *a priori* beliefs that large-font words are easier to remember or more important than small-font words; and secondly, people incorporate these beliefs into their JOLs. Mueller et al. (2014) reported that some people believe that large-font words are more important than small-font words, Castel (2007) found that perceived importance can mediate JOLs, and Rhodes and Castel (2008) proposed that participants might believe that a large font signals the importance of a study item within the context of an experiment. It is therefore possible that the difference in perceived importance between large- and small-font words may produce the font size effect on JOLs (Rhodes & Castel, 2008). Mueller et al. (2014) also found that some people believe that large-font words are easier to remember, and therefore suggested that people apply this belief in forming their JOLs (Mueller & Dunlosky, 2017). Moreover, Hu et al. (2015) found that the font size effect on JOLs is significantly predicted by variability in people’s beliefs about the difficulty of remembering large and small words. Altogether, these findings support the belief theory (based either on beliefs about importance or about ease of remembering) as an account for the font size effect on JOLs.

The alternative explanation postulates that large words are processed with greater perceptual fluency than small words: The experience of fluency during encoding produces a subjective feeling-of-knowing, and this subjective feeling acts as a basis for assessing learning status (Koriat and Bjork, 2006, Koriat and Ma'ayan, 2005, Mueller et al., 2013, Undorf et al., 2017). Previous studies have supplied convincing evidence for a fluency effect on JOLs, that is, more fluent processing produces higher JOLs (Ball et al., 2014, Besken and Mulligan, 2013, Hertzog et al., 2003, Magreehan et al., 2016, Undorf et al., 2017, Yang et al., 2017b).

However, only two empirical studies thus far have directly examined the role of fluency in the font size effect on JOLs. In Rhodes and Castel’s (2008) Experiment 6, some words were presented in a standard format (e.g., ‘computer’) and others in a format with alternating lowercase and uppercase letters (e.g., ‘gArDeN’). The experiment revealed a font size effect on JOLs in the standard format condition but not in the alternating format condition. The authors proposed that differences in perceptual fluency between large- and small-font words were disrupted in the alternating format condition. However, Mueller et al. (2014) argued that Rhodes and Castel’s (2008) Experiment 6 cannot provide unequivocal evidence to support the fluency theory, and that prior beliefs can equally well explain the results: Participants may simply not believe that large but alternating font words are easier to remember than small alternating font words.

Mueller et al. (2014) conducted a further study to test the fluency theory by employing a lexical decision task in their Experiment 1. Words (e.g., ‘chicken’) and non-words (e.g., ‘arage’) were sequentially presented in large or small font sizes. Participants were instructed to decide, as quickly and accurately as they could, whether the presented item was a word or a non-word. Mueller et al. (2014) found no difference in RTs between large and small words, and hence suggested that “processing fluency, as measured by the lexical decision task, is not mediating the font-size effect” (p. 4).

The questions of whether processing fluency can affect JOLs, and whether processing can affect the font size effect on JOLs, both remain controversial. The strength of evidence against the fluency theory will be further reviewed and discussed in Chapter 6. Experiments 4-6 will measure fluency *during encoding* using a continuous identification (CID) task (i.e., during the word presentation and prior to the JOL rating of each presented word). This task is nearly structurally identical to the CID-R task from Experiments 1-3b, except that no recognition testing is involved. To the best of our knowledge, no previous metamemory research has used the CID task to measure fluency, and we anticipated that this task would be more sensitive than previously-employed paradigms (e.g., lexical decision task) to variations in perceptual fluency. Through manipulating font size, Experiments 4-6 employed the CID task to determine whether or not perceptual fluency underlies the stimulus size effect on JOLs: by first exploring whether there is a difference in perceptual fluency between large- and small-font words, and then assessing whether perceptual fluency mediates the font size effect on JOLs. If perceptual fluency indeed differs according to font size *and* mediates the font size effect, the present experiments will support the fluency theory as an account for the font size effect on JOLs, which will also imply that perceptual fluency can affect JOLs.

1.6 General Aims

The broad aim of this thesis is to further explore the relationships or distinctions between implicit and explicit elements of memory: specifically, through examining the role of processing fluency in source memory and metamemory judgements. The mutual links between fluency, item memory, source memory, and metamemory will also be discussed throughout the following chapters.

Chapter 2: Multidimensional Source Memory and Processing Fluency

According to Johnson et al.'s (1993) Source Monitoring Framework, there are multiple cues to source, including information from the sensory/perceptual, contextual (spatial/temporal), semantic, affective, and cognitive-operational dimensions. Multidimensional source memory is ecologically important because episodic context is rarely constrained by a single source dimension in everyday life. To illustrate, we can often simultaneously retrieve multiple source information pertaining to an item which we own, such as where and when we bought the item. Despite of this, and despite source memory being regarded as a multidimensional construct in the Source Monitoring Framework, the vast majority of source memory experiments have tested only one dimension of source information (e.g., colour) with two attributional variations (e.g., red/green). A main purpose of this chapter is to address the lack of research on memory involving multiple source attributes which are crossed at encoding and jointly retrieved (see Hicks & Starns, 2015, for a review).

One example of multidimensional source memory research is Meiser and Bröder (2002). In their first experiment, Meiser and Bröder (2002) crossed the font size (small/large) of study words with presentation location (upper/lower screen location) and demonstrated that participants’ memory for one source dimension of an item was better when they correctly remembered the other source dimension than when they did not. This demonstration of stochastic dependence between memory for individual source dimensions suggests that remembering a source attribute on one dimension may facilitate or cue the retrieval of a source attribute on another dimension. A subsequent study (Meiser & Sattler, 2007) showed that, under perceptual encoding conditions, the proportion of items receiving correct source judgements on both dimensions was greater for R than K items, whereas the proportion of items receiving a correct source judgement on just one of the two dimensions was greater for K than R items.

In light of these findings, Experiment 1 investigated the relationship between identification RTs and source memory accuracy for font size and location. If the joint retrieval of features from two contextual dimensions depends exclusively on conscious recollection, it is unlikely that identification RTs on the CID-R task corresponding to items with correct responses on both source dimensions will differ significantly from those corresponding to correctly recognised items with incorrect responses on both dimensions. However, it is uncertain the extent to which the item-level identification RTs could be expected to differ between the correct retrieval of both, or just one, of the source dimensions.

In addition, as R judgements are previously reported to be associated with greater subjective confidence ratings and more accurate source memory on unidimensional source memory tests (e.g., Wixted, Mickes, & Squire, 2010), the present study explored how R and K responses correspond to source memory confidence ratings collapsed across two contextual dimensions. Given the findings of previous experiments, one could expect that R responses would receive the highest source confidence ratings followed by K then G responses, and that the proportions of trials with zero, one, or two sources correctly retrieved would vary as a function of R, K, and G item ratings.

2.1 Experiment 1

The aim of Experiment 1 was to investigate how fluency relates to source memory responses as well as the more prevalently used R/K or R/K/G responses, in the context of having two source dimensions (i.e., font size, and location on screen) which are crossed at encoding and retrieved jointly at test. It is known that R judgements correspond to both higher confidence ratings, faster identification RTs on the CID-R task (Berry et al., 2012), and greater accuracy on unidimensional source retrieval (Wixted et al., 2010). Therefore, we predicted that when at least one of the two source dimensions is correctly retrieved, memory at the item level will tend to receive an R rating, identification RTs will be shorter, and confidence ratings higher, as compared to when both source dimensions are incorrectly retrieved.

2.1.1 Method

2.1.1.1 Participants

Based on an *a priori* power estimation calculated using G*Power, a minimum of 48 participants[1] would be required to detect a significant one-way difference (α = .05; two-tailed), of a medium effect size (f = 0.25; Cohen, 1992) at 0.80 power, between mean identification response times of trials with one, two, or zero correct source attributions, assuming a conservative

---

[1] We also referred to Meiser and Bröder's (2002) Experiment 2, which had a sample of N = 43.

correlation of $r = .10$ amongst repeated measures. Fifty University College London (UCL) students were recruited to participate in the experiment for partial course credit or cash payment (£7.50). However, two participants had to be discarded from data analyses for failure to follow task instructions, leaving an effective total sample of 48, $M = 21.7$ years old, $SD = 3.14$; 38 females, 10 males. For this and all subsequent experiments, all participants spoke English fluently, reported normal or corrected-to-normal vision, and provided written consent for taking part in the experiment which was approved by the Ethics Committee of the UCL Department of Experimental Psychology.

2.1.1.2 Design and Materials

The design of the study was adapted from the paradigms developed by Meiser and Bröder (2002) and Stark and McClelland (2000). The source attributes and study phase procedures were nearly identical to those used by Meiser and Bröder. The test phase procedures mostly followed Stark and McClelland's CID-R procedure. However, instead of using old/new questions, participants were asked to make a remember/know/guess/new judgement after identifying each test item, and were asked to provide confidence ratings for their memory of the two source dimensions associated with the test item if the item had received a “remember” (R), “know” (K), or “guess” (G) judgement. The G response option was included in an effort to reduce guessing-related noise in K responses (Eldridge, Sarfatti, & Knowlton, 2002).

Experimental materials and instructions for this and all subsequent experiments were presented on a Dell PC monitor using Psychtoolbox 3 extensions (Brainard, 1997; Kleiner et al, 2007; http://psychtoolbox.org/; Pelli, 1997) for MATLAB (Mathworks, Natwick, MA). A total of 134 monosyllabic English nouns were selected from the MRC Psycholinguistic Database (Coltheart, 1981). Each word had 4-6 letters, a Kučera-Francis Frequency score of 3-20, a Concreteness score of 400-670, and an Imageability score of 424-600. Sixty-four of these words served as targets, which were presented at both study and test, and 64 served as lures, which appeared only at test. Half of the remaining six words were used as primacy buffers which appeared at the beginning of the study phase and the other half were used as recency buffers which appeared at the end of the study phase. The buffer words did not appear on the test.

During the study phase, the appearance of each word could vary on two source dimensions: font size (small/large) and location on the screen (lower/upper location). The vertical axes of the words presented at the upper and lower locations were approximately 15 cm apart, with equal distance to the centre of the screen. The character heights for the small and large fonts were 0.7 cm (20 pt.) and 1.8 cm (51 pt.), respectively. The 64 target words were randomly selected, and their presentation was programmed such that a total of 16 words appeared in each of the four presentation formats (i.e., small font, lower location; large font, lower location; small font, upper location; and large font, upper location). Experimental instructions and all words appearing in the test phase (both targets and lures) appeared in a medium font size of approximately 1.2 cm (34 pt.) character height. The Courier New font style was used for all stimuli and instructions in the experiment. A microphone headset was used for audio recording during the test phase. All statistical analyses were conducted using RStudio running R version 3.3.1 (R Core Team, Vienna, Austria) in this and all subsequent experiments.

2.1.1.3 Procedure

Participants went through the procedure individually in a single session which lasted approximately 50 minutes. Before the study phase began, they were informed that words would appear on screen one at a time and were asked to “try to remember as many of them as [they] can” and that they would be tested during the second part of the experiment. The instructions for the study phase did not make any reference to the varying appearance of the words, nor did the instructions provide any further detail on the nature of the memory test. The 64 target words were displayed on screen one at a time for 4 s with an interstimulus interval of 1 s.

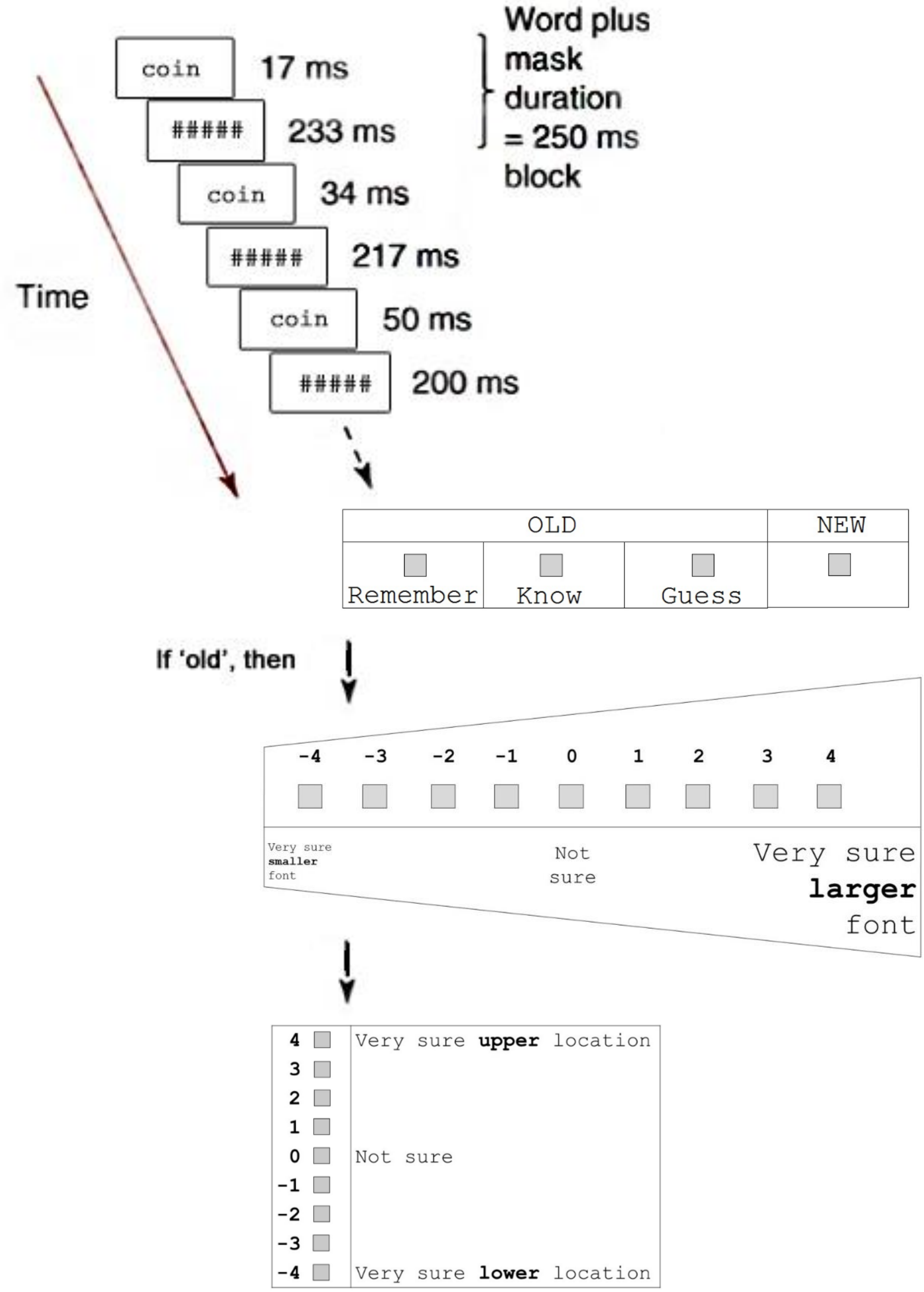

Figure 1. Diagram of the test procedures in Experiment 1.

After the end of the study phase, instructions describing the test phase procedures appeared on screen. A typical test trial began with a CID-R task which required participants to identify a word gradually presented on screen via progressive demasking (see Figure 1). In the task, a mask (i.e., a string of six hash symbols: ######) was displayed at the centre of the screen for 500 ms in medium font size, then a target word or a lure word was presented for 17 ms at the same location and in the same font size as the mask, followed by another presentation of the mask for 233 ms. The word was immediately presented again for 34 ms, and the mask was then displayed for 216 seconds. Thus, the 250 ms word-mask block was repeated with the display duration of the word increased by 17 ms and the duration of the mask decreased by 17 ms in each subsequent repetition of the block until the mask duration was 0 (i.e., 14 blocks) or until participants made an identification response (see Figure 1 for an illustration of the CID-R task and test questions).

To make an identification response, participants pressed the space bar as soon as they could identify the test word, and the RT to identifying the word was recorded. The word was replaced with the mask immediately after participants' key press, and participants were instructed to read aloud the word which they had just identified. The microphone automatically recorded participants’ identification responses which were verified offline by the experimenter. Participants then made a recognition judgement by clicking to indicate whether they remembered, knew, or guessed that the tested word had appeared in the study phase, or whether the word was new. The instructions for the four recognition response options were adapted from those used by Gardiner (1988) and Eldridge et al. (2002), and were described to the participant as follows:

> “If your recognition of the word is accompanied by a conscious recollection of your prior experience of the word during the study phase, select *REMEMBER*.
>
> If you can recognise the word as having occurred in the study phase, but your recognition is not accompanied by a conscious revival of the event when the word was presented to you earlier, *KNOW*.
>
> There might also be times when you do not remember the word, nor do you know it, but you might want to *GUESS* that it was one of the words you saw during the study phase.
>
> If you think the word was NOT presented during the study phase, select *NEW*.”

If participants made an R, K, or G response, they proceeded to answer two source memory questions which asked them to rate on a 9-point confidence scale whether the test word had been studied in a larger or smaller font (-4 indicated “very sure smaller font”, 0 indicated “not sure”, and 4 indicated “very sure larger font”), and whether it had appeared at the upper or lower screen location during the study phase (-4 indicated “very sure lower location”, 0 indicated “not sure”, and 4 indicated “very sure upper location”). The order of presentation of the two source questions was counterbalanced between participants.

For trials where participants reported the word being “new”, they proceeded to the next test trial without being presented with the source questions. On any particular trial, if participants were unable to identify the test word during the 14 presentation blocks (= 3500 ms), a “trial timeout” message appeared on screen, and participants directly proceeded to the next test trial. In total, there were 128 trials in the test phase per participant. All misidentified trials were excluded from subsequent data analyses.

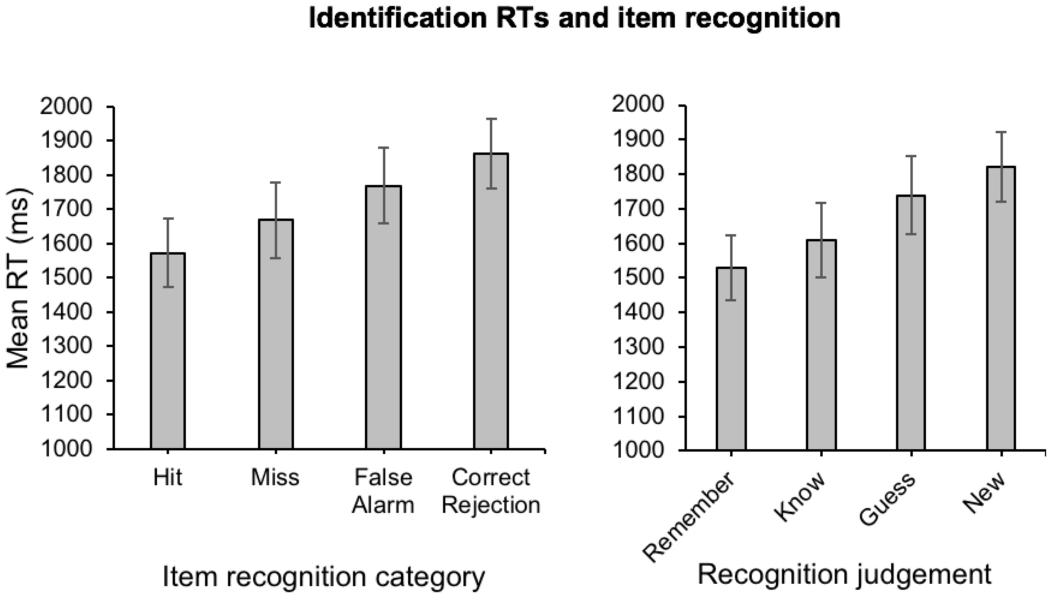

Figure 2. Left panel: mean item identification RTs (ms) for hits, misses, false alarms, and correct rejections in Experiment 1. Right panel: mean identification RTs for item hit trials according to “remember”, “know”, “guess”, and “new” response categories. Error bars indicate 95% confidence intervals of the mean.

2.1.2 Results

A total of 382 trials (6.22% of all trials) were excluded from the subsequent analyses due to misidentification of the test word or lack of an identification response. All 48 participants correctly identified the test word on at least 85% of their trials.

2.1.2.1 Recognition memory

Test trials receiving R, K, and G were considered as “old” responses. Across all valid test trials, the hit rate was .80 and the false alarm rate was .25.

Since the present experiment only collected RTs in the CID-R portion of the test phase, all references to identification RTs indicate RTs to the perceptual identification of the test item being presented at the beginning of each test trial. The left panel of Figure 2 shows mean identification RTs to hit, miss, false alarm, and correct rejection trials. A one-way repeated-measures analysis of variance (ANOVA)[2] revealed a significant difference between the identification RTs of the four recognition categories, $F(3,135) = 17.73$, $p < .001$, $\eta_p^2 = 0.28$. Post-hoc *t*-tests with Bonferroni correction revealed that identification RTs did not differ significantly between hits and misses, $t(45) = 2.01$, $p = .2152$, $d = 0.27$, or between misses and false alarms, $t(45) = 2.10$, $p = .1380$, $d = 0.31$, but identification RTs to false alarms were faster than those to correct rejections, $t(45) = 2.71$, $p = .0288$, $d = 0.40$.

Mean RTs to R, K, and G responses across item memory hit trials (Figure 2, right panel) were significantly different as determined by a one-way repeated-measures ANOVA, $F(1.59, 78.87) = 10.35$, $p < .001$, $\eta_p^2 = 0.18$. Post-hoc *t*-tests with Bonferroni correction revealed that identification RTs were faster for trials with R than K responses, $t(47) = 2.53$, $p = .0294$, $d = 0.37$, and K trials were associated with faster identification RTs compared to G trials, $t(47) = 2.55$, $p = .0281$, $d = 0.37$.

For 46 out of 48 participants, mean identification RTs to old items were at least 10 ms faster than mean RTs to new items. The mean identification RT for new items minus the mean identification RT for old items was calculated for each participant, and the overall difference indicated a significant priming effect, $M = 237$ ms, $SEM = 36$ ms, $t(47) = 7.87$, p < .001.

---

[2] Two participants were excluded from this analysis for having no false alarm trials.

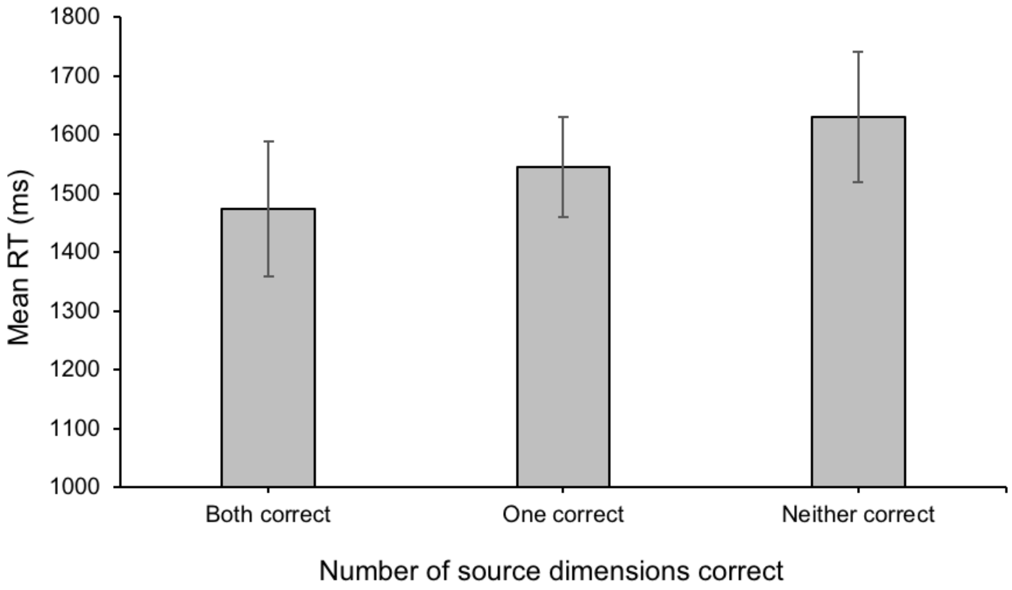

Figure 3. Mean identification RTs (ms) for recognition hit trials with correct source judgements on both, one, or none of the source dimensions in Experiment 1. Error bars indicate 95% confidence intervals of the mean.

2.1.2.2 Source memory

Source judgements in item recognition hit trials were categorised as correct only if participants responded with an absolute confidence rating of 1 or more on the correct source. Participants' mean identification RT difference scores for correct versus incorrect source trials showed no significant difference between the size (*M* = 99 ms, *SD* = 240 ms) and location dimensions (*M* = 114 ms, *SD* = 243), *t*(47) = 0.59, *p* = .56. Twenty-nine percent of responses had both source dimensions correct and 33% had one source correct. Figure 3 shows mean identification RTs according to the number of correct source identifications per trial.

A one-way repeated measures ANOVA[3] indicated a significant difference in identification RTs across trials with correct judgements on both source dimensions, trials with one correctly-identified source, and trials with no correct source judgements, *F*(1.43, 65.58) = 5.50, *p* = .0127, $\eta_p^2 = 0.11$. Post-hoc pairwise comparisons with Bonferroni correction revealed that identification RTs for trials with both sources correct were faster than identification RTs for trials with no sources correct, *t*(46) = 2.67, *p* = .0319, *d* = 0.39, and identification RTs for trials with one source correct were faster than identification RTs for trials with no sources correct, *t*(46) = 2.73, *p* = .0272, *d* = 0.40. However, the difference in identification RTs between trials with two correctly identified sources versus trials with only one correct source was not significant, *t*(46) = 1.49, *p* = .4332, *d* = 0.22.

A sign test was also conducted to examine the proportion of participants showing faster mean identification RTs for trials with two correct source judgements than for trials with no correct source judgements. In addition to the two participants excluded for having no trials with both sources correct, another participant was excluded for having tied mean identification RTs (i.e., a mean identification RT difference of less than 10 ms between trials with two correct source judgements and trials with two incorrect source judgements). Thirty participants showed faster mean RTs for trials in which both source judgements were correct compared to trials in which both source judgements were incorrect, and 15 participants showed the reverse pattern (*Z* = 2.09, *p* = .0369).

---

[3] For this analysis, one participant was excluded for not having any trials with both sources correct.

Analyses were conducted to examine the relationship between item recognition responses, source accuracy, and source confidence ratings. For R, K, and G trials, Table 1 shows response frequencies according to the number of correct source identifications per trial, as well as the mean total confidence ratings (sum of the absolute values of the confidence ratings for font size and location) per trial. A one-way repeated-measures ANOVA revealed a significant difference in confidence ratings per trial (collapsed across the two source dimensions) between trials receiving R, K, or G responses, $F(2, 87) = 104.7$, $p < .001$, $\eta_p^2 = 0.71$. Post-hoc pairwise comparisons with Bonferroni correction demonstrated that R trials received higher overall source confidence ratings than K trials, $t(44) = 6.27$, $p < .001$, $d = 0.93$, and K trials received higher source confidence than G trials, $t(42) = 5.20$, $p < .001$, $d = 0.79$.

	Source accuracy			*M* summed confidence rating
	Both sources correct	One source correct	Neither source correct	
Remember	509	510	412	4.14 (3.68, 4.60)
Know	122	199	243	2.38 (1.87, 2.90)
Guess	43	66	225	1.15 (0.72, 1.59)

Table 1. Source accuracy frequencies and confidence ratings for R, K, and G trials according to number of correct source judgements in Experiment 1. Source response frequencies represent the total number of trials with two, one, or zero correct source judgements across all participants. Mean confidence ratings were obtained from the summed absolute confidence ratings for font size and location per trial (95% confidence interval of the mean is shown in brackets).

Further analyses were conducted to examine the relationship between item recognition responses and source memory accuracy. Two participants were excluded from the analyses due to having no K or G responses. Across 46 participants, there was a significant and medium correlation between item-level R/K/G responses and source accuracy, Spearman's $\rho = .31$, $p < .025$. This indicated a tendency for more source dimensions to be correctly remembered when the corresponding recognition response was R, followed by K then G. Thirty-eight participants had positive correlations of $\rho \geq .10$, five had weak positive correlations of $0 \leq \rho < .10$, and the remaining three participants had weak negative correlations of $0 > \rho > -.10$.

2.1.3 Discussion

A main objective of the present study was to investigate the connection between familiarity-related processes such as processing fluency and source memory, and in particular, their relationship to memory for multidimensional source information. Another objective of the study was to explore how the accuracy of multidimensional source memory might be reflected in subjective confidence ratings. In pursuit of these objectives, trials with correct item recognition responses (item hits) were classified into three categories of source accuracy according to whether they received correct responses on both, one, or neither of the source dimensions. Identification RTs were compared between the three categories of source accuracy in order to evaluate the relationship between fluency, source memory performance, and subjective confidence ratings (as summed across the two source dimensions per trial). The pattern of item recognition shown in the present study generally replicated the patterns reported previously in Berry et al. (2012). Overall, there was a linear trend for identification RTs to increase across hits, misses, false alarms, and correct rejections, and RTs increased linearly across R, K, and G trials. A priming effect (i.e., faster mean RTs to target items than lures) was observed in the vast majority of participants.

These results demonstrate that familiarity can contribute to R decisions, and contradict the interpretation of R responses as a process-pure index of recollection.

The present results concerning the proportions of R/K/G judgements corresponding to source accuracy categories differed somewhat from previous cross-dimensional source memory studies (Meiser & Bröder, 2002; Meiser & Sattler, 2007). Within R trials, we did not find a greater proportion of trials with two correctly identified source dimensions compared to the proportion with only one correctly identified dimension. Instead of finding a greater proportion of K than R judgements in trials with one correctly retrieved source, the opposite pattern was shown in the present results. These differences may be due to the use of a one-step R/K/G/N question procedure here instead of the two-step procedure (i.e., old/new, then R/K) at test employed in previous experiments. Even though the inclusion of a G response option has been demonstrated to reduce noise in K decisions by alleviating guessing-induced false alarm rates (Eldridge et al. 2002), it has been shown that K hit rates in one-step recognition tests remain lower than in two-step procedures, even after the addition of a G option (Bruno & Rutherford, 2010). Consequently, despite accounting for the number of G item hits, the proportion of K item hit trials in the present study is disproportionately smaller than the proportions reported in previous studies, which may have contributed to a floor effect. However, it would have been impractical to use the two-step recognition procedure. Apart from the aforementioned differences, the present findings largely agree with previous findings demonstrating a relationship between R/K/G judgements and multidimensional source accuracy and a linear trend for confidence ratings to increase across R, K, and G trials. This suggests that fluency can correspond to a subjective feeling of remembering the experience of an episodic event.

Of particular interest to the question of whether multidimensional source memory judgements are related to familiarity-based fluency, results from the present experiment indicated that identification RTs differed depending on source accuracy status (i.e., the number source of dimensions correctly retrieved per trial). The finding that trials with both sources correct tended to have faster RTs than trials with neither of the sources correct suggests that fluency can be linked to memory for more complete and accurate source information similar to how familiarity, when operationalised as K responses, can contribute to partial source retrieval (Hicks et al., 2002; Meiser & Bröder, 2002; Meiser & Sattler, 2007).

Despite obtaining slower identification RTs for trials with no correct source responses compared to trials with one and two correct source responses, the results revealed no significant RT difference between trials with both versus one of the source dimensions correct. Although this finding might appear to suggest that familiarity contributes to multidimensional source memory in a threshold manner, it is important to note the numerically smaller proportion of responses correct on both source dimensions, as well as the high probability of participants being able to correctly guess one of the two source dimensions, providing that a correct 'old' response was given earlier in a given test trial. Further research is needed to determine the precise phenomenological similarities and differences between partial versus complete retrieval of source information, and this may require the use of more than two attributes per source dimension in order to reduce the likelihood of correctly guessing one of the sources (Cansino, 2009).

The present experiment presented two distinct source dimensions crossed at encoding, in an attempt to achieve a closer approximation of source monitoring in real-life situations. However, real-life source memories are undoubtedly more complex and involve a richer variety of features that characterise the circumstances under which information is encoded.

Chapter 3: Effects of Levels of Processing on Source Memory and Fluency

A primary finding from the previous chapter is that accurate memory for source information is related to greater processing fluency as indexed by identification RTs on the CID-R task. The experiments in this chapter aimed to extend this finding to other encoding conditions and source modalities, by investigating the extent to which a levels of processing (LOP) manipulation at encoding might affect fluency and memory for temporal and spatial source information. Although items that are processed deeper and more elaborately at study are generally remembered better in explicit memory tests (Craik & Lockhart, 1972), this LOP effect may vary depending on the perceptual, conceptual, explicit, and implicit nature of the test (Roediger, Weldon, & Challis, 1989). Numerous studies have found no effect of LOP on implicit perceptual tests such as perceptual priming (e.g., Jacoby & Dallas, 1981) and word pair free-associations (Schacter & McGlynn, 1989). Contrarily, Challis and Brodbeck (1992) found a LOP effect on priming in a between-subjects word fragment completion task, whereas Roediger, Weldon, Stadler, and Riegler (1992) observed reversed LOP effects in both word fragment and word stem completion tasks involving pictorial stimuli, such that greater priming occurred in the graphemic encoding condition relative to the pleasantness rating condition.

LOP effects have also been demonstrated on explicit tests of associative memory for word pairs (Cohn & Moscovitch, 2007) and face-name pairs (Troyer, Häfliger, Cadieux, & Craik, 2006, Experiment 2), but fewer studies have examined how LOP specifically affects source memory. A study by Ragland et al. (2006) demonstrated an LOP effect on an internal source monitoring task which instructed participants to identify at test whether a word was a target presented during the deep or shallow encoding condition, or was new, such that participants showed better internal source discrimination for semantically versus perceptually processed words. However, given the dissociations in performance between external and internal source monitoring (e.g., ageing; Degl'Innocenti & Bäckman, 1996), it is unclear whether Ragland et al.'s (2006) results would generalise to external source monitoring tasks such as the one used in our Experiment 1.

Recent studies using R/K and R/K/G reports as an index of familiarity and recollection found that memory for the order (Easton, Webster, & Eacott, 2012) and sequence (Persson, Ainge, & O'Connor, 2016) of stimulus presentation could be retrieved accurately based on either familiarity alone or a combination of familiarity and recollection, although accurate source memory for other contextual dimensions was found to depend exclusively on recollection. Persson et al.'s (2016) experiments were conducted in an immersive virtual environment such that each item in the study phase was presented in the context of one of six possible weather conditions (e.g., rainy weather was characterised by the use of visual and sound effects of raindrops in the background), and temporal source judgements involved six possible sequence position options. Their results showed that contextual source memory performance was above chance for R responses but not F responses, whereas performance for temporal source memory was above chance for both R and F. On the basis of those results, Persson et al. (2016)

supported the dual process account but proposed that the role of familiarity in temporal source memory is an exception, since memory strength at retrieval can reflect the amount of time elapsed since the presentation of that stimulus at encoding.

3.1 Experiment 2a

The aim of the present experiment was to investigate whether identification RTs to trials with correct source responses would be faster for identification RTs on incorrect source responses for temporal and spatial dimensions, and how LOP might affect these identification RT differences. Although each target word had a temporal placement (first/second half) and screen location (upper/lower) at study, the experiment was not designed to test multidimensional source memory as in Experiment 1, thus each test trial focused only on one of the two available dimensions. Based on the findings from previous experiments on associative memory (Cohn & Moscovitch, 2007; Troyer et al., 2006), temporal source memory (Easton et al., 2012; Persson et al., 2016), and Experiment 1, we predicted that deeper processing at encoding would produce more accurate source memory responses, and that trials with correct source responses would have faster identification RTs, for both time and location, compared to trials with incorrect source responses.

3.1.1 Method

3.1.1.2 Participants

A total of 112 undergraduate students at UCL participated in the experiment as part of a laboratory class. Two participants were excluded from data analyses for failing to correctly identify more than 50% of target items at test, leaving an effective sample of *N* = 110; *M* = 19.00 years old, *SD* = 1.24. Half of all participants were randomly assigned to the shallow encoding condition of the experiment and the other half to the deep encoding condition, with 42 females and 13 males in each condition. Assuming that a medium effect size of *f* = 0.25 (Cohen, 1992) could be detected at the .05 *α* level (two-tailed), this sample size would provide an estimated post-hoc power of 0.92 as computed through G*power (Version 3.1.9.2; Faul, Erdfelder, Buchner, & Lang, 2009).

3.1.1.3 Design and Materials

The experiment had level of processing during encoding (shallow vs. deep) as a between-subjects factor and source memory dimension (time vs. location) as a within-subjects factor. Two orienting questions were used in each encoding condition. Shallow orienting questions which emphasised orthographic aspects of the item included “Does this word contain the letter 'a'?” and “Is this word exactly 5 letters long?”. Deep orienting questions which emphasised semantic aspects of the item included “Is this word bigger or smaller than a shoebox?” and “Is this word living or non-living?”.

Unlike Experiment 1 which tested participants’ memory for both source dimensions on each test trial, the present experiment only tested one dimension (i.e., time of presentation or screen location) per test trial. The selected source dimension was pre-allocated by the computer program to each test trial such that if a participant was able to progress to the memory judgement stage on all test trials, there would be an equal number of trials testing memory for time and for location. Differing from Experiment 1, the R/K/G/N question appearing after the CID-R component in each test trial was omitted and replaced by a one-step question involving a single O/N judgement with the O subcategorised into a ratings scale consisting of six options based on source identification and confidence. Another difference from Experiment 1 is that the zero-confidence option (“not sure”) was no longer available.

All experimental materials and instructions were presented on a Dell PC monitor in the Courier New font style with a font size of 1.2 cm (34 pt.). A total of 70 monosyllabic English nouns were selected from the MRC Psycholinguistic Database (Coltheart, 1981). Each word had 4 or 5 letters, a Kučera-Francis Frequency score of 1-82, a Concreteness score of 487-648, and an Imageability score of 335-617. Thirty-two of these words served as targets, 32 served as test lures, three served as primacy buffers, and three served as recency buffers. Half of the words presented at study were five-letter words, contained the letter 'a', were living objects, or were smaller than a shoebox. During the study phase, the vertical axes of the words presented at the upper and lower screen locations were approximately 15 cm apart and were equidistant from the centre of the screen.

3.1.1.4 Procedure

The experiment lasted approximately 15 minutes, and participants were tested in individual cubicles. Before the study phase began, they were informed that words would appear on screen one at a time along with questions, and were asked to answer those questions whilst trying to “remember as many of the words as [they] can”. The instructions for the study phase did not make any reference to the varying appearance of the words, nor provide any further detail on the source memory aspects of the upcoming test. During the study phase, the 32 target words and 6 buffer words were displayed one at a time in black font for a total of 4 s each, and 500 ms after the onset of each word, one of the two orienting questions within each LOP condition was alternately selected per trial to be presented at the centre of the screen in blue font. Participants in the shallow LOP condition used the “y” and “n” keys to respond “yes” or “no” respectively to the questions, and participants in the deep LOP condition used the keys “1” or “0” to respond “bigger”/“living” or “smaller”/“non-living” respectively. Once the responses to the orienting questions were recorded, the question then disappeared from the screen whereas the word remained on screen until 4 s had elapsed since the presentation onset of the word. The interstimulus interval between each word was 1 s. After participants had been presented with 19 words, the message “You are now halfway through the study list. Press <spacebar> to continue” was displayed in a green font colour at the centre of a blank screen.

After participants had studied the second half of the study list, instructions explaining the test phase procedures were presented on screen. In total, there were 64 trials in the test phase per participant, each beginning with a CID-R task involving a target or lure word. The CID-R task was identical to the one used in Experiment 1 with two exceptions: the mask now consisted of five hash symbols instead of six, and participants were instructed to type the word they had identified rather than speak their answers aloud due to equipment constraints. If the participant could accurately identify the word on time, they were asked to make a single-step judgement of item recognition and temporal or spatial source memory. The six “old” recognition options were grouped as a confidence scale with -3 indicating “very sure lower location”/“very sure first half” and 3 indicating “very sure upper location”/“very sure second half”. In place of “very sure”, lower confidence ratings were labelled “probably” (for -2 and 2) and “guess” (for -1 and 1). The “new” option was located to the right of the “old” options.

The source-O/N question was not presented on any test trial in which participants failed to correctly identify the word on time. If the participant was unable to identify the test word during the 14 presentation blocks (= 3500 ms), a “trial timeout” message appeared on screen, and the participant directly proceeded to the next test trial. If the participant pressed the space bar on time but entered the wrong word, they saw an “incorrect word entered” message before being directed to the next trial. All misidentified trials were excluded from subsequent data analyses.

3.1.2 Results

Across all participants, 335 trials (4.8% of all trials) were excluded from the subsequent analyses due to misidentification of the test word or lack of an identification response. Participants correctly identified the test word on at least 71.9% of their trials.

3.1.2.1 Recognition memory

The item recognition hit and false alarm rates were .85 and .20 respectively across all valid trials. The mean number of item hit trials did not differ significantly between the deep (*M* = 26.51, *SD* = 4.31) and shallow (*M* = 25.60, *SD* = 3.86) conditions across participants, *t*(108) = 1.67, *p* = .2464, *d* = 0.22, but there was a significant difference in the number of false alarm trials, $M_{\text{(deep)}}$ = 3.81, $SD_{\text{(deep)}}$ = 4.26, $M_{\text{(shallow)}}$ = 8.44, $SD_{\text{(shallow)}}$ = 6.31, *t*(108) = 4.18, *p* < .001, *d* = 0.86. Corrected recognition scores (i.e., Hits + Correct Rejections) were significantly higher for participants in the deep condition (*M* = 51.84, *SD* = 7.19) versus the shallow condition (*M* = 46.49, *SD* = 7.23), *t*(108) = 3.88, *p* < .0001, *d* = .74 .

There was a significant mean difference between identification RTs to new and old items (*M* = 142 ms, *SD* = 173) which indicated a reliable priming effect across participants, *t*(109) = 8.62, *p* < .0001, *d* = 0.82. Across all participants, 89 provided evidence of priming with mean identification RTs to old items being at least 10 ms faster than mean RTs to new items, 18 showed RT differences in the opposite direction, and three had tied RTs (i.e., differences of 10 ms or less). Priming did not differ between the deep (*M* = 127 ms, *SD* = 168) and shallow (*M* = 152 ms, *SD* = 175) conditions, *t*(108) = 0.42, *p* = .6725, *d* = 0.14.

3.1.2.2 Source memory

All analyses on source memory performance were carried out using data from item hit trials. Correct source responses occurred on 60% of the shallow location trials, 59% of the shallow temporal trials, 61% of the deep location trials, and 62% of the deep temporal trials. A three-way mixed ANOVA was conducted to examine the effects of LOP, source dimension, and source memory accuracy on identification RTs (Figure 4, panels a and

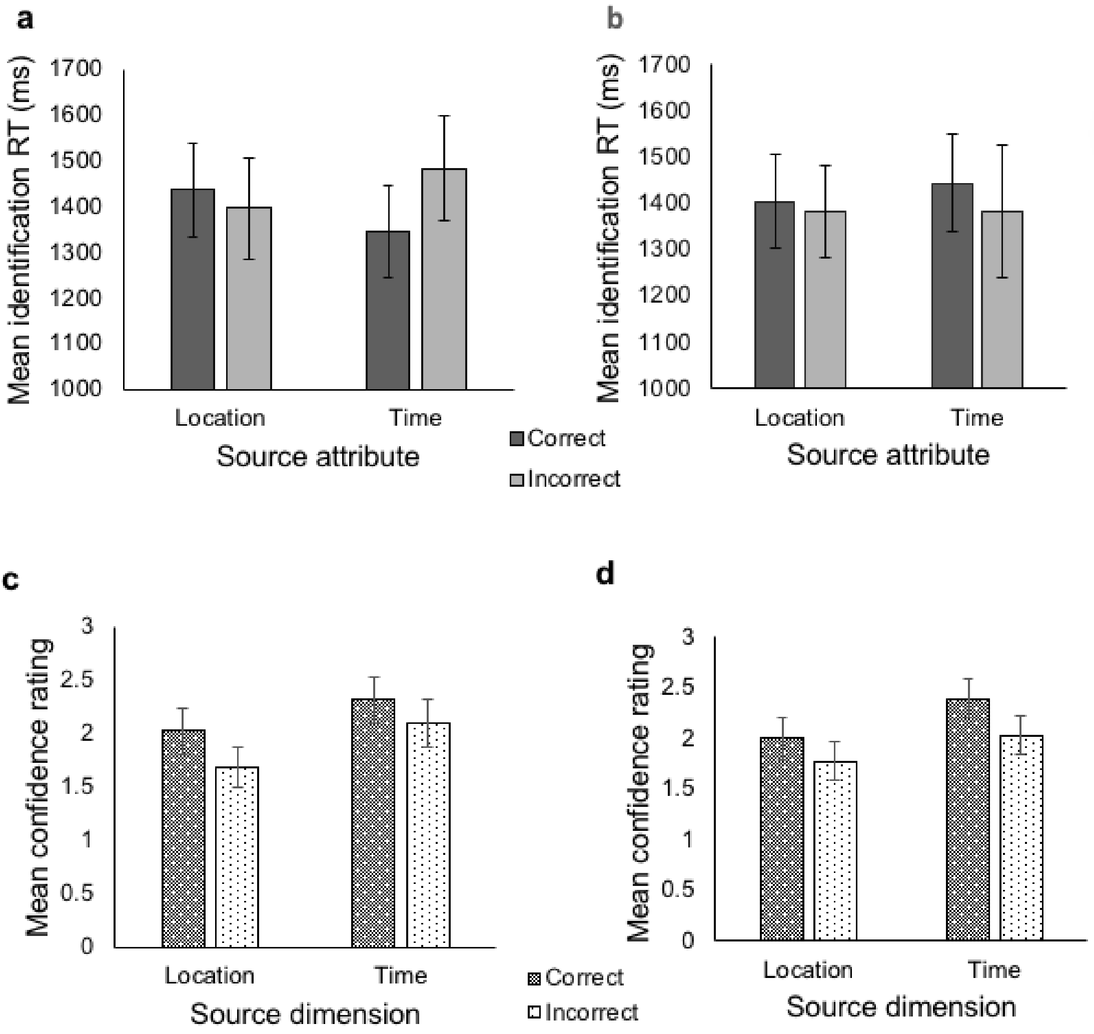

Figure 4. Effects of LOP, source dimension, and source memory accuracy on identification RTs (panels a and b) and source confidence ratings (panels c and d) in the shallow (panels a and c) and deep (panels b and d) LOP conditions of Experiment 2a. Error bars indicate 95% confidence intervals of the mean.

b). The ANOVA did not reveal any main effects but there was a significant three-way interaction, *F*(1, 105) = 6.70, *p* = .0110, η_p^2 = 0.06, which was followed up with two separate two-way ANOVAs for each LOP condition. The shallow condition ANOVA revealed a significant interaction between source dimension and source accuracy, *F*(1, 54) = 11.79, *p* = .0012, η_p^2 = 0.18, and tests of simple main effects found that mean identification RTs in the shallow condition were significantly longer when the temporal source dimension was incorrectly remembered than when it was correctly remembered, *F*(1, 54) =

14.01, $p < .001$, $\eta_p^2 = 0.22$, whereas no significant difference in RTs to correct vs. incorrect source trials was observed when the location dimension was tested, $F(1, 54) = 1.24$, $p > .05$, $\eta_p^2 = 0.01$. On trials testing the temporal dimension in the shallow LOP condition, 33 participants showed slower mean identification RTs when the dimension was incorrectly remembered, 20 showed the opposite pattern, and 2 had tied identification RTs with RT differences of 10 ms or less. There were no significant main effects or interaction effects in the ANOVA for the deep condition.

Another three-way mixed ANOVA examined the effects of LOP, source dimension, and source memory accuracy on confidence ratings (Figure 4, panels c and d). There was a significant main effect of source dimension on source confidence ratings such that higher mean ratings were given on trials testing the temporal source dimension versus trials testing the location dimension, $F(1, 104) = 60.73$, $p < .0001$, $\eta_p^2 = 3.69$. There was also a significant main effect of source memory accuracy on confidence ratings, such that higher mean ratings were associated with correct source responses compared to incorrect source responses, $F(1, 106) = 65.60$, $p < .00001$, $\eta_p^2 = 3.82$.

3.1.3 Discussion

The item memory results from Experiment 2a were consistent with most of the recognition and priming results reported in previous LOP studies. Although deeper processing during encoding did not produce higher item recognition hit rates compared to shallow processing, it resulted in greater recognition accuracy as measured by the total number of hits and correct rejections. Priming at the item level was also unaffected by LOP, which corroborated previous findings employing perceptual identification tests (e.g., Jacoby & Dallas, 1981).

On a broader level, the source confidence results supported Experiment 1's finding that trials with correct source responses also tended to receive higher confidence ratings. Despite participants' higher confidence ratings on the temporal source dimension compared to the location source dimension, their performance on the source memory task did not indicate better recognition for temporal attributes versus source attributes, which suggested that the difficulty of remembering the two source dimensions were similar in this task. However, the analyses on the relationship between LOP, source memory accuracy, and source dimension on identification RTs revealed an unexpected interaction between the three factors. Experiment 1's pattern of faster identification RTs associated with correct-source trials compared to incorrect-source trials was only obtained in the temporal source dimension within the shallow LOP group.

In the shallow condition, it was unexpected that the faster identification RTs to trials with accurate versus inaccurate source responses occurred in the temporal source dimension but not in the location dimension, given that Experiment 1 demonstrated faster identification RTs to trials with accurate versus inaccurate source responses in the visual source modality. A possible explanation for this interaction is Persson et al.'s (2016) proposal that the contribution of familiarity to source memory is only exclusive to temporal forms of source memory. However, there were many fewer test trials on either source dimension compared to Experiment 1 (i.e., 16 vs. 64

trials, respectively), and thus some participants in Experiment 2a might not have had sufficient time to familiarise themselves with the test format.

3.2 Experiment 2b

The aim of this experiment was to use a larger number of study and test trials per participant to replicate Experiment 2a's finding that identification RTs were faster to trials with correct time judgements compared to trials with incorrect time judgements in the shallow LOP condition, whereas identification RTs did not differ between trials with correct and incorrect location source judgements.

3.2.1 Method

3.2.1.1 Participants

Based on the effect size (d_z = 0.70) corresponding to the standardised identification RT difference scores between trials with correct versus incorrect temporal source responses in the shallow LOP condition of Experiment 2a, a power analysis using G*Power (Version 3.1.9.2; Faul, Erdfelder, Buchner, & Lang, 2009) indicated that a minimum of 18 participants would be required to detect a significant difference (α = .05) at 0.80 power. Twenty-one UCL students participated in the experiment for partial course credit or cash payment (£5.00). One participant was removed from data analyses for not following task instructions, leaving an effective sample of *N* = 20; *M* = 24.45 years old, *SD* = 6.07, 13 females, and 7 males.

3.2.1.2 Design and Materials

The within-subjects factor of interest was source memory dimension (time vs. location). This experiment had the same design as that of Experiment 2a, except that orienting questions were omitted during encoding, and the stimulus set included 64 target and 128 test items (twice as many as in Experiment 2a).

A total of 134 monosyllabic English nouns were selected from the MRC Psycholinguistic Database (Coltheart, 1981). These included the 70 nouns from Experiment 2a and additional nouns with similar properties, and thus all nouns had 4 or 5 letters, a Kučera-Francis Frequency score of 1-312, Concreteness scores of 487-670, and Imageability scores of 335-643. For each participant, each of the nouns was randomly assigned to be one of the 64 targets, one of the 64 lures, one of the 3 primacy buffers, or one of the 3 recency buffers. All instructions and stimuli were presented in the same font styles, font sizes, and format as in Experiment 2a. A microphone was used to record participants' word identification answers at test, which were spoken aloud (see Experiment 1) rather than typed.

3.2.1.3 Procedure

Participants took part in the experiment individually in a single session which lasted approximately 40 minutes. As in Experiments 1 and 2a, the study phase procedure involved intentional encoding of the word items and incidental encoding of the source attributes. The same presentation durations and interstimulus intervals were used as in Experiment 1.

After the end of the study phase, further instructions explaining the test procedures appeared on screen. The CID-R task was the same as the version used in Experiment 2a, except that participants were asked to identify words aloud after pressing the space bar instead of typing their answers. The recognition and source memory questions were also identical to the version used in Experiment 2a.

3.2.2 Results

Across all participants, 181 trials (7.1% of all trials) were excluded from the subsequent analyses due to misidentification of the test word or lack of an identification response. Each participant correctly identified the test word on at least 68.8% of their trials.

3.2.2.1 Recognition memory

The item recognition hit and false alarm rates were .79 and .25 respectively across all valid trials. The overall difference between mean identification RTs to new (M = 1755 ms, SD = 529) and old (M = 1592 ms, SD = 496) items indicated a significant priming effect across participants, $t(19) = 6.45$, $p < .0001$, $d = 0.32$, with 18 participants showing evidence of priming with mean identification RTs to old items being at least 10 ms faster than mean RTs to new items, and two participants showing RT differences in the opposite direction.

3.2.2.2 Source memory

All analyses on source memory performance were carried out using data from item hit trials. The mean identification RTs to trials with correctly and incorrectly responded source questions on time and location are shown

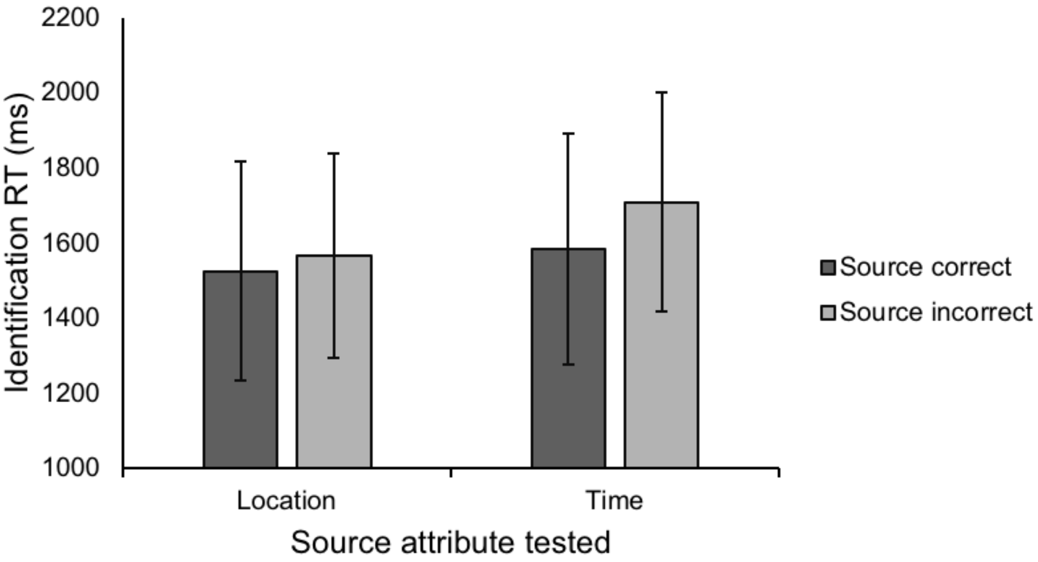

Figure 5. Mean identification RTs according to source memory accuracy and tested source dimension in Experiment 2b. Error bars indicate 95% confidence intervals of the mean.

in Figure 5. Sixty-six percent of source decisions were correct on the location trials and 63% were correct on the temporal trials. Since our sample size was planned based on Experiment 2a's effect of source accuracy on identification RTs on the temporal dimension (rather than the interaction effect), the present source memory results focused on *t*-tests.[4] Consistent with the results from Experiment 2a, identification RTs were significantly faster on trials where participants correctly answered whether the item had appeared in the first or second half of the study phase, compared to when the presentation time was incorrectly answered, *t*(19) = 2.52, *p* = .02, *d* = 0.56. For 15 participants, mean identification RTs to trials with correct temporal source responses were at least 10 ms faster than to trials with incorrect temporal responses, four participants showed the reverse pattern, and one

---

[4] An unplanned two-way ANOVA indicated a significant interaction effect of source dimension and source memory accuracy on identification RTs, *F*(1, 19) = 6.23, *p* = .02, η_p^2 = 0.06, concurring with Experiment 2a.

had tied identification RTs. There was no significant identification RT difference between trials with correct location responses versus trials with incorrect location responses $t(19) = 0.60$, $p > .05$, $d = 0.14$.

3.2.3 Discussion

Using a larger number of study and test trials per participant, Experiment 2b replicated the pattern of findings in the shallow LOP condition of Experiment 2a. Both experiments showed that, on the temporal source trials, mean identification RTs to source-correct trials were faster than to source-incorrect trials, but the same pattern was not observed on the location source trials. Based on an unplanned analysis, the size of this interaction effect in Experiment 2b was also equivalent to the effect size of the interaction found in 2a.

3.3: Summary and Discussion of Experiments 2a and 2b

The lack of a LOP effect on source memory accuracy in Experiments 2a and 2b contradicted previous results from tests of associative memory (Cohn & Moscovitch, 2007; Troyer et al., 2006) and internal source monitoring (Ragland et al., 2006). This could be due to the possibility that the deep or elaborative encoding instructions used in those studies tended to promote more unitisation (i.e., the encoding of an item with its surrounding contextual elements as a single, coherent unit of information), which can enhance associative memory (Graf & Schacter, 1989) and source attribution accuracy (Yonelinas, 1999). For example, Cohn and Moscovitch (2007) instructed participants in their deep associations condition to “produce a sentence, aloud, that contained the two words, was meaningful, and maintained both the form (i.e., singular) and order as they appeared on the screen”, and Troyer et al.’s (2002) participants generated “a definition or association for the name and then generated an activity for the face that was semantically related to the name”. Since Ragland et al. (2006)'s source dimension of interest was the encoding orientation (i.e., specifically whether items were processed shallowly or deeply during study), such information was highly salient at the time of encoding, and could have already been unitised with the items as part of the encoding process. In contrast, our deep encoding instructions did not direct any attention towards the items' screen location or time of presentation, and their focus on the semantic aspects of the items could have discouraged participants from unitising items and their perceptual source information.

Of greater interest was the finding that correct retrieval of source information on the location dimension did *not* correspond to faster identification RTs, as compared to when location information was not accurately retrieved. Taken together, the results of Experiments 2a and 2b are consistent with those of Persson et al. (2016). According to Persson et al.'s (2016) proposal, the presence of an identification RT difference between correct and incorrect temporal source trials in Experiments 2a and 2b supports the contribution of familiarity (and fluency) in temporal source judgements, whereas the absence of identification RT differences on location trials indicates a lack of familiarity to visuospatial source judgements. However, our findings could not provide definite evidence in favour of this conclusion, especially since identification RTs were shown to be faster to trials with correct versus incorrect source judgements on location and font size in Experiment 1.

An alternative explanation for the pattern of results in Experiments 2a and 2b could be provided by the source of activation confusion (SAC) model of memory (Reder, Donavos, & Erickson, 2002). According to the SAC model, a memory trace is used to encode the fact that a particular item was studied in a particular experimental context during the study phase. When the item is presented at test, activation will spread to the relevant item and source information stored in memory. It is easier to retrieve the encoding context when there are fewer competing contextual associations (i.e., low contextual fan conditions), as the source information has higher distinctiveness and a higher amount of the source activation would be directed towards memory for its associated encoding event. Context retrieval would be more difficult in high contextual fan conditions, as there is greater associative interference, and the memory trace linking an item with its associated source attribute would become weaker as it is saturated by the greater number of other items also associated with the same source attribute. Supporting this model, Reder et al. (2002) found that perceptual match effects (i.e., better memory for items that are presented with similar physical attributes at study and test) are enhanced for words presented in font styles not shared with other words.

In our Experiments 2a and 2b, the location source dimension had a high fan since there were only two possible screen locations, and each location was shared with half of the study items, whereas the temporal dimension could potentially have had a much lower fan since each word had occupied a unique sequential position in time. If contextual fan can affect memory for item and source information in the manner suggested by the SAC, it is possible that the higher contextual fan in the location dimension could have attenuated the relationship between fluency and source judgements.

Chapter 4: Effects of Contextual Fan on Source Memory and Fluency

4.1 Experiment 3a

The results from Experiments 2a and 2b showed faster identification RTs to trials with correct versus incorrect temporal source responses, and no difference in identification RTs between trials with correct and incorrect location responses. The objective of Experiment 3a was to test the possibility that the relationship between fluency and source memory accuracy can be affected by the contextual fan of the source information associated with studied items. According to the SAC (Reder et al., 2002) context retrieval would be more difficult in high contextual fan conditions for the following reason: the memory trace linking an item with its associated source attribute would become weaker due to increased associative interference when it is saturated by a greater number of other items also associated with the same source attribute. Whereas under a low-fan condition, the memory trace connecting an item to an associated source attribute would be relatively stronger, due to less competition between a fewer number of other items also associated with that particular source attribute. For this purpose, we manipulated the number of attributional variations within two visuospatial source dimensions: screen location and font colour.

Based on the predictions of the SAC, we expected a greater identification RT difference between source-correct and source-incorrect trials when the tested source dimension comprised more (low fan) variations in source attributes at study, in comparison to a source dimension with fewer source attribute variants (high fan).

4.1.1 Method

4.1.1.1 Participants and design

An a priori power analysis using G*Power (Version 3.1.9.2; Faul, Erdfelder, Buchner, & Lang, 2009) indicated that a minimum of 34 participants would be required to detect a significant difference ($\alpha = .05$; two-tailed) of a medium effect size ($d_z = 0.50$) at 0.80 power. In total, 40 volunteers were recruited via the UCL Psychology Subject Pool to participate in exchange for £5.00. The design had contextual fan (low vs. high) as the main within-subjects factor of interest. This was achieved through two conceptually equivalent between-subject groups. Both groups had one of the two available source dimensions presented with high contextual fan and their other source dimension presented with low contextual fan: in the *low-location-fan* group ($n = 20$; $M = 25.10$ years old, $SD = 5.09$; 14 females, 6 males), words appeared at multiple locations on the top or bottom of the screen in either a vermilion or cerulean font colour (high colour fan) at study, and the *low-colour-fan* group ($n = 20$; $M = 24.10$ years old, $SD = 3.39$; 14 females, 6 males) was presented with words displayed in a variety of blue-green and red-orange font colours at either the top-centre or bottom-centre of the screen (high location fan).

4.1.1.2 Materials

The present experiment used the same set of 134 monosyllabic English nouns as used in Experiment 2b. All stimuli and instructions were presented in a black 34-point Courier New font as in Experiments 2a and 2b. The spectrum of font colours of the low-colour-fan group were generated using RGB Color Gradient Maker (Bang, n.d.; https://www.perbang.dk/rgbgradient/), with the 32 colours of the blue-green gradient ranging from vivid cobalt blue (#004ED2) to moderate spring green (#37A86B), and the 32 red-orange colours ranging from moderate crimson (#AA062E) to vivid orange (#D66C00). The two gradients were of approximately equal saturation and luminosity. In the low-location-fan group, the vertical axes of the words presented at the upper and lower locations were all approximately 15 cm apart, but each of the top-location and bottom-location words were assigned to be presented in one of the 32 horizontal axis points which were equidistant across the width of the screen. The two font colours used in the low-location-fan group were moderate cerulean (#1A79A0) and moderate vermilion (#BF3717) which were selected from the respective midpoints of the blue-green and red-orange gradients. A microphone was used to record participants' word identification answers at test.

4.1.1.3 Procedure

The study phase instructions given to participants at the beginning of the experiment were identical to the version used in Experiment 2b. During the study phase, the 3 primacy buffers, 64 target items, and 3 recency buffers were each displayed on screen one at a time for 4 s, and between each word, a blank screen with a fixation cross at the centre was displayed for 1 s. For both contextual fan groups, primacy and recency buffers were randomly selected to appear at the top-centre or bottom-centre of the screen in either strong cerulean or vivid vermilion. Once the study phase ended, participants received instructions on the CID-R task and source memory questions which they would complete in the test phase. The test phase procedures were identical to those in Experiment 2a and 2b, the test items in the CID-R task were presented in a black font colour at the centre of the screen, and the source memory options for font colour were displayed in the same format as the temporal source options in those experiments.

4.1.2 Results

Across all participants, 182 trials (3.56% of all trials) were excluded from the subsequent analyses due to misidentification of the test word or lack of an identification response. Each participant correctly identified the test word on at least 81.25% of their trials.

4.1.2.1 Recognition memory

The item recognition hit and false alarm rates were .84 and .28 respectively across all valid trials. The overall difference between mean identification RTs to new (*M* = 1480 ms, *SD* = 328) and old (*M* = 1395 ms, *SD* = 355) items indicated a significant priming effect across participants, $t(39) = 4.23$, $p < .001$, $d = 0.67$, with 30 participants showing evidence of priming with mean identification RTs to old items being at least 10 ms faster than mean RTs to new items, eight participants showing RT differences in the opposite direction, and two participants with tied mean identification RTs for old and new items.

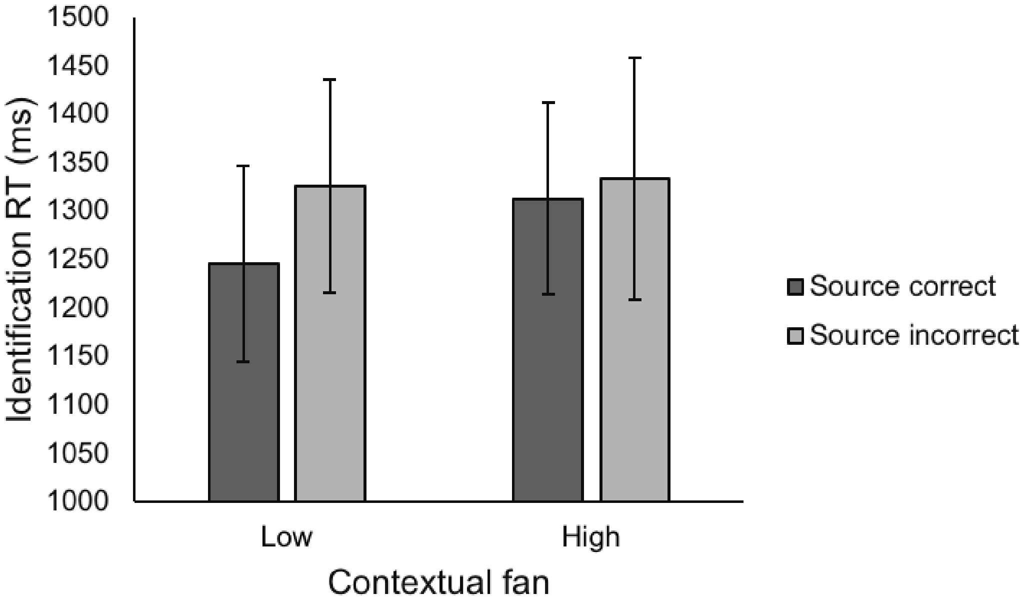

Figure 6. Mean identification RTs according to source memory accuracy in Experiment 3a. Error bars indicate 95% confidence intervals of the mean.

4.1.2.2 Source memory

For all source memory analyses, low-location-fan and low-colour-fan trials were collapsed across participants as low-contextual-fan trials, and high-location-fan and high-colour-fan trials were collapsed as high-contextual-fan trials (Figure 6). Correct source decisions occurred on 59% of the low-fan trials and on 58% of the high-fan trials. A repeated measures ANOVA was conducted to examine the effects of context fan and source accuracy on identification RTs. The ANOVA indicated a significant main effect of source accuracy, $F(1, 38) = 5.26$, $p = .0274$, $\eta_p^2 = 0.12$, but neither the context fan effect, $F(1, 37) = 2.12$, $p = .15$, $\eta_p^2 = 0.05$, nor the interaction, $F(1, 37) = 0.32$, $p = .57$, $\eta_p^2 = 0.01$, were significant. Twenty-nine participants had faster mean identification RTs to source-correct trials than to source-incorrect trials, nine showed the reverse pattern, and two had tied mean RTs (i.e., with mean RT differences of less than 10 ms).

4.1.3 Discussion

Consistent with Experiments 1 and 2, identification RTs were faster for trials with correct source responses compared to trials with incorrect source responses. Contrary to our expectations, there was no significant effect of contextual fan. However, this may have been due to an artefact related to the within-participants manipulation of contextual fan. It is possible that the contextual fan information could have been summed across word colour and location whenever any degree of unitisation between the item and its two source dimensions had occurred during encoding, thus resulting in equivalent net contextual fan across multiple trials.

4.2 Experiment 3b

Although lower contextual fan did not correspond to faster identification RTs in Experiment 3a, the results were not conclusive due to the possibility that any unitisation between item, location, and colour during encoding could have weakened the within-subjects contextual fan manipulation. To eliminate this possibility, Experiment 3b aimed to replicate the results of Experiment 3a by varying only one source dimension at the time of encoding rather than two, and by manipulating contextual fan within participants in two separate blocks.

4.2.1 Method

4.2.1.1 Participants and design

Fourty-four volunteers participated in the experiment in exchange for partial course credit or £7.50 (*M* = 22.55 years old, *SD* = 3.55; 30 females, 14 males). Each participant was randomly assigned to either the location or the colour version of the task, and there were equal numbers of participants for each version. The experiment had contextual fan as the within-participants factor of interest. In the colour version of the task, this was achieved by presenting words in different shades of blue-green and red-orange in the study phase of the low-fan block, and in one shade of blue-green and red-orange in the high-fan block. In the location version of the task, the low-fan block involved words being presented in various top and bottom locations of the screen, and at only the top-centre location or the bottom-centre screen location in the high-fan block.

4.2.1.2 Materials and procedure

A total of 204 monosyllabic English nouns were selected from the MRC Psycholinguistic Database (Coltheart, 1981). The words had 4-5 letters, a Kučera-Francis Frequency score of 1-213, Concreteness scores of 406-646, and Imageability scores of 431-647. For each participant, each of the nouns was randomly assigned to appear in either the first or the second block of the experiment, acting as one of the 48 targets presented at study, one of the 48 lures at test, one of the 3 primacy buffers, or one of the 3 recency buffers. All instructions and stimuli were presented in the same font styles, font sizes, and format as in Experiment 3a.

The blue-green and red-orange colour spectrums were generated using the same procedure as in Experiment 3a, but there were 24 instead of 32 shades in each spectrum. The screen locations of the words were generated with the vertical axis location of the word being randomly sampled from 24 equidistant points across the width of the screen, and the horizontal axis location fixed at approximately 15 cm from the screen centre for all top and bottom words. A microphone was used to record participants' spoken word identification answers at test.

The experiment session consisted of two study-test blocks: one high-fan block and one low-fan block, with the order of the blocks counterbalanced across participants. Before starting the experiment, the instructions given to participants were generally the same as in Experiment 3a, although they were additionally informed about the two-block structure of the session, and that the two blocks had no relation to each other (i.e., they can forget about all the words from the first block once they finished it, as the second block would not test them on those words). In the low-fan block, participants studied words presented in one of 24 variations of blue-green /red-orange shades (top/bottom locations), whereas words were presented in only one blue-green/red-orange shade (top-centre/bottom-centre location) in the high-fan group. The CID-R task and question format of the test phase were also identical to those of Experiment 3a. After finishing the test phase of the first block, participants completed a 2-minute word search puzzle as a filler activity before starting the second block. The study and test phases of the second block were conducted in the same procedural format as the first block.

4.2.2 Results

Across all participants, 324 trials (3.84% of all trials) were excluded from the subsequent analyses due to misidentification of the test word or lack of an identification response. Each participant correctly identified the test word on at least 70.83% of their trials.

4.2.2.1 Recognition memory

Across all valid trials, item recognition hit and false alarm rates were .77 and .28 respectively. The overall difference between mean identification RTs to new (M = 958 ms, SD = 387) and old (M = 863 ms, SD = 383) test items indicated a significant priming effect across participants, $t(43) = 8.78$, $p < .00001$, $d = 1.32$, with 39 participants showing evidence of priming with mean identification RTs to old items being at least 10 ms faster than mean RTs to new items, two participants showing RT differences in the opposite direction, and three participants with tied mean identification RTs for old and new items.

4.2.2.2 Source memory

Correct source decisions occurred on 65% of low-fan trials and 63% on high-fan trials. A repeated measures ANOVA was conducted to examine the effects of context fan and source accuracy on identification RTs (Figure 7). The ANOVA indicated a significant main effect of source accuracy, $F(1, 42) = 8.56$, $p = .0055$, $\eta_p^2 = 0.17$, but neither the context fan effect, $F(1, 42) = 0.58$, $p = .45$, $\eta_p^2 = 0.01$, nor the interaction, $F(1, 42) = 0.53$, $p = .47$, $\eta_p^2 = 0.01$, were significant. Thirty-one participants showed faster mean identification RTs to source-correct trials than to source-incorrect trials, 10 showed the reverse pattern, and three had tied mean RTs (i.e., with mean RT differences of less than 10 ms).

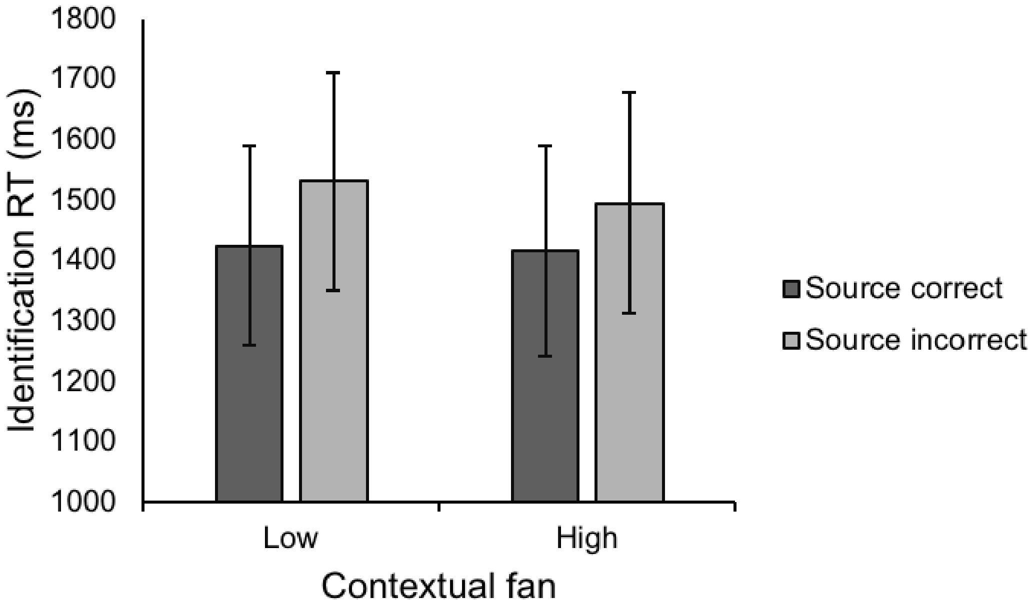

Figure 7. Mean identification RTs according to source memory accuracy in Experiment 3b. Error bars indicate 95% confidence intervals of the mean.

4.2.3 Discussion

In line with the previous experiments in this and the previous chapters, trials with correct source responses tended to have faster item identification RTs than trials with incorrect source responses. Replicating the findings from Experiment 3a, the current results show no evidence that contextual fan directly affects identification RTs or moderates the relationship between identification RTs and source memory accuracy, even with a within-subjects manipulation of contextual fan.

4.3 Summary and Discussion of Experiments 3a and 3b

Neither Experiment 3a nor 3b suggested that contextual fan affects identification RTs or the relationship between identification RTs and source accuracy. One possibility is that, even though participants might have perceived words as having different shades of font colours and occupying distinct locations on the screen, the colours and locations could still have been amalgamated as “greenish colours” or “top of the screen location” during the encoding process. If that had been the case, then future source memory studies manipulating contextual fan may benefit from using even more categorically distinct source attribute variations.

Although Reder et al. (2002) found that low fan encoding conditions can further enhance perceptual match effects, it is possible that, as a factor on its own, contextual fan might not have as much of a direct influence on item-source recognition. In Buchler, Light, and Reder’s (2008) study on memory for paired associates, their high fan manipulation adversely affected the retrieval of associations but not of items. Since identification RTs were made at the item level in Experiments 2a and 2b, it is possible that the SAC’s predictions did not apply here.

It would also warrant further investigation to determine the extent that the 64 locations or colours were, in reality, perceived as 64 functionally distinct source attributes by participants. Although much care was put into selecting as diverse a range of locations and colours as possible, Miller (1956) has suggested that people are able to make accurate absolute identifications for an approximate maximum of 5-7 equally-spaced stimuli along a perceptual continuum. Experiments 2a and 2b would then only have had 10-14 effective source variants in the low-fan conditions (i.e., 5-7 perceivable variations in each of the upper and lower, or blue-green and red-orange, sources) versus two source variants in the high-fan conditions, as opposed to 64 versus two, and could thus have rendered the fan manipulations considerably weaker.

Another explanation for the lack of a fan effect on identification RTs or source accuracy could be due the difference in the direction of fan manipulations of the source attributes (colour/location) and source memory cues (word items) used in this study, as compared to Reder et al. (2002). In our Experiments 2a and 2b, experiment-wide contextual fan was manipulated by arranging more or fewer source variants whilst keeping the number of word items constant. In contrast, Reder et al. (2002) manipulated contextual fan by varying the number of words associated with each individual font style. Although both types of manipulation ultimately alter contextual fan in terms of the experiment-wide source-to-cue ratio, there have been no studies directly comparing the effects of these two approaches, to the best of our knowledge.

Chapter 5: Processing Fluency and Source Memory Accuracy: A Meta-Analysis

5.1 Meta-Analysis of Experiments 1-3b

The results across Experiments 1-3b suggested that fluency (identification RTs) is related to source memory accuracy under certain conditions. Experiment 1 showed faster identification RTs to trials with both source dimensions (font size and screen location) correctly identified versus trials with no correctly identified source dimensions. Experiment 2a found faster identification RTs to source-correct trials than source-incorrect trials under the shallow encoding condition, but only when the source attributes were temporal, and no such identification RT differences were shown when the source attributes were visuospatial (i.e., screen location). These results were replicated in Experiment 2b. However, visuospatial source attributes were used again in Experiments 3a and 3b (by collapsing the font colour and screen location dimensions in the analyses), and the results of both showed faster identification RTs to source-correct trials than source-incorrect trials.

Using the metafor package (Viechtbauer, 2010) in R, a meta-analysis with a random-effects model was conducted to examine the overall relationship between identification RTs and source memory accuracy. Across the aforementioned instances in Experiments 1-3b where the source memory accuracy effects on identification RTs were found, the meta-analytic effect size was $d_z = 0.46$, 95% CI = [0.26, 0.66], a medium-sized effect. The heterogeneity between the experiments was non-significant, $Q(4) = 2.50$, $p = .64$.

5.2 Discussion of Experiments 1-3b

The finding of the meta-analysis is broadly consistent with Kelley et al.'s (1989) results, which showed a dependence between perceptual identification at test with source modality judgements. It is also consistent with Kurilla's (2011) findings that participants' tendency to report that an item had been studied in the same source format is influenced by artificially manipulating perceptual processing fluency. The present results additionally suggested that fluency is related to the accuracy of the source judgements. Furthermore, our participants had no access to the original source attributes at test, as all test items were presented in a neutral colour, font, or location (unlike in Kelley et al.'s and Kurilla's studies where participants judged whether or not the source format of the test item matched its original source at study). Consequently, our participants would have had no access to familiarity via the perceptual match between the source attribute(s) of study and test items. This would have made the test even more dependent on recollection, yet the identification RT results imply that recollective processes were not the only contributors to test performance. It is unknown to what extent participants were able to access their stored representations of the item and source information during the stimulus demasking phase of the CID-R task at test, and this could be an avenue for future research.

The relationship between item-level identification RTs and source accuracy and confidence challenges versions of dual-system models that specifically assume complete independence between the bases of implicit and explicit memory (e.g., priming and recognition, respectively; Tulving, Schacter, & Stark, 1982), given that recollection and the retrieval of source information are conventionally considered to be even more reliant on explicit processes than item recognition. This finding is compatible, in contrast, with single-system models, although it may not necessarily contradict dual-system models which allow for implicit and explicit memory to be correlated (i.e., the MS2 model; Berry et al., 2012). For example, when attention fluctuates during encoding, some items may accordingly be better encoded than others, resulting in both faster identification RTs and more accurate source retrieval for those items at test. Thus, dual-system models which include a free correlation parameter between implicit and explicit memory strengths would allow for fluency and source memory performance to be partially related even if the two measures rely on separate systems. A zero-correlation is implied by dual-system models that assume complete independence implicit and explicit bases, whereas a correlation value of 1 is implied by single-system models (i.e., both bases of memory strength signals can be represented as a single base).

In Berry et al.’s (2012) study on fluency and recognition memory, the best fitting correlation parameter estimate was found to be .93 after applying the MS2 model to data across their three experiments. This near-maximal correlation suggested that empirically, the MS2 model’s performance was virtually identical to that of the single-system model, and that any influence of correlations across items at encoding (e.g., as a result of attentional fluctuation) is minimal. Future studies might directly manipulate attention or distinctiveness during encoding in order to examine their influences on fluency, source accuracy, and the correlation between them.

On the basis of the definition for source memory, having access to recollective details is commonly assumed to be crucial to determining the success or failure of source information retrieval (e.g., Guttentag & Carroll, 1997; Perfect et al., 1996; Yonelinas, 1999). Nonetheless, the present findings are in line with the view that familiarity-related processes and other types of information can also support source memory (e.g., Hicks et al., 2002; Johnson et al., 1993). Although there is still ongoing research on the circumstances that would enable or promote the contribution of familiarity to source memory, Yonelinas (2001) proposed this contribution can occur especially when item and source information are encoded as a single unit. More recent studies have demonstrated an important role of item-source unitisation in moderating familiarity's contribution to source memory (e.g., Diana, Yonelinas, & Raganath, 2008). For their study phase, Diana et al. instructed participants in the high-unitisation condition to imagine items as being in their corresponding background colours, and participants in the low-unitisation condition to imagine items associated with background-coloured objects (e.g., if the background is red, associate the item with a red stop sign). Based on converging behavioural and ERP results, the study suggested that unitisation instructions increase familiarity’s connection to source memory.

It is possible that the items and source attributes in the present experiments were unitised to some degree at the time of encoding. Yet it is also possible that very little unitisation had occurred given that most of the item-source associations would have been rather arbitrary (e.g., the word “torch” presented in a green font colour), and that the presentation duration might not have been sufficient to allow for participants to generate unitised images themselves. According to the Levels of Unitisation Framework (Parks & Yonelinas, 2015), there can be lower (e.g., “this torch was found in a green field”) versus higher (e.g., “this torch emits green light”) degrees of unitisation. Despite the former being a unitised image it still consists of an arbitrary association between two separate entities, whereas the latter forms a single entity in its own right. Whether or not unitisation is necessary in order to observe a relationship between fluency and source memory still needs to be further investigated, as well as the level of unitisation that would have been required for this relationship to occur.

To conclude, the present findings established fluency as an important contributor to memory for source information at least on some dimensions. The exact nature of the relationship between fluency and source memory awaits additional study, as does the contribution of fluency to source monitoring in more ecological settings, but the present findings are amongst the first to reveal directly that familiarity-based processes are linked to source memory accuracy. Given the vital part source information plays in our social interactions and episodic remembering, research should continue to elucidate the mechanisms underlying source memory.

Chapter 6: Implications of Source Information and Processing Fluency on Metamemory: The Font Size Effect

Metamemory, a subtype of metacognition, refers to an individual's self-awareness and self-monitoring of their own memory processes. Due to the fallibility of human memory, metamemory is necessary for auditing and controlling learning and other related behaviours, according to the extent to which a given piece of information is available or accessible in one's memory (e.g., Tulving & Pearlstone, 1966). Much the same as source memory, metamemory also plays a crucial role in informing decisions and actions. For example, based on metamnemonic judgements, a strategic decision can be made on whether to attempt to retrieve an answer from memory or to instead formulate an answer through reasoning (Miner & Reder, 1994). Quite similar to source memory, there has also been considerable interest in the past two decades concerning how (and whether) aspects of a target stimulus, the target's context, or the learner's own mental processes can become determinants of metamemory decisions (Schwartz, 1994).

Broadly speaking, metamnemonic monitoring includes both prospective and retrospective metamemory judgements. According to this broad definition of metamemory, even the kinds of R/K judgements, source monitoring, and confidence judgements used in Experiments 1-3b would fall under the scope of retrospective metamemory (Johnson et al., 1993; Kelley & Lindsay, 1993; Rajaram, 1993). However, the classic focus of metamemory research has been on prospective metamnemonic monitoring, which is concerned with predicting the memorability of information retrieved at a future time (Nelson & Narens, 1990). Judgements of learning (JOLs) made during the encoding of information represent only one of the many types of prospective metamemory measures, and there are also several varieties of JOLs, including ease-of-learning judgements (Underwood, 1966), paired-associate JOLs (Arbuckle & Cuddy, 1969), recognition JOLs (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989), and free-recall JOLs (Groninger, 1979). Experiments 4-6 in the following chapters will concentrate solely on JOLs for free-recall, also known as *memorability ratings* (Mazzoni, Cornoldi, & Marchitelli, 1990) and *the “feeling-that-I-will-know” phenomenon* (Groninger, 1979).

As mentioned in the General Introduction of this thesis, the font size effect on free-recall JOLs has been revealed in studies where participants tend to assign higher JOLs to larger font words than smaller font words even though font size does not actually affect memorability (e.g., Rhodes & Castel, 2008), and the two main classes of theories proposed for this effect are belief-based theories (e.g., Castel, 2007; Hu et al., 2015; Mueller et al., 2014; Mueller & Dunlosky, 2017; Rhodes & Castel, 2008) and fluency-based (e.g., Besken & Mulligan, 2013; Koriat & Bjork, 2006; Koriat & Ma'ayan, 2005; Magreehan et al., 2016; Undorf, et al., 2017) theories.

To reiterate, few studies to date have directly assessed the role of processing fluency in the font size effect on JOLs. One of them (Mueller et al., 2014, Experiment 1) employed small and large word and non-word items in a lexical decision task to test whether lexical decision RTs underlie participants’ JOLs for each item. On the basis of the lack of difference in processing fluency between small and large items (as derived from lexical decision RTs), the authors suggested that fluency does not mediate the font size effect.

Prior to Mueller et al.'s (2014) study, the general consensus in the literature was that perceptual fluency does underlie the font size effect on JOLs, and indeed many researchers had offered the font size effect on JOLs as evidence that perceptual fluency can affect JOLs (e.g., Bjork, Dunlosky, & Kornell, 2013; Diemand-Yauman, Oppenheimer, & Vaughan, 2011; Kornell et al., 2011; Miele et al., 2011; Rhodes & Castel, 2008). It is important to note that Muller et al. (2014) did not completely reject the fluency theory. Instead, they suggested that their results were inconsistent with the fluency theory and they encouraged future research to further explore the fluency theory (p. 9). However, after the publication of Mueller et al. (2014), the consensus began to shift towards the view that fluency may play no role in the font size effect on JOLs (e.g., Ball et al., 2014; Finn & Tauber, 2015; P. Li, Jia, Li, & Li, 2016; Magreehan et al., 2016; Mueller & Dunlosky, 2017; Mueller, Dunlosky, & Tauber, 2016; Susser, Jin, & Mulligan, 2016; Susser, Panitz, Buchin, & Mulligan, 2017; Undorf et al., 2017). Taking a more neutral position, Hu et al. (2015) claimed that, "Although Mueller et al. (2014) suggest that fluency does not differ... There may be other types of fluency that differ significantly between large and small words" (p. 10).

There are at least three possible reasons for the lack of a difference in RTs between large and small words in Mueller et al.'s (2014) Experiment 1. The first, as proposed by Mueller et al. (2014), is that there is truly no difference in perceptual fluency between large and small words. Secondly, their null result might be a false negative, because the number of trials (18 large and 18 small words) and sample size (31 participants) might have combined to render their experiment underpowered. It is well-known that small sample size and number of trials can lead to false negative results (Vadillo, Konstantinidis, & Shanks, 2016). The third possibility concerns the research method Mueller et al. employed, specifically, their use of RTs obtained from a lexical decision task as an index of perceptual fluency. The lexical decision task is complex (Yap, Sibley, Balota, Ratcliff, & Rueckl, 2015): Participants need to read or identify the letter string first, judge whether it is a word or a non- word, and then select which button to press to indicate their response before the judgement RT is recorded. Participants may check the letter string letter-by-letter, and their lexical decisions may be conservative and time-consuming. Therefore, there could be considerable noise in the RTs obtained from the lexical decision task. Access to word meaning is also assumed to be involved in the lexical decision task (Chumbley & Balota, 1984). Consequently, RTs derived from Mueller et al.'s (2014) Experiment 1 might be driven by semantic processing *in addition* to perceptual processing of the words, and thus it is unclear to what extent their findings contradict accounts claiming that perceptual fluency underlies the font size effect on JOLs. In short, lexical decision may be a poor tool for measuring variations in perceptual fluency.

Mueller et al. (2014) tested the fluency theory more indirectly by measuring study time allocation in their Experiment 2. Participants were allowed to spend as much time as they wanted to study each word. Mueller et al. (2014) hypothesised that participants would spend less time studying large compared to small words if large words are processed more fluently than small words. However, they observed no difference between study times allocated to large and small words, and proposed that “the lack of an effect of font size on study time allocation is inconsistent with the hypothesis that encoding fluency is responsible for the font-size effect on JOLs” (p. 5).

Yet again, this result does not provide strong motivation to reject the fluency theory because, besides fluency, many other factors could have affected participants’ study time allocation (e.g., motivation, curiosity). Participants might believe that large words are more important than small words (Mueller et al., 2014; Rhodes & Castel, 2008), and allocate more time to them accordingly (Noh, Yan, Vendetti, Castel, & Bjork, 2014). A fluency advantage for large words (leading them to be studied for less time) may have operated in opposition to a belief that large words are important (leading them to be studied for longer), thus contributing to the overall null result. Yang, Potts, and Shanks (2017a) found that participants decreased their study times across a study phase when they were allowed to spend as much time as they wanted to study each item (e.g., Euskara-English word pairs in Yang et al.’s Experiment 1 and face-name pairs in their Experiment 2), again implying that self-regulated study time allocation can be affected by other factors besides fluency.

Moreover, recent research has found that self-regulated study time allocation is not a sensitive measure of fluency in certain situations. For example, Witherby and Tauber (2017) found that participants responded faster to concrete (e.g., apple) than to abstract (e.g., idea) words in a lexical decision task, but there was no difference in study times between concrete and abstract words when participants were allowed to spend as much time as they wanted to study them. Therefore, Mueller et al.’s (2014) Experiment 2 cannot be taken as providing indirect evidence against the fluency theory because self-regulated study time allocation can be affected by many other factors besides fluency, and is an insensitive measure of fluency. Overall, Mueller et al.’s (2014) Experiments 1 and 2 fall short of providing compelling evidence against the fluency theory and it remains unclear whether perceptual fluency contributes to the font size effect on JOLs.

After Mueller et al.’s (2014) study, researchers raised two other important questions. The first question is whether – moving beyond the standard font size manipulation – there exists evidence that perceptual fluency can affect JOLs (e.g., Besken, 2016; Frank & Kuhlmann, 2016; Price & Harrison, 2017; Susser et al., 2016; Undorf et al., 2017). Susser et al. (2016) addressed this question by employing an identity-priming paradigm. Participants were asked to name and make item-by-item JOLs for words (e.g., phone) which were preceded by either matched (phone) or mismatched (e.g., doctor) primes. Susser and colleagues found that matched priming produces greater perceptual fluency than mismatched priming, as reflected by a difference in naming latencies. They also found that higher JOLs were given to matched words than to mismatched words – a priming effect on JOLs. But a mediation analysis revealed that naming latencies did not mediate the priming effect on JOLs. Thus Susser and colleagues concluded (p. 660) that “effects of perceptual fluency on JOLs do not exist.”

On the other hand, Undorf et al.’s (2017) results contradicted Susser et al.’s (2016) conclusion. Undorf et al. (2017) instructed participants to identify stimuli (objects, faces, or words in different experiments) and make item-by-item JOLs. For each stimulus, 30 images were created in which the object became progressively larger and larger: Image size increased with image number. In the slow clarification condition, images were presented for 1 s each, in the following number sequence: 1, 2, 3 ….30; in the fast condition the images were presented in the sequence: 1, 3, 5….29. Thus the maximum image size occurred after 15 image presentations in the fast condition and after 30 images in the slow condition. The results showed that stimuli were identified faster in the fast condition than in the slow condition, and the size level at which a stimulus was identified was larger in the fast condition than in the slow condition. The results also showed that higher JOLs were given to stimuli in the fast condition than in the slow condition – a clarification speed effect on JOLs. Most importantly, Undorf et al. (2017) found that identification RTs significantly mediated the clarification speed effect on JOLs (for similar findings, see Besken, 2016). Evidently, Undorf et al.'s (2017) and Susser et al.'s (2016) results support mutually conflicting conclusions. Therefore, it is still controversial whether perceptual fluency can affect JOLs and more research is needed to explore this question.

The second question is whether perceptual fluency underlies the stimulus size effect on JOLs. For example, after Mueller et al.'s study, Undorf et al. (2017) noted that "there is no evidence that perceptual fluency contributes to the stimulus size effect on JOLs" (p. 294), and they further investigated this question by manipulating stimulus clarification speed. Nonetheless, Undorf et al.'s (2017) study cannot provide direct evidence that perceptual fluency underlies the stimulus size effect on JOLs because it manipulated the rate of change in the sizes of their stimuli, rather than directly manipulating the stimulus size. All stimuli in their study had the same (dynamically-changing) size, except that the identified size was determined by the participants’ response. For example, on a slowly-identified trial, the stimulus size displayed on screen would be larger *at the moment* of identification relative to the stimulus size displayed on screen if the participant could identify the stimulus more rapidly. This means that the relationship between identification RTs and JOLs is confounded by the different levels of stimulus size at which the words were identified across the two clarification conditions.

Undorf et al. suggested that the greater JOLs in the fast clarification condition relative to the slow condition could be mediated by greater perceptual fluency (i.e., shorter RTs). However, since stimulus identifications tended to be made at a larger size in the fast condition than in the slow condition, an alternative explanation for the aforementioned finding is that the higher JOLs observed in the fast condition occurred as a direct consequence of their larger stimulus size at identification. Similarly in the slow condition, for a given trial with a fast identification RT, stimulus size would have been smaller at the moment of identification compared to the size corresponding to the same RT if the trial had been in the fast condition. Direct evidence should demonstrate that a large (versus small) stimulus size, which is processed with greater perceptual fluency, produces higher JOLs, and that perceptual fluency mediates that stimulus size effect on JOLs. This demands an explicit experimental manipulation of stimulus size – something which was not part of Undorf et al.’s method. Therefore, despite Undorf et al.'s (2007) demonstration of perceptual fluency contributing to the effect of stimulus enlargement speed on JOLs, there is still no direct evidence that perceptual fluency underlies the stimulus size effect on JOLs when stimulus sizes are pre-determined and stationary.

To summarise, lexical decision and self-regulated study time allocation are the two most widely-used methods to measure fluency in metamemory research (e.g., Ball et al., 2014; Jia et al., 2015; Mueller et al., 2016; Mueller et al., 2014; Mueller et al., 2013; Undorf & Erdfelder, 2014; Witherby & Tauber, 2017). By employing these two methods, Mueller et al. (2014) found no difference in fluency between large and small words. However, as discussed, the null outcomes could have been produced by alternative factors. Following Mueller et al.'s study, researchers examined whether perceptual fluency can affect JOLs. By employing different experimental methods and types of stimuli, Undorf et al. (2017) and Susser et al. (2016) observed different results supporting mutually conflicting conclusions. Undorf et al. (2017) investigated whether perceptual fluency underlies the stimulus size effect on JOLs by manipulating stimulus classification speed, but their study cannot provide conclusive evidence because they did not experimentally manipulate processing fluency independently of stimulus size at the point of classification.

The main aim of Experiments 4-6 is to further test whether perceptual fluency underlies the font size effect on JOLs by employing a CID task, a variety of perceptual identification task (Sanborn, Malmberg, & Shiffrin, 2004). The task has frequently been used in memory (e.g., repetition priming) research (e.g., Berry, Shanks, Speekenbrink, & Henson, 2012; Stark & McClelland, 2000; Ward, Berry, & Shanks, 2013), but to the best of

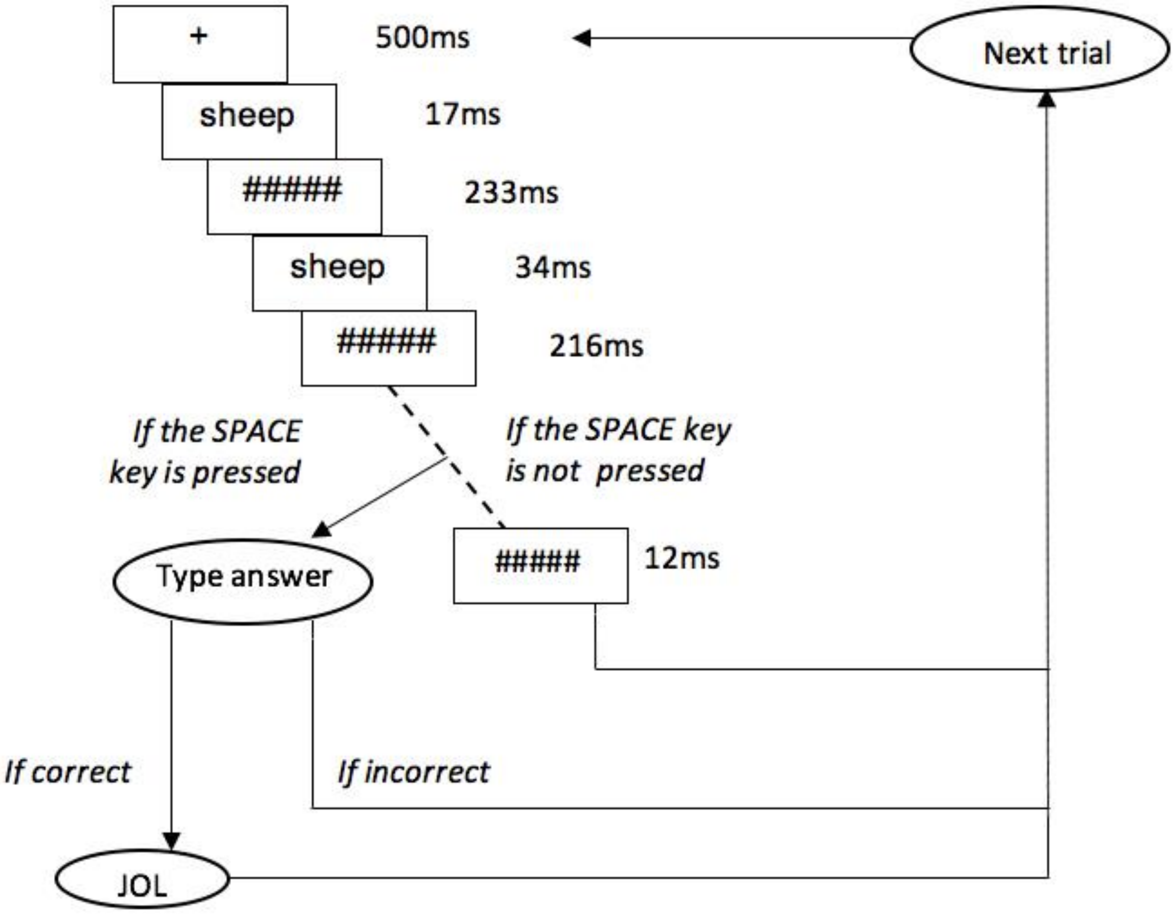

Figure 8. Experimental design schema of Experiment 4's study phase. For each identification trial, a word and a mask were alternatively presented in the same font size, which was randomly decided by the computer. Participants' task was to identify each word as quickly and as accurately as they could by pressing the space bar.

our knowledge, no previous prospective metamemory research has employed the CID task to measure fluency. It is important to re-emphasise here, that although the CID task used in Experiments 4-6 is nearly identical to the CID-R task used in Experiments 1-3b, fluency is recorded through the CID task during encoding instead of during retrieval, and the test is of free-recall rather than recognition.

In the CID task of Experiments 4-6, a word and a mask are alternately presented, with the presentation time of the word increasing and the presentation time of the mask decreasing in each fixed-duration cycle (see

Figure 8). Across cycles, the word gradually becomes clearer and easier to perceive as the stimulus-to-mask ratio increases via progressive demasking. Participants’ task is to identify the presented word as quickly and accurately as possible, and their identification RT is used as an index of fluency. On the basis of prior research (Ferrand et al., 2011; Grainger & Segui, 1990), we anticipated that the CID task would be more sensitive than lexical decision to variations in perceptual fluency. By employing the CID task, we tested for any difference in perceptual fluency between large and small words, and explored whether perceptual fluency mediates the font size effect on JOLs. The presence of a difference in fluency between the font sizes in addition to a mediating effect of fluency on JOLs will support the fluency theory as an account for the font size effect on JOLs. At the same time, through directly manipulating font size, the current research will determine whether or not perceptual fluency underlies the stimulus size effect on JOLs.

6.1 Experiment 4

Experiment 4 involved the use of the CID task to investigate whether perceptual fluency underlies the font size effect on JOLs. As discussed, the small number of trials in Mueller et al.’s (2014) Experiment 1 might have contributed to their null result. We therefore increased the number of trials to 100.

6.1.1 Method

6.1.1.1: Participants

The required sample size was determined *a priori* through a power analysis conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

Based on the effect sizes from previous studies in which Cohen's *d*s ranged from 0.58 to 0.74 (Hu et al., 2016; Rhodes & Castel, 2008), 22-34 participants are required to observe a significant (α = .05) font size effect on JOLs at 0.9 power. Therefore, we recruited 28 participants[5] (*M* = 22.21 years old, *SD* = 7.10; 21 females, 7 males) from the UCL Psychology Subject Pool. All reported normal or corrected-to-normal vision, received a £3 cash payment or partial course credit as compensation, and were tested individually in cubicles.

6.1.1.2: Design

The experiment had font size (large vs. small) as the within-subjects factor. To prevent any potential item effects, the program randomly selected half the words to be presented in large and the other half in small font sizes for each participant, and the presentation sequence of words was also randomly determined.

6.1.1.2: Materials

The principal stimuli were 110 monosyllabic English nouns selected from the MRC Psycholinguistic Database (Coltheart, 2007), each with 5 letters, a Kučera-Francis Frequency score of 3-50, a Concreteness score of 300-670, and an Imageability score of 300-600. We strictly controlled the letter length to 5 in order to ensure that the mask (#####) would completely cover each word. Ten words were used for practice and the other 100 were

---

[5] This sample-size estimation is conservative. Morey (2016) showed that effect sizes change with varying numbers of experimental trials, because a larger number of trials yields a smaller mean squared error (MSE) and hence a greater effect size. As we have increased the number of trials compared to previous studies, we expect to observe a greater effect size. Thus the power to detect a significant font size effect on JOLs is expected to be greater than specified.

used in the main experiment. Stimuli were displayed on an LCD monitor (1920 × 1080 resolution at 60 Hz) via the MATLAB *Psychtoolbox* package (Kleiner, Brainard, & Pelli, 2007).

6.1.1.3: Procedure

The experiment consisted of three components: study, distraction, and test. In the study phase, a fixation cross was presented at the centre of the screen in a medium font size (30-point) for 500 ms. Then a word and a mask were alternately presented in the same Arial font style and the same font sizes (48 or 18-point) as in Mueller et al. (2014). For each identification trial, there were 14 cycles in total. At the first cycle, the word was presented for 17 ms followed by the mask for 233 ms. At the second cycle, the word was presented for 34 ms, followed by the mask for 216 ms. Thus across cycles, the presentation duration of the word increased in 17 ms steps with the duration of the mask decreasing in 17 ms steps. The word-mask cycle was repeated until participants responded or until the end of the 14th cycle. Participants were instructed to press the space bar as soon as they could identify the word. If they did not respond before the end of the 14th cycle, the next identification trial began. If they responded, the word and mask disappeared, and participants typed in their answer (the word) via the keyboard. Then the computer automatically checked whether or not their answer was correct. If correct, a slider ranging from 0 (“I’m sure I’ll not remember it”) to 100 (“I’ll definitely remember it”) was presented at the centre of the screen for participants to predict the likelihood that they would remember that word at a later test. If incorrect, the next trial began (see experiment design schema of the study phase in Figure 8). After participants identified all 100 words, they were asked to solve as many arithmetic problems (e.g., 24+32 = ___?) as they could in 2 minutes. Then, they were instructed to recall as many words as possible in any order and to type their answers. Their answers were shown on screen in a medium-sized font (30-point).

All experimental instructions were presented in the medium font size. Participants were told to place their left hand above the space bar while they used the mouse to make JOLs, which enabled them to press the space bar as soon as they could identify the word. They were allowed to freely adjust their distance from the monitor.

6.1.2: Results

Table 2 reports participants' identification accuracy which was similar for large and small words, M_{diff} = -1.1%, 95% CI = [-3.3%, 1.0%], $t(27) = 1.08$, $p = .29$, $d = 0.20$. All data from incorrectly identified trials were removed from the subsequent analyses.

Participants' recall accuracy for large and small words was calculated using the formula:

$$\text{Recall accuracy} = \frac{\text{Number of words correctly recalled}}{\text{Number of words correctly identified}} \times 100\%$$

Consistent with previous studies, we found no difference in recall accuracy between large and small words, difference = 0.9%, 95% CI = [-3.3%, 5.1%], $t(27) = 0.44$, $p = .66$, $d = 0.08$ (see the right pair of bars in Figure 9A). In contrast participants gave significantly higher JOLs to large ($M = 51.56$, $SD = 14.90$) than to small words ($M = 47.50$, $SD = 14.63$), $M_{\text{diff}} = 4.05$,

	Large	Small
Experiment 4	93.4% (6.6%)	94.6% (3.5%)
Experiment 5		
CID	94.9% (4.4%)	94.4% (5.3%)
Lexical Word	97.2% (3.8%)	96.8% (5.0%)
Lexical Non-word	89.4% (14.3%)	92.6% (7.2%)
Experiment 6	94.3% (5.8%)	93.2% (6.3%)

*Table 2***.** *M* (*SD*) identification and judgment accuracy in Experiments 4-6

95% CI = [1.98, 6.13], $t(27) = 4.01$, $p < .001$, $d = 0.76$ (see the left pair of bars in Figure 9A), reflecting a font size effect on JOLs.

The key data concern the measure of perceptual fluency. As can be seen in Figure 9B, participants' median identification RTs were significantly shorter for large ($M = 1.19$ s, $SD = 0.34$) than for small ($M = 1.44$ s, $SD = 0.30$) words, $M_{\text{diff}} = -0.25$ s, 95% CI = [-0.33, -0.17], $t(27) = -6.60$, $p < .001$, $d = -1.25$. Twenty-seven participants responded faster to large than to small words while only one showed the reverse pattern, $\chi^2(1) = 24.14$, $p < .001$. This is a very substantial effect of font size on perceptual fluency, as measured via the CID task.

To explore the statistical relationship between identification RTs and JOLs, we conducted a multilevel regression analysis using the R *lme4* package (Bates, Mächler, Bolker, & Walker, 2015), with RTs as the independent variable and JOLs as the dependent variable. The results showed that the fixed effect of RTs on JOLs was -4.35,

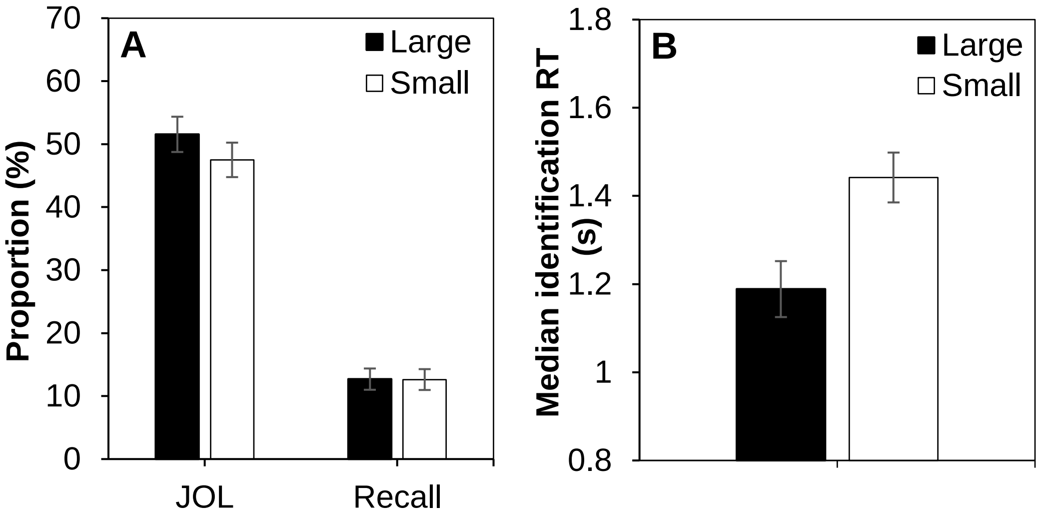

Figure 9. Experiment 4. Panel A: Judgments of learning (JOLs) and recall for large and small words. Panel B: Median identification RTs for large and small words. Error bars represent ± 1 standard error.

95% CI = [-6.64, -2.11], indicating that every decrease of 1 s in identification RTs increases JOLs by 4.35 points on the 100-point scale. These results revealed a fluency effect on JOLs – that is, the faster a word is identified, the higher the JOL it is given.

To directly test the fluency theory, we explored whether identification RTs mediate the font size effect on JOLs using a multilevel mediation analysis method with the *bmlm* package (Vuorre, 2017) in R. The package provides a Bayesian estimation of multilevel mediation models (Vuorre, 2017) and the mediation effect was estimated with 4 Markov Chain Monte Carlo (MCMC) chains and 10,000 iterations for each chain. In this multilevel mediation analysis, we took font size (small = 0; large = 1) as the independent variable, identification RTs as a mediator, and JOLs as the dependent variable. Table 3 reports the mediation results.

*Table 3***.** Multilevel mediation analysis results in Experiments 4 and 6.

	b	*SE*	95% CI
Experiment 4: Font size-RTs-JOL			
Effect of font size on RTs	-0.21	0.03	[-0.28, -0.15]
Effect of RTs on JOLs	-3.70	1.06	[-5.87, -1.69]
Total effect of font size on JOLs	4.11	1.02	[2.10, 6.12]
Direct effect of font size on JOLs	3.27	0.99	[1.34, 5.22]
Indirect effect of font size on JOLs through RTs	0.84	0.30	[0.31, 1.50]
Proportion of the total effect of font size on JOLs mediated by RTs	21%	15%	[8%, 42%]
Experiment 6: Font size-RTs-sJOLs			
Effect of font size on RTs	-0.20	0.04	[-0.28, -0.13]
Effect of RTs on sJOLs	-2.81	0.65	[-4.09, -1.51]

Total effect of font size on sJOLs	4.30	0.91	[2.51, 6.09]
Direct effect of font size on sJOLs	3.69	0.91	[1.88, 5.49]
Indirect effect of font size on sJOLs through RTs	0.60	0.18	[0.30, 0.99]
Proportion of the total effect of font size on sJOLs mediated by RTs	15%	6%	[7%, 28%]
Experiment 6: RTs-oJOLs-sJOLs			
Effect of RTs on oJOLs	1.40	2.92	[-4.29, 7.17]
Effect of oJOLs on sJOLs	0.07	0.02	[0.02, 0.11]
Total effect of RTs on sJOLs	-3.33	0.7	[-4.69, -1.98]
Direct effect of RTs on sJOLs	-3.21	0.66	[-4.49, -1.91]
Indirect effect of RTs on sJOLs through oJOLs	-0.12	0.38	[-0.94, 0.58]
Proportion of the total effect of RTs on sJOLs mediated by oJOLs	3%	12%	[-22%, 26%]
Experiment 6: Font size-oJOLs-sJOLs			
Effect of font size on oJOLs	17.23	3.02	[11.31, 23.24]

Effect of oJOLs on sJOLs	0.02	0.03	[-0.04, 0.07]
Total effect of font size on sJOLs	4.30	1.01	[2.32, 6.31]
Direct effect of font size on sJOLs	3.91	1.05	[1.88, 6.00]
Indirect effect of font size on sJOLs through oJOLs	0.39	0.57	[-0.65, 1.63]
Proportion of the total effect of font size on sJOLs mediated by oJOLs	9%	15%	[-17%, 38%]
Experiment 6: Font size-(RTs, oJOLs)-sJOLs			
Indirect effect of font size on sJOLs though RTs	0.62	0.17	[0.29, 0.95]
Indirect effect of font size on sJOLs though oJOLs	0.64	0.44	[-0.22, 1.50]
Difference between the indirect effect through RTs and that through oJOLs	-0.02	0.45	[-0.90, 0.86]

Note: JOL= judgment of learning; sJOL = study phase judgment of learning; oJOL = observation phase judgment of learning.

The total effect of font size on JOLs was 4.11, 95% CI = [2.10, 6.12]. The indirect effect of font size on JOLs through RTs was 0.84, 95% CI = [0.31, 1.50], indicating that large fonts increase JOLs indirectly by increasing perceptual fluency. Fluency (RTs) explained 21%, 95% CI = [8%, 42%], of the font size effect on JOLs. The direct effect of font size on JOLs was 3.27, 95% CI = [1.34, 5.22], indicating that fluency did not explain all of the font size effect on JOLs: The direct effect of font size on JOLs was still significant when identification RTs were controlled.

6.1.3: Discussion

Perceptual fluency differs between large and small words, as reflected by the significant difference in identification RTs on the CID task. The faster a word is identified, the higher the JOL given to that word, as revealed by the inverse relationship between RTs and JOLs. Most importantly, perceptual fluency contributes to the font size effect on JOLs, as shown by the significant mediation results. In sum, these results demonstrate that perceptual fluency can affect JOLs and provide direct evidence that perceptual fluency underlies (at least in part) the stimulus size effect on JOLs.

6.2: Experiment 5

As previously discussed, the null result observed in Mueller et al.’s (2014) Experiment 1 might be due to a range of factors. In Experiment 5, we directly compared the lexical decision and CID tasks in the same participants, with the same number of trials and the same materials, to explore whether the CID task is more sensitive to variations in perceptual fluency than the lexical decision task.

6.2.1: Method

6.2.1.1: Participants

Twelve volunteers from the UCL Psychology Subject Pool participated in exchange for partial course credit or £2, all of whom reported normal or corrected-to-normal vision (*M* = 21.67 years old, *SD* = 3.17; 8 females 4 males).

With regards to the *N*=12 sample size we have planned for this experiment, we have additionally taken into account the number of trials per participant, as effect sizes will change as a function of this (Morey, 2016). This is important because in Experiment 5 we have decreased the number of trials to 36 compared to 100 in Experiment 4. In order to determine the required sample size for Experiment 5, we re-analyzed the RT data from Experiment 4. In Experiment 4, participants successfully identified about 94% of words, therefore we expected that participants in Experiment 5 would each successfully identify about 17 (94% × 18) large and small words. Based on this estimate, we calculated the median RTs for the first 17 large and small words which were correctly identified by each participant in Experiment 4.

Then we conducted a paired-sample *t* test, which showed that participants responded faster to large than to small words on these restricted sets, difference = -0.31 s, 95% CI = [-0.43, -0.19], *d* = 1.04. Consistent with Morey’s analysis, this is appreciably smaller than the effect size (*d* = 1.25) computed across all trials. Using this effect size, we therefore determined that Experiment 5 requires about 12-13 participants to detect a significant (α = .05) difference in RTs between large and small words in the CID task at 0.90 power.

6.2.1.2: Materials, Design, and Procedure

Eighty words were selected from Experiment 4 and 40 non-words (e.g., *dralp*) from the English Lexicon Project (Balota et al., 2007), following Mueller et al. (2014). The length of the non-words was 5 and all were monosyllabic. The items were randomly divided into two sets, one assigned to the CID task and the other to the lexical decision task. Set assignment to tasks was counterbalanced across participants. In the CID task, four words were used for practice and 36 for the main experiment. For each participant, the program randomly selected half the words to be presented in large and the remainder in small font sizes. In the lexical decision task, four words and four non-words were used for practice and 36 words and 36 non-words for the main experiment. For each participant, half the words and half the non-words were randomly chosen to be presented in large and the remainder in small font sizes. In both the CID and lexical decision tasks, the presentation sequence of items was randomly determined.

Experiment 5 involved a 2 (font size: large/small) × 2 (task: CID/lexical decision) within-subjects design. Half of the participants performed the CID task first followed by the lexical decision task, and the task order was reversed for the remainder of the participants. The procedure in the CID task was identical to that in Experiment 4 except that participants did not make item-by-item JOLs and did not take a free recall test. In the lexical decision task, words and non-words were randomly presented, one at a time; half in large and half in small font sizes. Participants were asked to judge whether the presented item was a word or a non-word as rapidly and accurately as they could by pressing the ‘f’ (word) or ‘j’ (non-word) key.

One reason for omitting item-by-item JOLs was that participants experienced non-words in the lexical decision task but not in the CID task. In the lexical decision task, the word type (word/non-word) might affect JOLs as well as the font size. As the aim of the experiment was specifically to explore whether the CID task is more sensitive to variations in perceptual fluency than the lexical decision task, omitting both the requirement for participants to make JOLs and the final memory test allowed us to compare the sensitivities of these two tasks to perceptual fluency while minimising influences from other task demands.

6.2.2: Results

In the CID task, there was no significant difference in identification accuracy between large and small words, M_{diff} = 0.5%, 95% CI = [-3.4%, 4.3%], $t(11) = 0.27$, $p = .80$, $d = 0.08$ (see Table 2). In the lexical decision task, a repeated measures ANOVA, with word type (word/non-word) and font size as the within-subjects independent variables and judgement accuracy as the dependent variable, showed that words were judged more accurately than non-words, $F(1,11) = 5.27$, $p = .04$, $\eta_p^2 = .32$, but there was no main effect of font size, $F(1,11) = 1.44$, $p = .26$, $\eta_p^2 = .12$, and no significant interaction between font size and word type, $F(1,11) = 0.85$, $p = .38$, $\eta_p^2 = .07$ (see Table 2). All incorrectly identified trials in the CID task and incorrectly

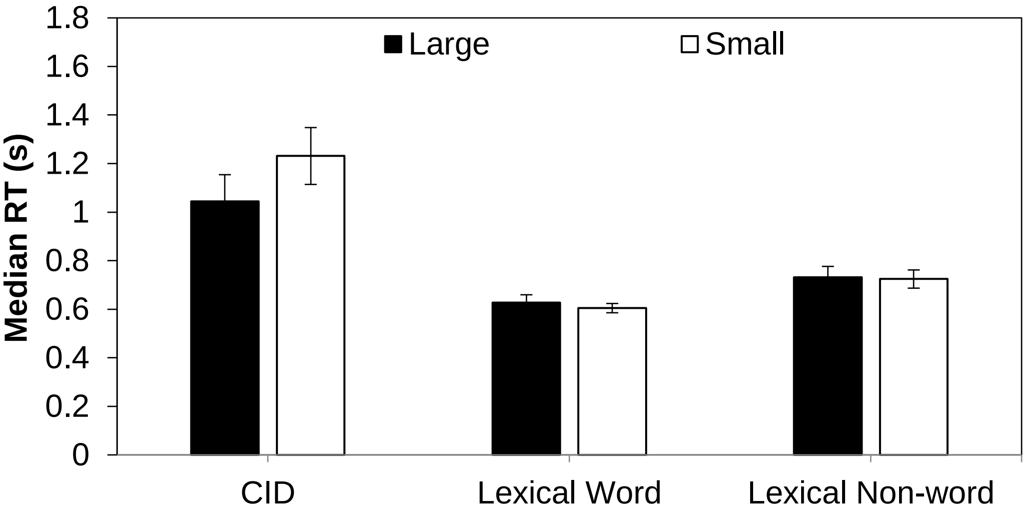

Figure 10. Experiment 5. Median identification RTs in the CID task and median judgment RTs in the lexical decision task for the word and non-word trials. Error bars represent ± 1 standard error.

judged trials in the lexical decision task were removed from this and subsequent analyses.

Figure 10 shows participants' median identification RTs in the CID and lexical decision tasks. In the CID task, participants identified large (M = 1.04 s, SD = 0.38) words faster than small (M = 1.23 s, SD = 0.41) words, M_{diff} = -0.19 s, 95% CI = [-0.28, -0.09], $t(11)$ = 4.40, p = .001, d = -1.27 (see the left pair of bars in Figure 10). All participants responded faster on average to large than to small words, $\chi^2(1)$ = 12.00, p < .001. Overall identification RTs were slightly faster than in Experiment 4, probably caused by a combination of two factors: (i) The requirement to make a JOL on each trial in Experiment 4 may have induced participants to delay making their identification response while they formed their judgement; (ii) In Experiment 5 participants were free to choose which hand to use to make their response (in Experiment 4 they used their left hand).

For the lexical decision task, a repeated measures ANOVA, with word type (word/non-word) and font size as the within-subjects variables and decision RTs as the dependent variable, showed that participants responded faster to words than to non-words, $F(1,11) = 22.19$, $p = .001$, $\eta_p^2 = .67$, but there was no main effect of font size, $F(1,11) = 1.06$, $p = .33$, $\eta_p^2 = .09$, and no interaction between word type and font size, $F(1,11) = 0.59$, $p = .46$, $\eta_p^2 = .05$ (see the middle and right pairs of bars in Figure 10). There was no difference in decision RTs between large ($M = 0.63$, $SD = 0.11$) and small ($M = 0.60$, $SD = 0.07$) words, $M_{\text{diff}} = 0.02$ s, 95% CI = [-0.02, 0.06], $t(11) = 1.24$, $p = .24$, $d = 0.36$. Five participants responded faster to large words than to small words while seven showed the reverse pattern, $\chi^2(1) = 0.33$, $p = .57$. These results replicate Mueller et al.'s (2014) finding that there is no reliable difference in RTs between large and small words in a lexical decision task.

The critical question of interest is whether a significant interaction is shown between task and font size in RTs. A repeated measures ANOVA, with task (CID vs. lexical decision) and font size as the within-subjects independent variables, and RTs in the CID task and RTs to words in the lexical decision task as the dependent variable, showed that participants responded faster to large words than to small words, $F(1,11) = 10.60$, $p = .008$, $\eta p^2 = .49$, faster in the lexical decision task than in the CID task, $F(1,11) = 27.19$, $p < .001$, $\eta_p^2 = .71$, and there was a significant interaction between task and font size, $F(1,11) = 24.97$, $p < .001$, $\eta_p^2 = .69$.

6.2.3: Discussion

By employing the same participants with the same number of trials and with the same materials, we found a significant difference in RTs between large and small words in the CID task but not in the lexical decision task. These results indicate that the CID task provides a more sensitive measure of perceptual fluency than the lexical decision task.

6.3 Experiment 6

In Experiment 4, we observed an inverse relationship between identification RTs and JOLs (i.e., the fluency effect on JOLs), and that fluency (i.e., identification RTs) partly mediates the font size effect size on JOLs. The first aim of Experiment 6 is to replicate these findings, and the second aim is to explore how fluency affects JOLs. There are two possibilities. The first is that fluency affects JOLs directly: Fluency produces a feeling- of-knowing, which acts as a basis for JOLs. The second possibility is that fluency affects JOLs indirectly through people’s beliefs about fluency: People believe that fluently processed items are easier to remember, and therefore they give higher JOLs to fluently processed items (for detailed discussion, see Dunlosky, Mueller, & Tauber, 2014; Mueller & Dunlosky, 2017). For example, Mueller and Dunlosky (2017) recently proposed that “a *belief about processing fluency* appears to produce the font-size effect (on JOLs) (Mueller et al., 2014) and not differential processing fluency per se.” (p. 11). However, recent research has also provided evidence that beliefs about fluency cannot explain the fluency effect on JOLs (e.g., Undorf et al., 2017). Therefore, Experiment 6 aims to explore whether fluency directly affects JOLs or affects them indirectly through beliefs about fluency.

The third aim of Experiment 6 is to test the *analytic processing* (Mueller & Dunlosky, 2017; Mueller et al., 2016) and *dual-basis* theories (Koriat, Bjork, Sheffer, & Bar, 2004; Mueller et al., 2016). The analytic processing theory proposes that people's beliefs play a dominant role in JOLs whereas fluency plays a much smaller or even no role. In contrast, the dual-basis theory claims that both fluency and beliefs contribute importantly to JOLs (Koriat et al., 2004; Mueller et al., 2016). A few previous studies have tested these two theories, with inconclusive results (e.g., Mueller & Dunlosky, 2017; Mueller et al., 2016; Mueller et al., 2014; Undorf et al., 2017; Witherby & Tauber, 2017).

To conceptually replicate Experiment 4's findings, in Experiment 6 we asked participants to perform the same study task as that in Experiment 4, in which they identified 100 words, half in small and half in large font sizes, and made item-by-item JOLs. To test whether fluency directly affects JOLs or affects them indirectly through beliefs, we need to measure the latter and explore to what extent beliefs can explain the fluency effect on JOLs (Undorf & Erdfelder, 2011; Undorf et al., 2017). We employed the learner-observer paradigm (Undorf & Erdfelder, 2011) to measure beliefs about fluency. Immediately following the study task, participants were asked to perform an observation task, in which they were instructed to view identification responses purportedly from another participant and make item-by-item JOLs to predict the likelihood that that participant would remember the item. In the observation task, each word was replaced by a letter string (i.e., *abcde*), and the letter string and the mask were presented in the same font size and duration as a corresponding item in the study task (see below for details).

Because participants in the observation task did not explicitly experience the identification process, JOLs can only be based on beliefs. This observation task can be regarded as a measure of both participants' beliefs about the font size effect on memory (i.e., whether they believe that large words are more likely to be remembered than small words) and their beliefs about fluency (i.e, whether they believe that more rapidly identified items are more likely to be remembered), because in the observation task they viewed each item's font size and identification speed.

In the data analysis, we conducted a multilevel mediation analysis to explore whether beliefs mediate the fluency effect on JOLs. To test the analytic processing and dual-basis theories, we asked whether beliefs (both about the font size effect on memory and about processing fluency) can explain a greater proportion of the font size effect on JOLs than fluency, or the reverse.

Previous studies showed that participants may adjust their beliefs across a study phase (e.g., Susser et al. 2017; Undorf & Erdfelder, 2011). Putting the observation task after the study task for all participants allows us to measure the beliefs that they developed and applied in the study task.

6.3.1 Method

6.3.1.1: Participants

We planned the same sample size as in Experiment 4 in order to conceptually replicate it. Thirty participants were recruited from the UCL

Psychology Subject Pool. One participant's data were not recorded due to computer failure, leaving an effective final sample of 29 participants (*M* = 20.72 years old, *SD* = 2.45; 21 females, 8 males). All reported normal or corrected-to-normal vision, and received a cash payment (£5) or partial course credit in exchange for participation.

6.3.1.2: Materials, Design, and Procedure

The same stimuli were employed as in Experiment 4. Experiment 6 consisted of three tasks: study, observation, and test. The study task was same as in Experiment 4: Participants identified 100 words with the CID procedure (Figure 8), half in large and half in small font sizes, and made item-by-item JOLs. Following the study phase, participants were given the following instructions for the observation task:

> "You will observe the responses of another participant who had undergone the same learning task. However, instead of seeing the exact words which the participant identified, you will see the letter string "abcde" in place of all the words. On each trial, the mask and the letter string will be displayed to you in the same FORMAT as in the learning phase, and for the same DURATION that the participant took to identify the word. Please CAREFULLY observe the participant's identification process, put yourself in his or her perspective, and judge the likelihood that he or she would remember that word at a later test."

Although we told participants that they would observe another participant's identification trials, in fact they observed their own study phase trials replayed without the word information. Ten practice trials were presented in the same font size and duration as the practice trials in the study task, but in a new random order. In the main observation phase, they observed their own identification trials in a new random order. On each trial, the letter string (i.e., *abcde*) and mask were alternately presented. No response was required to terminate the identification part of the trial. Following the presentation of the letter string and mask, participants made a JOL to predict the likelihood that the “other participant” would remember that word later. They then pressed the ENTER key to trigger the next trial.

To summarise, participants observed their own identification trials during the observation task, but we informed them that they were observing another participant’s trials. In addition, we presented all items in a new random order. The aim was to prevent participants from realizing that they were observing their own identification trials and then explicitly recalling their corresponding JOLs from the study task.

Following the observation task, all participants completed a free recall test, which was the same as in Experiment 4. We also measured how much effort participants put into the study and observation tasks. After completing each of these phases, participants reported how much effort they had exerted on a scale ranging from 1 (no effort at all) to 7 (full effort).

6.3.2: Results and Discussion

Participants’ effort ratings were greater than the midpoint of the rating scale (i.e., 4) in both the study (M = 5.10, SD = 0.94) and observation (M = 4.93, SD = 1.16) tasks. There was no difference in effort ratings between the two tasks, M_{diff} = 0.17, 95% CI = [-0.21, 0.55], $t(28)$ = 0.93, p = .36, d = 0.17. These results suggest that participants engaged in both tasks to an approximately equal extent.

In the study task there was no significant difference in identification accuracy between large and small words, M_{diff} = 1.1%, 95% CI = [-0.1%,

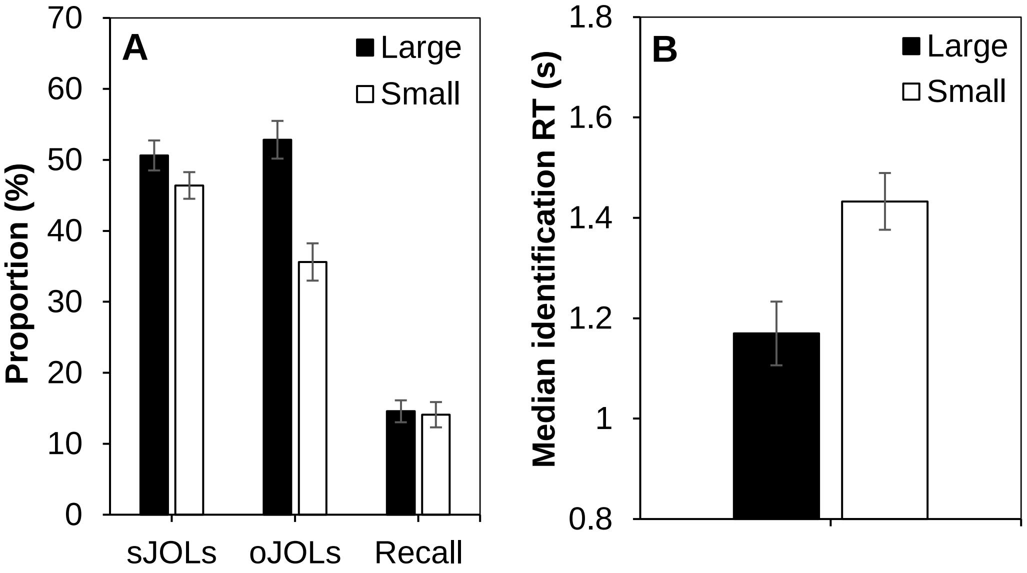

Figure 11. Experiment 6. Panel A: Study judgments of learning (sJOLs), observation JOLs (oJOLs), and recall for large and small words. Panel B: Median identification RTs for large and small words. Error bars represent ± 1 standard error.

2.3%], *t*(28) = 1.86, *p* = .07, *d* = 0.35 (see Table 2). All data from misidentified trials were removed from the subsequent analyses.

There was no difference in recall accuracy between large (*M* = 14.6%, *SD* = 8.31%) and small (*M* = 14.1%, *SD* = 9.6%) words, M_{diff} = 0.5%, 95% CI = [-2.2%, 3.2%], *t*(28) = 0.36, *p* = .72, *d* = 0.07 (see the right pair of bars in Figure 11A). In contrast, in the study task participants gave significantly higher sJOLs (i.e., JOLs in the study task) to large (*M* = 50.63, *SD* = 11.38) than to small words (*M* = 46.40, *SD* = 10.10), M_{diff} = 4.23, 95% CI = [2.47, 5.99], *t*(28) = 4.92, *p* < .001, *d* = 0.91 (see the left pair of bars in Figure 11A), reflecting a font size effect on sJOLs. In the observation task, participants gave significantly higher oJOLs (i.e., JOLs in the observation task) to large (*M* = 52.84, *SD* = 14.32) than to small words (*M* = 35.62, *SD* = 14.16), M_{diff} = 17.22, 95% CI = [11.35, 21.09], *t*(28) = 6.01, *p* < .001, *d* = 1.12 (see the middle pair of bars in Figure 11A). As can be seen in Figure 11B, participants' median identification RTs were significantly faster for large (*M* = 1.17 s, *SD* = 0.38) than for small (*M* = 1.43 s, *SD* = 0.56) words, M_{diff} = -0.26 s, 95% CI = [-0.36, -0.16], *t*(28) = -5.42, *p* < .001, *d* = -1.00. Twenty-five participants responded faster on average to large than to small words while only four showed the reverse pattern, $\chi^2(1) = 15.21$, *p* < .001.

6.3.2.1: Does fluency contribute to the font size effect on JOLs?

We conducted a multilevel mediation analysis using the *bmlm* package in R, with font size as the independent variable, fluency (identification RTs) as the mediator, and sJOLs as the dependent variable, to explore whether fluency mediates the font size effect on JOLs (see Table 3 for detailed results). The total effect of font size on sJOLs was 4.30, 95% CI = [2.51, 6.09]. The indirect effect of font size on sJOLs through RTs was 0.60, 95% CI = [0.30, 0.99], slightly smaller than in Experiment 4 but nonetheless again indicating that large fonts increase JOLs indirectly by increasing perceptual fluency. Fluency (identification RTs) explained 15%, 95% CI = [7%, 28%], of the font size effect on sJOLs. The direct effect of font size on sJOLs was 3.69, 95% CI = [1.88, 5.49]. These results successfully replicated Experiment 4's findings: Font size affects JOLs, at least partially, through perceptual fluency.

6.3.2.2: Does fluency affect JOLs through beliefs about fluency?

In the following analyses, we explored whether fluency affects JOLs through beliefs about fluency. We first conducted a multilevel regression of RTs on sJOLs to quantify the fluency effect on sJOLs. The results showed an inverse relationship between RTs and sJOLs, fixed effect = -3.34, 95% CI

= [-4.49, -2.17], indicating that every decrease of 1 s in RTs increases sJOLs by 3.34. Then we conducted a multilevel regression of RTs on oJOLs to explore people's beliefs about fluency. This analysis found no significant relationship between RTs and oJOLs, fixed effect = 1.41, 95% CI = [-4.41, 7.22], hence revealing a dissociation between the fluency effect on sJOLs and beliefs about fluency, contradicting the claim that fluency affects JOLs through beliefs. Thus while the identification RT for a word in the learning task predicts the sJOL given to it, it does not predict the oJOL given to the letter string "*abcde*" when the latter is presented in the observation phase for the same duration as the word had been in the learning phase.

Although we observed no relationship between RTs and oJOLs, we also conducted a multilevel mediation analysis to explore whether beliefs mediate the fluency effect on sJOLs. This analysis was conducted using the *bmlm* package in R, with fluency (identification RTs) as the independent variable, beliefs (oJOLs) as the mediator, and sJOLs as the dependent variable (see Table 3 for detailed results). The results show that the indirect effect of fluency (RTs) on sJOLs through beliefs (oJOLs) was -0.12, 95% CI = [-0.94, 0.58], which is non-significant (because the 95% CI includes 0) and again counter to the claim that fluency affects JOLs via beliefs.

6.3.2.3: Do beliefs play a more important role than fluency in the font size effect on JOLs, or vice versa?

We also explored whether beliefs play a more important role than fluency in the font size effect on JOLs, or whether the reverse is true. First, we conducted a multilevel mediation analysis to determine whether beliefs (oJOLs) mediate the font size effect on sJOLs. This mediation analysis was conducted using the R *bmlm* package, with font size as the independent variable, beliefs (oJOLs) as the mediator, and sJOLs as the dependent variable (for detailed results, see Table 3). The results show that the total effect of font size on sJOLs was 4.30, 95% CI = [2.32, 6.31] and the indirect effect of font size on sJOLs through beliefs was 0.39, 95% CI = [-0.65, 1.63]. The proportion of the effect of font size on sJOLs mediated by beliefs (oJOLs) was 9%, 95% CI = [-17%, 38%]. The direct effect of font size on sJOLs was 3.91, 95% CI = [1.88, 6.00]. Overall, this multilevel mediation analysis shows little evidence supporting the claim that font size affects JOLs via beliefs.

In the next multilevel mediation analysis, font size was assigned as the independent variable, fluency (RTs) and beliefs (oJOLs) as two mediators, and sJOLs as the dependent variable. The analysis was conducted using the Mplus program (Muthén & Muthén, 1998-2010; Preacher, Zyphur, & Zhang, 2010)[6]. Table 3 reports the detailed results. The indirect effect of font size on sJOLs through fluency (identification RTs) was 0.62, 95% CI = [0.29, 0.95], again indicating that font size affects JOLs at least partially through fluency. The indirect effect of font size on sJOLs though beliefs (oJOLs) was 0.64, 95% CI = [-0.22, 1.50], which again shows little evidence that font size affects JOLs through beliefs. The difference between the indirect effect of font size on sJOLs via fluency (identification

---

[6] We switched to Mplus (https://www.statmodel.com/) because the R *bmlm* package is not yet applicable to multilevel mediation analyses with multiple mediators (we thank Matti Vuorre for confirming this). We also conducted multilevel mediation analyses using Mplus to replicate the ones reported above conducted with the *bmlm* package. All the results showed the same patterns. We report results from the R *bmlm* package in the Results section because it provides Bayesian estimation (Vuorre, 2017).

RTs) and via beliefs (oJOLs) was -0.02, 95% CI = [-0.90, 0.86], indicating no difference between the indirect effects. Overall, these results are inconsistent with the claim of analytic processing theory that beliefs dominate fluency, but are in line with the dual-basis theory.

6.4 Summary and Discussion of Experiments 4-6

Until recently the font size effect on JOLs was widely taken as direct evidence that perceptual fluency can affect JOLs. However, Mueller et al. (2014) found no difference in fluency between large and small words when assessed by means of lexical decision and self-regulated study time allocation and hence suggested that the fluency theory is unlikely to provide an adequate account of the font size effect on JOLs. Subsequently, many researchers began to question the role that perceptual fluency plays in the font size effect on JOLs. For instance, Mueller and Dunlosky (2017) interpreted font size experiments as revealing that JOLs are mainly based on the deliberate application of people’s beliefs. We suspected that the null result in RTs in the lexical decision task in Mueller et al.’s (2014) Experiment 1 might be caused by task insensitivity to variations in perceptual fluency. In addition, the null result in the self-regulated study time allocation task in Mueller et al.’s (2014) Experiment 2 might have resulted from the fact that this dependent measure can be affected by many other possible factors besides fluency.

In the present studies, we directly tested the fluency theory by employing a CID task. In Experiments 4 and 6, we found a substantial font size effect on JOLs as reflected by a significant difference in JOLs between large and small words, while font size had no effect on actual recall performance. These results replicate the classic font size effect on JOLs (Hu et al., 2015; Hu et al., 2016; Mueller et al., 2014; Rhodes & Castel, 2008). Our results show that large words are processed with greater perceptual fluency than small words, as revealed by a significant difference in identification RTs. There was a significant fluency effect on JOLs, supported by an inverse relationship between RTs and JOLs. More importantly, we also found that large font size increases JOLs indirectly by increasing perceptual fluency, as reflected by a significant mediation of RTs in the font size effect on JOLs. These results bring the fluency theory back to the foreground as an account for the font size effect on JOLs. Going beyond Undorf et al. (2017), these results also provide direct evidence that perceptual fluency partly underlies the stimulus size effect on JOLs.

Experiments 4 and 6 contradict Susser et al.'s (2016) proposal that effects of perceptual fluency on JOLs do not exist. Our findings, corroborating those of Undorf et al. (2017), support the conclusion that perceptual fluency can affect JOLs. The differences in perceptual fluency in our Experiments 4 and 6 and those in Undorf et al.'s (2017) Experiments 4-6 were greater than in Susser et al.'s (2016) Experiment 5. The perceptual fluency effects on JOLs might have been too small to be detected in Susser et al.'s (2016) study which had only 36 trials, compared to 100 trials in our Experiments 4 and 6 and 64 trials in Undorf et al.'s (2017) Experiments 4-6. Therefore, lack of power resulting from the small number of trails might have contributed to the null result in Susser et al.'s study.

Consistent with Undorf et al.'s (2017) findings, our Experiments 4 and 6 also challenge the proposal that beliefs play the dominant role in the formation of JOLs (Mueller & Dunlosky, 2017; Mueller et al., 2014; Mueller et al., 2013). Experiment 4 supports the dual-basis theory, which proposes that JOLs are based on both beliefs and fluency (Koriat, 1997; Undorf et al., 2017). Furthermore, Experiment 6 directly compared the contributions of fluency and beliefs (both beliefs about the effect of font size on memory and about fluency) to the font size effect on JOLs. The results revealed no difference in the roles (importance) of fluency and beliefs in the font size effect on JOLs, which is inconsistent with the analytic processing theory but in line with the dual-basis theory.

However, it is important to note that we do not entirely reject the analytic processing theory. In Experiment 6, we also explored whether fluency affects JOLs directly or indirectly through beliefs about fluency. We observed an inverse relationship between identification RTs and sJOLs but no relationship between identification RTs and oJOLs, indicating a dissociation between the fluency effect on JOLs and beliefs about fluency. In addition, the multilevel mediation analysis found no evidence that beliefs about fluency mediate the fluency effect on JOLs. There are two potential explanations of these results. The first possibility is that participants in Experiment 6 simply had no beliefs about fluency. The second possibility is that they had such beliefs but did not apply them when forming their oJOLs in the observation task (Koriat et al., 2004; Kornell & Hausman, in press; Kornell et al., 2011). Participants in the observation task might regard font size as a more salient cue than identification speed, and therefore base their oJOLs on font size rather on identification speed. If they did not apply beliefs about fluency to form their oJOLs, then there is little reason to expect that they applied beliefs about fluency when forming their sJOLs in the study task, because they experienced the difference in font sizes in both the study and observation tasks. Therefore, regardless of whether participants had no beliefs about fluency or had such beliefs but did not apply them, we propose that, at least in the current research, beliefs about fluency play no role in the fluency effect on JOLs.

In recent years, the roles of fluency and beliefs in JOLs has received a great deal of attention among researchers (e.g., Dunlosky et al., 2014; Frank & Kuhlmann, 2016; Mueller & Dunlosky, 2017; Mueller et al., 2016; Mueller et al., 2014; Undorf & Ackerman, 2017; Undorf & Erdfelder, 2011, 2014; Yang et al., 2017b). How to measure and compare the roles (and importance) of fluency and beliefs in JOLs has been a key concern. Experiment 6 provides an demonstration of how to achieve this using the same participants with the same items.

In Experiment 5, we directly compared the CID and lexical decision tasks by employing the same participants, same number of trials, and with the same materials. We found a significant difference in identification RTs between large and small words in the CID task, but no difference was found in decision RTs in the lexical decision task. These results are consistent with previous studies’ findings (Ferrand et al., 2011; Grainger & Segui, 1990) and clearly indicate that the CID task is more sensitive to variations in perceptual fluency than the lexical decision task. Although the principal implications of the results concern the effects of fluency on metacognitive judgements, they also have implications on the theoretical analysis of perceptual identification and lexical decisions. It is well-established that variables can have effects of very different magnitude on naming (identification) and lexical decision. For instance, word frequency has a much larger impact on lexical decision than on naming latencies (Schilling, Rayner, & Chumbley, 1998). The difference between these tasks is usually conceptualised in terms of the additional decision stage required to judge whether a lexical item is a word or a non-word. Models of lexical decision (e.g., Balota & Chumbley, 1984; Ratcliff, Gomez, & McKoon, 2004), which tend to focus on non-perceptual variables such as word frequency and concreteness, could be extended to incorporate variables such as size, colour, or font which have not traditionally been considered.

The lexical decision and self-regulated study time allocation tasks have both commonly been used in previous studies examining the role of fluency in metamemory. For instance, Jia et al. (2015) explored whether fluency underlies the word frequency effect on JOLs (i.e., higher JOLs to high frequency words than to low frequency words) by employing a self-regulated study time allocation task. They found no difference in study times allocated to high versus low frequency words. In another example, Mueller et al. (2016) explored whether fluency underlies the identity effect on JOLs (i.e., higher JOLs to identical word pairs, e.g., dog-dog, than to related pairs, e.g., dog-cat) by employing a self-regulated study time allocation task. Mueller et al. (2016) found that study times were shorter for identical pairs than for related pairs, but study times did not mediate the identity effect on JOLs. Witherby and Tauber (2017) also investigated whether fluency underlies the concreteness effect on JOLs (i.e., higher JOLs to concrete words than to abstract words). By employing a lexical decision task, Witherby and Tauber (2017) found that judgement RTs were shorter for concrete words than for abstract words, but RTs did not mediate the concreteness effect on JOLs. Using a self-regulated study time allocation task, Witherby and Tauber (2017) found no difference in study times between concrete and abstract words.

All of the aforementioned studies employed lexical decision or self-regulated study time allocation tasks to explore the role of fluency in some metamemory phenomena, failed to either find a significant difference in fluency or a significant mediation of fluency, and then concluded that fluency plays no role in these metamemory phenomena. We encourage future research to re-examine these metamemory phenomena by employing the CID task.

There are two main limitations in the present studies. The first limitation is that in both Experiments 4 and 6, participants made a JOL immediately following each correct identification. Such a procedure might draw participants' attention to fluency and inflate its influence on JOLs. Drawing participants' attention to fluency might also contribute to the null difference between the indirect effect through fluency and that through beliefs in Experiment 6. Another limitation is that in Experiment 6, fluency (identification RTs) and sJOLs were collected in the study task, but beliefs (oJOLs) were measured in the observation task, which might contribute to the null difference in the indirect effects. Therefore, we reiterate that we do not reject the analytic processing theory. Future research is recommended to develop more elegant procedures to avoid drawing participants’ attention to fluency (or to measure fluency less overtly) while measuring the role of fluency in JOLs. In addition, future research is encouraged to develop new methods to measure the roles of fluency and beliefs in JOLs simultaneously (in contrast to the different tasks in Experiment 6), allowing researchers to compare the roles of fluency and beliefs more precisely.

Chapter 7: General Discussion

The aim of the present thesis was to explore to what extent, and under what conditions, processing fluency is linked to complex memory processes, namely source memory and metamemory. We were specifically interested in whether fluency relates to source memory performance accuracy across various encoding and stimuli contexts, and how fluency contributes to JOL formation in the font size effect in comparison to metamnemonic beliefs.

Perceptual processing fluency is closely related to the accuracy of source memory judgements, at least across several conditions, as shown in Experiments 1-3b. This provides further evidence against versions of dual-system models that specifically assume complete independence, or minimal interrelatedness, between implicit and explicit bases of memory (e.g., between priming and recognition; Tulving, Schacter, & Stark, 1982). The findings are concordant with a single-system model of memory, and with dual-system models which assume some association at the latent level between implicit and explicit memory (Berry et al., 2012).

JOLs can also be affected by perceptual processing fluency. Large font size increases JOLs at least in part through increasing perceptual fluency, which implies that perceptual fluency contributes to the stimulus size effect on JOLs. Experiments 4-6 found little evidence that beliefs about fluency play a role in the fluency effect on JOLs. The results support the dual-basis theory (Koriat, 1997), but we reiterate that we do not reject the analytic processing theory (Mueller & Dunlosky, 2017). Additionally, the CID task provides a more sensitive measure of perceptual fluency than the lexical decision task.

The relationship between source memory accuracy and processing fluency is also consistent with recent experiments by Lange, Berry, and Hollins (2019; under review). Also using a modified CID-R task, Lange et al. (2009) demonstrated that both source confidence and accuracy corresponded to faster identification RTs on the task. This relationship was enhanced particularly when the overall memory strength on the task is increased (via shortening the study list; Experiment 2). Additionally, by measuring identification RTs and memory ratings in separate test phases, Lange et al. (under review; Experiments 3a and 3b) addressed the possibility that this effect of fluency on item and source recognition confidence ratings might be a consequence specific to the interleaved structure of the CID-R task itself (which is also a limitation of our Experiments 1-3b in the present Chapters 2-5). As items are identified more quickly, participants might be more inclined to attribute fluency to the prior exposure of the item at study as a function of the memory judgements being immediately preceded by the identification (Jacoby & Dallas, 1981). In other words, even if performance on the CID and the memory judgements are based on two separate systems, performance could be correlated simply due to the temporal proximity of the two tasks, which would still be consistent with dual-system models which allow for correlations between implicit and explicit bases of memory (e.g., the MS2 model; Berry et al., 2012). The results of Lange et al.’s Experiments 3a and 3b demonstrate that the relationship between identification RTs, source memory accuracy, and source confidence persists despite the separation of the CID and memory judgement components at test.

Kurilla’s (2010) Experiment 3 showed that, as for perceptual fluency, enhancing conceptual fluency increases the rate of responding that items were presented in the same source modality at study and test. Other research has shown that conceptual fluency is related to accuracy on source memory judgements, but results are conflicting as to whether this corresponds to greater false recollection (Henkel & Franklin, 1998) or to greater judgement accuracy (Lyle & Johnson, 2001). Just like for perceptual fluency, topics that await additional research include the contribution of conceptual fluency on retrieval accuracy in multidimensional source memory tasks, as well as the effects of other potential moderators (e.g., contextual fan and unitisation) on the relationship between conceptual fluency and source accuracy.

Experiments 4-6 were the first to find a non-inferior contribution of processing fluency (vs. beliefs) to the stimulus size effect on JOLs through assessing this relationship using a direct measurement of perceptual fluency. Although these results agree with dual-basis models of JOL formation, they do not entirely contradict the analytic processing model either, as some would consider the analytic processing model to be a subclass of the dual-basis models (Mueller & Dunlosky, 2017). Indeed, neither dual-basis nor analytic processing models completely rule out the contributions of either fluency or metacognitive beliefs to JOLs. However, the analytic processing model is more specific with its predictions. It asserts that people exert conscious control over the monitoring and search for features of the stimulus or task (e.g., size) which could be incorporated as metacognitive cues, and then proposes that beliefs based on the anticipated mnemonic impact of those cues (e.g., “bigger/faster = easier to remember”) are subsequently applied to make metacognitive judgments (Kelley & Jacoby, 1996).

Thus, analytic processing theory assumes a substantially greater role of beliefs than fluency in JOL formation: It allows for the possibility of fluency to directly influence JOL formation, but only when no beliefs, neither formed on-line nor *a priori* ones retrieved from memory, are constructed by the learner (Mueller & Dunlosky, 2017). Even though our Experiment 6 did not find any evidence for the presence of beliefs about processing fluency (“more fluent = better remembered”; as indexed through oJOLs), Experiments 4 and 6 still showed that participants held metamnemonic beliefs about font size. Assuming that beliefs are constructed and applied in the same way to both the study (sJOLs) and observation (oJOLs) tasks, then the prevalence of the direct effect of fluency on JOLs in these experiments, despite the presence of font size beliefs, challenges the analytic processing assumption that fluency may only play a role in JOL formation when no beliefs are available to fulfil this role.

New lines of work that could extend from the present studies can continue to scrutinise the role of processing fluency in other kinds of memory decisions or metacognitive effects in the context of the intersections between source memory, item memory, and metamemory. One particularly relevant avenue is research on judgements of source (JOS), which are similar to JOLs but refer specifically to the likelihood of remembering source information associated with an item at a later test, assuming that the item could also be remembered (Carroll, Mazzoni, Andrews, & Pocock, 1999). Recent experiments by Schaper, Kuhlmann, and Bayen (2019) studied the roles of processing fluency versus beliefs in the expectancy effect (i.e., the tendency for expected item-source pairs to be judged as easier to remember than unexpected pairs), looking at both JOLs and JOSs. Their results suggest that the two types of metamemory judgements weigh fluency and beliefs differently – the expectancy effect on JOSs appeared to be less moderated by fluency (as measured via study time allocation) compared to JOLs, despite the prevalence of expectancy effects on both JOLs and JOSs. It would be interesting to further examine whether beliefs would still have the greater contribution to the expectancy effect on JOSs, if the CID or another more perceptually oriented measure of fluency is used instead of study time allocation.

The utility of processing fluency as a diagnostic cue to the accuracy of memory or metamemory performance is complicated, and can depend on numerous task, stimulus, and situational factors. Experiments 1-3b suggested that fluency could be diagnostic of performance accuracy on both item and source memory judgements, but the font size effect of Experiments 4-6 represented a situation where fluency does affect metamemory judgements but might not be diagnostic of metamemory accuracy *per se* (i.e., although people expected larger words to be recalled better, this expectation was inaccurate as large and small words tend to be recalled equally well). Up until recently, research on the use of the fluency heuristic in memory and metacognitive processes has focused mostly on its potential negative impacts on performance accuracy, such as fluency illusions, confabulation, fluency misattributions, and false memories (Koriat, Goldsmith, & Pansky, 2000, Sera & Metcalfe, 2009; c.f., Kelley & Rhodes, 2002). Interventions have also been designed to improve metamemory accuracy across the lifespan through decreasing the use of fluency-based metacognitive strategies (e.g., von der Linden, Löffler, & Schneider, 2016).

However, work by Undorf and Zimdahl (2019) suggested that font size *does in fact* affect memory accuracy: In three out of four of their experiments using a wider range of font sizes (i.e., 6 point to 500 point, and with four sizes, instead of two as conventionally used in font size effect studies), memory performance actually improved with increasing font size, consistent with participants' metamnemonic beliefs. Participants' JOLs also continued to increase monotonically with font size, even beyond the point where a large font impaired perceptual fluency (measured with lexical decision RTs). Based on this observation, the authors concluded that beliefs still make a greater overall contribution to JOLs than perceptual fluency. Setting aside discussions on the appropriateness of lexical decision RTs as a measure of perceptual fluency, Undorf and Zimdahl's (2019) findings point to the importance of continuous inquiry on whether -- and under what circumstances -- processing fluency may help inform better memory or metamemory performance in ecological settings.

Taken together, the present thesis spotlights the importance of processing fluency across a wide range of mnemonic situations, including and beyond item retrieval. However, the importance of familiarity-based cues such as processing fluency has been somewhat neglected by certain models of memory and metamemory, where a greater emphasis is placed on the importance of analytic cues such as recollection or belief. Future research should delineate the extent to which the contribution of perceptual (and perhaps also conceptual) fluency to source and metacognitive judgements might vary when measured at different phases in memory, such as encoding (e.g., Experiments 4-6) versus test (e.g., Experiments 1-3b). Other fruitful avenues of future research include continuing to explore how source memory and source monitoring informs other types of metamemory, and vice versa (Dodson, Kawa, & Krueger, 2007; Schwartz, 1994; Scoboria, Talarico, & Pascal, 2014), and whether any fluency-based interventions could be applied to assist memory in educational and gerontological settings. Along with the present findings, such aforementioned investigations may pave the way towards a more comprehensive understanding of the rich and complex ways that memory operates in our day-to-day lives.

References

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*(2), 97–123. doi:10.1037/h0033773

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*(1), 126-131. doi:10.1037/h0027455

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance* (Vol. 4). New York, NY: Academic Press.

Ball, B. H., Klein, K. N., & Brewer, G. A. (2014). Processing fluency mediates the influence of perceptual information on monitoring learning of educationally relevant materials. *Journal of experimental Psychology: Applied, 20*(4), 336-348. doi:10.1037/xap0000023

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*(3), 340-357. doi: 10.1037/0096-1523.10.3.340

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behavior research methods, 39*(3), 445-459. doi: 10.3758/BF03193014

Bang, P. (n.d.). RGB Color Gradient Maker [web application]. Retrieved from http://www.perbang.dk/rgbgradient/

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*(5), 610-632. doi:10.1016/0749-596X(89)90016-8

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*(4), 446-458.

Berry, C. J., Shanks, D. R., Speekenbrink, M., & Henson, R. N. A. (2012). Models of recognition, repetition priming, and fluency: Exploring a new framework. *Psychological Review*, *119*(1), 40–79. doi:10.1037/a0025464

Besken, M. (2016). Picture-perfect is not perfect for metamemory: Testing the perceptual fluency hypothesis with degraded images. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(9), 1417-1433. doi:10.1037/xlm0000246

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), 897-903. doi:10.3758/s13421-013-0307-8

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444. doi:10.1146/annurev-psych-113011-143823

Brainard, D. H. (1997). The Psychophysics Toolbox, *Spatial Vision 10*(4), 433-436. doi:10.1163/156856897X00357

Brown, A. A., & Bodner, G. E. (2011). Re-examining dissociations between remembering and knowing: Binary judgments vs. independent ratings. *Journal of Memory and Language*, *65*(2), 98–108. doi:10.1016/j.jml.2011.04.003

Bruno, D., & Rutherford, A. (2010). How many response options? A study of remember-know testing procedures. *Acta Psychologica*, *134*(2), 125–129. doi:10.1016/j.actpsy.2010.01.002

Buchler, N. G., Light, L. L., & Reder, L. M. (2008). Memory for items and associations: Distinct representations and processes in associative recognition. *Journal of Memory and Language*, *59*(2), 183-199. doi:10.1016/j.jml.2008.04.001

Cabeza, R., Mengels, J., Nyberg, L., Habib, R., Houle, S., McIntosh, A. R., & Tulving, E. (1997). Bran regions differentially involved in remembering what and when: a PET study. *Neuron*, *19*(4), 863-870. doi:10.1016/S0896-6273(00)80967-8

Cansino, S. (2009). Episodic memory decay along the adult lifespan: A review of behavioral and neurophysiological evidence. *International Journal of Psychophysiology*, *71*(1), 64–69. doi:10.1016/j.ijpsycho.2008.07.005

Carroll, M., Mazzoni, G., Andrews, S., & Pocock, P. (1999). Monitoring the future: Object and source memory for real and imagined events. *Applied Cognitive Psychology*, *13*(4), 373-390. doi:

10.1002/(SICI)1099-0720(199908)13:4<373::AID-ACP605>3.0.CO;2-F

Castel, A. D. (2007). The adaptive and strategic use of memory byolder adults: Evaluative processing and value-directed remembering. *Psychology of Learning and Motivation, 48*, 225-270. doi:10.1016/s0079-7421(07)48006-9

Challis, B. H., & Brodbeck, D. R. (1992). Level of processing affects priming in word fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 595-607. doi:10.1037/0278-7393.18.3.595

Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition, 12*(6), 590-606. doi:10.3758/BF03213348

Cohen, J. (1988). *Statistical power analysis for the behavioral science*s (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *12*(1), 155-159.

Cohn, M., & Moscovitch, M. (2007). Dissociating measures of associative memory: Evidence and theoretical implications. *Journal of Memory and Language, 57*(3), 437-454. doi:10.1016/j.jml.2007.06.006

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497-505. doi:10.1080/14640748108400805

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2006). Source memory in the absence of successful cued recall. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 32(4), 828-835. doi:10.1037/0278-7393.32.4.828

Craik, F. I., Moscovitch, M., & McDowd, J. M. (1994). Contributions of surface and conceptual information to performance on implicit and explicit memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 864-875. doi:10.1037/0278-7393.20.4.864

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* New York, NY: Routledge.

Davidson, P. S. R., McFarland, C. P., & Glisky, E. L. (2006). Effect of emotion on item and source memory in young and older adults. *Cognitive, Affective, & Behavioral Neuroscience*, *6*(4), 306-322. doi:10.3758/CABN.6.4.306

Degl'Innocenti, A., & Bäckman, L. (1996). Aging and source memory: Influences of intention to remember and associations with frontal lobe tests. *Aging, Neuropsychology, and Cognition, 3*(4), 307-319. doi:10.1080/13825589608256632

Diana, R. A, Yonelinas, A. P., & Ranganath, C. (2008). The effects of unitisation on familiarity-based source memory: Testing a behavioral prediction derived from neuroimaging data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 730–40. doi:10.1037/0278-7393.34.4.730

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the Italicized): Effects of disfluency on educational outcomes. *Cognition, 118*(1), 111-115. doi:10.1016/j.cognition.2010.09.012

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging*, *22*(1), 122-133. doi:10.1037/0882-7974.22.1.122

Duarte, A., Ranganath, C., Trujillo, C., & Knight, R. T. (2006). Intact Recollection Memory in high-performing older adults: ERP and behavioral evidence. *Journal of Cognitive Neuroscience*, *18*(1), 33–47. doi:10.1162/089892906775249988

Dunlosky, J., Mueller, M., & Tauber, S. K. (2014). *The contribution of processing fluency (and beliefs) to people's judgments of learning*. New York: NY: Psychology Press.

Easton, A., Webster, L. A. D., & Eacott, M. J. (2012). The episodic nature of episodic-like memories. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *19*(4), 146–150. doi:10.1101/lm.025676.112

Eldridge, L. L., Sarfatti, S., & Knowlton, B. J. (2002). The effect of testing procedure on remember-know judgments. *Psychonomic Bulletin & Review*, *9*(1), 139–145. doi:10.3758/bf03196270

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160. doi:10.3758/BRM.41.4.1149

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Meot, A., . . . Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: evidence from chronolex. *Frontiers in Psychology, 2*, 306. doi:10.3389/fpsyg.2011.00306

Finn, B., & Tauber, S. K. (2015). When confidence is not a signal of knowing: How students' experiences and beliefs about processing fluency can lead to miscalibrated confidence. *Educational Psychology Review, 27*(4), 567-586. doi:10.1007/s10648-015-9313-7

Frank, D. J., & Kuhlmann, B. G. (2016). More Than Just Beliefs: Experience and Beliefs Jointly Contribute to Volume Effects on Metacognitive Judgments. *Journal of Experimental Psychology: Learning, Memory & Cognition*. doi:10.1037/xlm0000332

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Consciousness and Cognition*, *16*(4), 309–313. doi:10.3758/BF03197041

Graf, P., & Schacter, D. L. (1989). Unitization and grouping mediate dissociations in memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 930–940. doi:10.1037/0278-7393.15.5.930

Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Attention, Perception, & Psychophysics, 47*(2), 191-198. doi:10.3758/BF03205983

Greifeneder, R., Alt, A., Bottenberg, K., Seele, T., Zelt, S., & Wagener, D. (2010). On writing legibly: Processing fluency systematically biases evaluations of handwritten material. *Social Psychological and Personality Science*, *1*(3), 230-237. doi:10.1177/1948550610368434

Groninger, L. D. (1979). Predicting recall: The “feeling-that-I-will-know” phenomenon. *American Journal of Psychology*, 45-58. doi:10.2307/1421478

Guttentag, R. E., & Carroll, D. (1997). Recollection-based recognition: Word frequency effects. *Journal of Memory and Language*, *37*, 502-516. doi:10.1006/jmla.1997.2532

Hanczakowski, M., Zawadzka, K., Collie, H., & Macken, B. (2017). Metamemory in a familiar place: The effects of environmental context on feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 59-71. doi:10.1037/xlm0000292

Henkel, L. A., & Franklin, N. (1998). Reality monitoring of physically similar and conceptually related objects. *Memory & Cognition*, *26*, 659-673. doi:10.3758/BF03211386

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1191-1206. doi:10.1037/a0013025

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(1), 22-34. doi:10.1037/0278-7393.29.1.22

Hicks, J. L., Marsh, R. L., & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 503–508. doi:10.1037//0278-7393.28.3.503

Hicks, J. L., & Starns, J. J. (2015). Using multidimensional encoding and retrieval contexts to enhance our understanding of stochastic dependence in source memory. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 62, pp. 101-140. New York, NY: Academic Press. doi:10.1016/bs.plm.2014.09.004

Higham, P. A., & Vokey, J. R. (2004). Illusory recollection and dual-process models of recognition memory. *Quarterly Journal of Experimental Psychology*, *57*(4), 714–44. doi:10.1080/02724980343000468

Hirshman, E., & Henzler, A. (1998). The role of decision processes in conscious recollection. *Psychological Science*, *9*(1), 61–65. doi:10.1111/1467-9280.00011

Hockley, W. E., Bancroft, T. D., & Bryant, E. (2012). Associative and familiarity-based effects of environmental context on memory. *Canadian Journal of Experimental Psychology*, *66*(2), 81-89. doi:10.1037/a0027136

Hu, X., Li, T., Zheng, J., Su, N., Liu, Z., & Luo, L. (2015). How much do metamemory beliefs contribute to the font-size effect in judgments of learning? *PLoS One, 10*(11), e0142351. doi:10.1371/journal.pone.0142351

Hu, X., Liu, Z., Li, T., & Luo, L. (2016). Influence of cue word perceptual information on metamemory accuracy in judgement of learning. *Memory, 24*(3), 383-398. doi:10.1080/09658211.2015.1009470

Inoue, C., & Bellezza, F. S. (1998). The detection model of recognition using know and remember judgments. *Memory & Cognition*, *26*(2), 299–308. doi:10.3758/bf03201141

Jacoby, L. L., & Brookes, L. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 1-47). New York, NY: Academic Press. doi:10.1016/S0079-7421(08)60358-8

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*(3), 306–340. doi:10.1037/0096-3445.110.3.306

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology*, *118*(2), 126–135. doi:10.1037/0096-3445.118.2.126

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia*, *27*(8), 1043–1056. doi:10.1016/0028-3932(89)90184-X

Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2015). The effect of word frequency on judgments of learning: contributions of beliefs and processing fluency. *Frontiers in Psychology, 6*, 1995. doi:10.3389/fpsyg.2015.01995

Johnson, M.K. (1991). Reality monitoring: Evidence from confabulation in organic brain disease patients. In G.P. Prigatano & D.L. Schacter (Eds.), *Awareness of deficit after brain injury* (pp. 176-197). New York, NY: Oxford University Press.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. doi:10.1037/0033-2909.114.1.3

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, *35*(2), 157-175. doi:10.1006/jmla.1996.0009

Kelley, C. M., Jacoby, L. L., & Hollingshead, A. (1989). Direct versus indirect tests of memory for source: Judgements of modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1101–1108. doi:10.1037/0278-7393.15.6.1101

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general-knowledge questions. Journal of Memory & Language, *32*, 1-24. doi:10.1006/jmla.1993.1001

Kelley, C. M., & Rhodes, M. G. (2002). Making sense and nonsense of experience: Attributions in memory and judgment. *Psychology of Learning and Motivation*, *41*, 293-320. doi:10.1016/S0079-7421(02)80010-X

Kirwan, C. B., Wixted, J. T., & Squire, L. R. (2008). Activity in the medial temporal lobe predicts memory strength, whereas activity in the prefrontal cortex predicts recollection. *Journal of Neuroscience*, *28*(42), 10541-10548. doi:10.1523/JNEUROSCI.3456-08.2008

Kleiner, M., Brainard, D., & Pelli, D. (2007, August). *What's new in Psychtoolbox-3?* Tutorial presented at the 30th European Convention on Visual Perception, Arezzo, Italy. Retrieved from http://pec.sagepub.com/content/36/1_suppl.toc

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental*

Psychology: General, 126(4), 349-370. doi:10.1037/0096-3445.126.4.349

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1133-1145. doi:10.1037/0278-7393.32.5.1133

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychololgy: General, 133*(4), 643-656. doi:10.1037/0096-3445.133.4.643

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology*, *51*(1), 481-537. doi: 10.1146/annurev.psych.51.1.481

Koriat, A., Levy-Sadot, R., Edry, E., & de Marcas, S. (2003). What do we know about what we cannot remember? Accessing the semantic attributes of words that cannot be recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1095–1105. doi:10.1037/0278-7393.29.6.1095

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language, 52*(4), 478-492. doi:10.1016/j.jml.2005.01.001

Kornell, N., & Hausman, H. (in press). Performance bias: Why judgments of learning are not affected by learning. *Memory & Cognition*. doi:10.3758/s13421-017-0740-1

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science, 22*(6), 787-794. doi:10.1177/0956797611407929

Kurilla, B. P. (2011). Enhanced processing fluency leads to biases in source memory. *Quarterly Journal of Experimental Psychology (2006)*, *64*(8), 1609–31. doi:10.1080/17470218.2011.561866

Lange, N., Berry, C. J., & Hollins, T. J. (2019, April). *Investigating the role of recognition in the association of priming and source memory*. Paper presented at the Tagung Experimentell Arbeitender Psychologen, London, UK.

Lange, N., Berry, C. J., & Hollins, T. J. (under review). Linking repetition priming, recognition, and source memory: A single-system signal-detection account.

Li, F., Xie, R., Li, X., & Li, W. (2015). The influence of perceptual information on control processes involved in self-regulated learning: Evidence from item selection. *Psychonomic Bulletin & Review, 22*(4), 1007-1013. doi:10.3758/s13423-014-0762-7

Li, P., Jia, X., Li, X., & Li, W. (2016). The effect of animacy on metamemory. *Memory & Cognition, 44*(5), 696-705. doi:10.3758/s13421-016-0598-7

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, *52*(3), 297–318. doi:10.1016/0022-0965(91)90065-Z

Lyle, K. B., & Johnson, M. K. (2007). Source misattributions may increase the accuracy of source judgments. *Memory & Cognition*, *35*,1024–1033. doi:10.3758/BF03193475

Macrae, C. N., Bodenhausen, G. V., & Calvini, G. (1999). Contexts of cryptomnesia: may the source be with you. *Social Cognition*, *17*(3), 273–297. doi:10.1521/soco.1999.17.3.273

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning, 11*(1), 35-56. doi:10.1007/s11409-015-9147-1

Mandler, G., Pearlstone, Z., & Koopmans, H. S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, *8*(3), 410–423. doi:10.1016/S0022-5371(69)80134-9

Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation?. *Memory & Cognition*, *18*(2), 196-204. doi: 10.3758/BF03197095

McCabe, D. P., & Geraci, L. D. (2009). The influence of instructions and terminology on the accuracy of remember–know judgments. *Consciousness and Cognition*, *18*, 401–413. doi:10.1016/j.concog.2009.02.010

Meiser, T., & Broder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 116–137. doi:10.1037//0278-7393.28.1.116

Meiser, T., & Sattler, C. (2007). Boundaries of the relation between conscious recollection and source memory for perceptual details. *Consciousness and Cognition*, *16*(1), 189–210. doi:10.1016/j.concog.2006.04.003

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15*(1), 174-179. doi:10.3758/pbr.15.1.174

Miele, D. B., Finn, B., & Molden, D. C. (2011). Does easily learned mean easily remembered? It depends on your beliefs about intelligence. *Psychological Science, 22*(3), 320-324. doi:10.1177/0956797610397954

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81-97. doi:10.1037/h0043158

Morey, R. (2016). Averaging can produce misleading standardized effect sizes. Retrieved from https://www.statsblogs.com/2016/01/07/averaging-can-produce-misleading-standardized-effect-sizes/

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language, 93*, 245-258. doi:10.1016/j.jml.2016.10.008

Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people's metamemory judgments: What are the contributions of processing fluency and beliefs about memory? *The*

Quarterly Journal of Experimental Psychology, 69(4), 781-799. doi:10.1080/17470218.2015.1058404

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language, 70*, 1-12. doi:10.1016/j.jml.2013.09.007

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review, 20*(2), 378-384. doi: 10.3758/s13423-012-0343-6

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide (Version 6.).* Los Angeles, CA: Muthén & Muthén.

Murdock, B. B., Jr., & Walker, K. D. (1969). Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*(5), 665-676. doi:10.1016/S0022-5371(69)80120-9

Nelson, T. O., & Narens, L. (1990). "Metamemory: A theoretical framework and new findings". In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-141). San Diego, CA: Academic Press. doi:10.1016/S0079-7421(08)60053-5

Noh, S. M., Yan, V. X., Vendetti, M. S., Castel, A. D., & Bjork, R. A. (2014). Multi-level induction of categories: Venomous snakes hijack the learning of lower category levels. *Psychological Science, 25*(8), 1592-1599. doi:10.1177/0956797614535938

Onyper, S. V, Zhang, Y. X., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of*

Experimental Psychology: General, *139*(2), 341–36. doi:10.1037/a0018926

Parks, C. M., & Yonelinas, A. P., (2015). The importance of unitization for familiarity-based learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 881-903. doi:10.1037/xlm0000068

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision 10*(4), 437-442. doi:10.1163/156856897X00366

Perfect, T. J., Mayes, A. R., Downes, J. J., & Van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *Quarterly Journal of Experimental Psychology*, *49A*(3), 797–813. doi:10.1080/713755644

Persson, B. M., Ainge, J. A., & O'Connor, A. R. (2016). Disambiguating past events: Accurate source memory for time and context depends on different retrieval processes. *Neurobiology of Learning and Memory, 132*, 40-48. doi:10.1016/j.nlm.2016.05.002

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*(3), 209-213. doi:10.1037/a0020141.supp

Price, J., & Harrison, A. (2017). Examining what prestudy and immediate judgments of learning reveal about the bases of metamemory judgments. *Journal of Memory and Language, 94*, 177-194. doi:10.1016/j.jml.2016.12.003

Price, J., McElroy, K., & Martin, N. J. (2016). The role of font size and font style in younger and older adults' predicted and actual recall performance. *Aging, Neuropsychology, and Cognition, 23*(3), 366-388. doi:10.1080/13825585.2015.1102194

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Ragland, J. D., McCarthy, E., Bilker, W. B., Rensinger, C. M. B., Valdez, J., Kohler, C., … Gur, R. C. (2006). Levels-of-processing effect on internal source monitoring in schizophrenia. *Psychological Medicine, 36*(5), 641-648. doi:10.1017/S0033291706007094

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*(1), 89–102. doi:10.3758/BF03211168

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*(1), 159-182. doi:10.1037/0033-295X.111.1.159

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?. *Personality and Social Psychology Review*, *8*(4), 364-382. doi:10.1207/s15327957pspr0804_3

Reder, L. M., Donavos, D. K., & Erickson, M. A. (2002). Perceptual match effects in direct tests of memory: The role of contextual fan. *Memory & Cognition, 30*(2), 312-23. doi:10.3758/BF03195292

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), 615-625. doi:10.1037/a0013684

Roediger, H. L., III., Weldon, M. S., Stadler, M. L., & Riegler, G. L. (1992). Direct comparison of two implicit memory tests: Word fragment and word stem completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1521-1269. doi:10.1037/0278-7393.18.6.1251.

Roediger, H. L., III., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger, III, & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3-41). Hillsdale, NJ: Erlbaum.

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal-detection model. *Psychological Review*, *11*(3), 588-616. doi:10.1037/0033-295X.111.3.588

Rutherford, A. (2004). Environmental context-dependent recognition memory effects: An examination of ICE model and cue-overload hypotheses. *Quarterly Journal of Experimental Psychology Section A*, *57*(1), 107-127. doi:10.1080/02724980343000152

Sanborn, A. N., Malmberg, K. J., & Shiffrin, R. M. (2004). High-level effects of masking on perceptual identification. *Vision Research, 44*(12), 1427-1436. doi:10.1016/j.visres.2004.01.004

Schacter, D. L., Kaszniak, A. W., Kihlstrom, J. F., & Valdiserri, M. (1991). The relation between source memory and aging. *Psychology and Aging*, *6*(4), 559–568. doi:10.1037/0882-7974.6.4.559

Schacter, D. L., & McGlynn, S. M. (1989). Implicit memory: Effects of elaboration depend on unitization. *The American Journal of Psychology, 102*(2), 151-181. doi:10.2307/1422950

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019). Metacognitive expectancy effects in source monitoring: Beliefs, in-the-moment experiences, or both?. *Journal of Memory and Language*, *107*, 95-110. doi:10.1016/j.jml.2019.03.009

Schilling, H. E., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26*(6), 1270-1281. doi:10.3758/BF03201199

Schwarz, N., & Reber, R. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, 8(3), 338-342. doi:10.1006/ccog.1999.0386

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review*, *1*(3), 357-375.

Scoboria, A., Talarico, J. M., & Pascal, L. (2015). Metamemory appraisals in autobiographical event recall. *Cognition*, *136*, 337-349. doi:10.1016/j.cognition.2014.11.028

Serra, M. J., & Metcalfe, J. (2009). Effective Implementation of Metacognition. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.).

Handbook of Metacognition in Education. (pp. 278-298). New York, NY: Routledge.

Shimura, A. P., & Squire, L. R. (1991). The relationship between fact and source memory: Findings from amnesiac patients and normal subjects. *Psychobiology*, *19*(1), 1–10. doi:10.1007/BF03337951

Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition*, *33*(1), 151–70. doi:10.3758/BF03195305

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*(2), 203-220. doi:10.3758/BF03196157

Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience*, *8*, 872-883. doi:10.1038/nrn2154

Stark, C. E. L., & McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology*, *26*(4), 945–972. doi:10.10371/0278-7393.26.4.945

Starns, J. J., Rotello, C. M., & Hautus, M. J. (2014). Recognition memory zROC slopes for items with correct versus incorrect source decisions discriminate the dual process and unequal variance signal detection models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1205–1225. doi: 10.1037/a0036846

Susser, J. A., Jin, A., & Mulligan, N. W. (2016). Identity priming consistently affects perceptual fluency but only affects metamemory when primes are obvious. *Journal of Experimental Psychology: Learning, Memory and Cognition, 42*(4), 657-662. doi:10.1037/xlm0000189

Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory & Cognition, 41*(7), 1000-1011. doi:10.3758/s13421-013-0323-8

Susser, J. A., Panitz, J., Buchin, Z., & Mulligan, N. W. (2017). The motoric fluency effect on metamemory. *Journal of Memory and Language, 95*, 116-123. doi:10.1016/j.jml.2017.03.002

Tauber, S. K., Dunlosky, J., Rawson, K. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: Do people interleave or block exemplars during study? *Psychonomic Bulletin & Review, 20*(2), 356-363. doi:10.3758/s13423-012-0319-6

Troyer, A. K., Häfliger, A., Cadieux, M. J., & Craik, F. I. M. (2006). Name and face learning in older adults: Effects of level of processing, self-generation, and intention to learn. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 61*(2), P67-74. doi:10.1093/geronb/61.2.P67

Troyer, A. K., Winocur, G., Craik, F. I., & Moscovitch, M. (1999). Source memory and divided attention: reciprocal costs to primary and secondary tasks. *Neuropsychology*, *13*(4), 467–474. doi:10.1037/0894-4105.13.4.467

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*(1), 1–12. doi:10.1037/h0080017

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior*, *5*, 381-391. doi:10.1016/S0022-5371(66)80048-8

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *8*, 336-342. doi:10.1037/0278-7393.8.4.336

Tunney, R. J., & Fernie, G. (2007). Repetition priming affects guessing not familiarity. *Behavioral and Brain Functions*, *3*(40), 40. doi:10.1186/1744-9081-3-40

Turk-Browne, N. B., Yi, D.-J., Chun, M. M. (2006). Linking implicit and explicit memory: Common encoding factors and shared representations. *Neuron*, *49*(6), 917-927. doi:10.1016/j.neuron.2006.01.030

Undorf, M., & Ackerman, R. (2017). The puzzle of study time allocation for the most challenging items. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-017-1261-4

Undorf, M., & Erdfelder, E. (2011). Judgments of learning reflect encoding fluency: Conclusive evidence for the ease-of-processing hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1264-1269. doi:10.1037/a0023719

Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology*, *71*(5), 673-679. doi:10.1037/h0023107

Undorf, M., & Erdfelder, E. (2014). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition*. doi:10.3758/s13421-014-0479-x

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language, 92*, 293-304. doi:10.1016/j.jml.2016.07.003

Undorf, M., & Zimdahl, M. F. (2019). Metamemory and memory for a wide range of font sizes: What is the contribution of perceptual fluency? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(1), 97-109. doi:10.1037/xlm0000571

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review, 23*(1), 87-102. doi:10.3758/s13423-015-0892-6

von der Linden, N., Löffler, E., & Schneider, W. (2016). Effects of a Short Strategy Training on Metacognitive Monitoring across the Life-span. *Frontline Learning Research*, *3*(4), 37-55. doi:10.14786/flr.v3i4.196

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1-48. doi:10.18637/jss.v036.i03

Vuorre, M. (2017). bmlm: Bayesian Multilevel Mediation. R package version 1.3.0., from https://cran.r-project.org/package=bmlm

Ward, E. V., Berry, C. J., & Shanks, D. R. (2013). An effect of age on implicit memory that is not due to explicit contamination: implications for single and multiple-systems theories. *Psychology and Aging, 28*(2), 429-442. doi:10.1037/a0031888

Williams, H. L., & Moulin, C. J. A. (2015). Know versus Familiar: Differentiating states of awareness in others' subjective reports of recognition. *Memory*, *23*(7), 981-990. doi:10.1080/09658211.2014.945460

Witherby, A. E., & Tauber, S. K. (2017). The concreteness effect on judgments of learning: Evaluating the contributions of fluency and beliefs. *Memory & Cognition, 45*(4), 639-650. doi:10.3758/s13421-016-0681-0

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176. doi:10.1037/0033-295X.114.1.152

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember / know judgments. *Psychonomic Bulletin & Review*, *11*(4), 616–641. doi:10.3758/BF03196616

Wixted, J. T., Mickes, L., & Squire, L. R. (2010). Measuring recollection and familiarity in the medial temporal lobe. *Hippocampus*, *20*(11), 1195–1205. doi:10.1002/hipo.20854

Yang, C., Potts, R., & Shanks, D. R. (2017a). The forward testing effect on self-regulated study time allocation and metamemory monitoring. *Journal of Experimental Psychology: Applied, 23*(3), 263-277. doi:10.1037/xap0000122

Yang, C., Potts, R., & Shanks, D. R. (2017b). Metacognitive unawareness of the errorful generation benefit and its effects on self-regulated learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(7), 1073-1092. doi:10.1037/xlm0000363

Yang, C., Sun, B., & Shanks, D. R. (in press). The anchoring effect in metamemory monitoring.

Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Learning, Memory and Cognition, 41*(3), 597-613. doi:10.1037/xlm0000064

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recogition and source-memory judgements: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/0278-7393.25.6.1415

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517. doi:10.1006/jmla.2002.2864

Yonelinas, A. P. (2001). Components of episodic memory: The contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *356*(1413), 1363–1374. doi:10.1098/rstb.2001.0939

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1341–1354. doi:10.1037/0278-7393.20.6.1341

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497-1517. doi:10.1111/j.1751-9004.2008.00109.x