Universidade do Minho
Escola de Psicologia

Pedro Simão Oliveira Mendes

(Meta)memory: Effects of word frequency on judgments of learning and recall

dezembro de 2020

(Meta)memory: Effects of word frequency on judgments of learning and recall
Pedro Simão Mendes
UMinho | 2020

FCT Fundação para a Ciência e a Tecnologia

Cofinanciado por:

COMPETE 2020

PORTUGAL 2020

UNIÃO EUROPEIA
Fundo Social Europeu

Universidade do Minho
Escola de Psicologia

Pedro Simão Oliveira Mendes

(Meta)memory: Effects of word frequency on judgments of learning and recall

Tese de Doutoramento
Doutoramento em Psicologia Básica

Trabalho efetuado sob a orientação do
Professor Doutor Emanuel Pedro Viana Barbas de Albuquerque
e do
Professor Doutor Karlos Luna Ortega

dezembro de 2020

Acknowledgments

I want to thank my supervisors for all their contributions to this dissertation. To Professor Pedro Albuquerque, for introducing me to memory research and inciting my interest in experimental psychology almost 10 years ago. To Professor Karlos Luna, for everything you taught me along this journey working on metamemory. Together, you provided me with the necessary tools to accomplish this work.

I would not have finished this journey alone. Thus, I must thank all of my lab and PhD colleagues, some of whom became close friends. Thank you: Joaquina Almeida, for all the laughs along the way – I hope Doctor Kermit still makes laugh in the future; Catarina, my lunch buddy and the most optimistic person I know, for always helping me put things into perspective; Marlene, for all the unwinding moments that boosted my motivation to endure this adventure – let us keep organizing scientific and social events together; Patrícia, for being my travel pal on memory research conferences; Ana Duarte, for sharing your dark sense humor with me, and helping me on psycholinguistics; Raquel, for your precious help with data collection; Daniela, Alexandrina, and Diogo, for all the adventures we shared, big and small. A special thanks to Helena, whose guidance helped me through the most challenging times.

A big special thanks to you, Monika, for the opportunity you gave me and all your patience. You welcomed me into Mannheim and taught me so much, making the experience of working with you one that I will never forget.

Thanks to all the friends I have made in Mannheim. Philipp, Tine, Michael, Lili and Nikoletta, my daily companions on the third floor, but also the rest of the gang, Alejandro, Malte, David, Ruben and Martin. You have made me feel at home. A special thanks to Sofia, who shared the feeling of being a stranger in Germany and many hours of swimming. Dear all, may our paths cross again in the future.

Obrigado à minha família, por todos os sacrifícios feitos ao longo dos anos para que pudesse chegar aqui. Pelos apoios e encorajamentos incondicionais. Sem vocês, nada disto teria sido possível.

Thank you, Daniel, the man of my life. For being there whenever I needed the most. For being by my side at my best and at my worst. For sharing the past four years with me. It has been a privilege for me, but a demanding time for both of us. Your patience and your love kept me going, and I am sure they will help me persevere in the future. Let this moment be the start of something new together.

Finally, I would like to thank the Portuguese Foundation for Science and Technology (FCT) for the grant (reference PD/BD/128460/2017) that made all this possible.

Statement of Integrity

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

(Meta)memória: Efeitos da frequência das palavras nos julgamentos de aprendizagem e na evocação

Resumo

As previsões do desempenho da memória – julgamentos de aprendizagem (JOLs, do Inglês *judgments of learning*) – são úteis para a monitorização da aprendizagem, uma capacidade integrante da nossa metamemória. A memória e a metamemória são afetadas pela frequência das palavras. Por exemplo, a nossa memória em tarefas de evocação é habitualmente melhor para palavras de alta frequência (comuns) do que para palavras de baixa-frequência (raras). A investigação tem mostrado que as pessoas julgam as palavras de alta-frequência como mais memoráveis (JOLs mais elevados) do que as de baixa-frequência. Os JOLs são baseados em processos guiados pela teoria, como crenças sobre como diferentes pistas afetam a memória, mas também em processos guiados pela experiência, como a facilidade em processar um item (i.e., fluência). Estudos prévios sobre o efeito da frequência das palavras nos JOLs defendem que este efeito se deve exclusivamente a crenças. Nesta dissertação, argumentámos contra a explicação deste efeito baseada somente em crenças, testando esta ideia num total de 12 experiências em que os participantes atribuíram JOLs a palavras de alta e baixa frequência. Num estudo piloto testámos se as pessoas identificariam a frequência ao atribuírem JOLs para listas de palavras de diferentes intervalos de frequência, mas a ausência de efeito indicou que a diferença entre palavras de baixa e alta frequência não foi adequada. No Estudo 1 testámos novamente a identificação da frequência usando extremos de frequência e testámos diretamente o contributo da fluência para o efeito da frequência nos JOLs. No Estudo 2, apresentámos informação (in)congruente sobre a frequência das palavras estudadas para explorar se os JOLs seriam mais baseados em crenças ou na experiência, tendo também manipulado as crenças dos participantes sobre os efeitos da frequência na memória. Finalmente, no Estudo 3 explorámos se a frequência das palavras afetaria os JOLs em conjunto com a pista adicional tamanho do tipo de letra. Os resultados desta dissertação sugerem que os processos guiados pela experiência (e.g., a fluência de processamento) explicam melhor o efeito da frequência das palavras nos JOLs do que processos baseados em crenças. Além disso, os resultados sugerem que as crenças sobre a frequência podem ser alteradas facilmente apresentando uma crença contrária, com impacto nos JOLs. Finalmente, este trabalho sugere que os efeitos da frequência prevalecem mesmo na presença da pista adicional tamanho do tipo de letra, indicando integração de diferentes pistas nos JOLs. O presente trabalho ajudou a clarificar as bases dos efeitos da frequência das palavras nos JOLs, contribuindo para uma melhor compreensão da metamemória.

Palavras-chave: frequência, julgamentos de aprendizagem, memória, metamemória

(Meta)memory: Effects of word frequency on judgments of learning and recall

Abstract

Predictions of future memory performance – judgments of learning (JOLs) – are a useful tool to monitor one's learning, an ability that is part of our metamemory. Both our memory and metamemory are affected by word frequency. For example, our memory in recall tasks is usually better for high-frequency (common) than low-frequency (rare) words. Previous research has shown that people judge high-frequency words as more memorable (with higher JOLs) than low-frequency words. People's JOLs are based on theory-driven processes, such as beliefs of how different cues affect memory, but also on experience-driven processes, such as the easiness of processing an item (i.e., fluency). Regarding the word frequency effect on JOLs, previous research has argued that it is exclusively based on beliefs. In the present dissertation, we argued against a belief-only explanation of the word frequency effect on JOLs and tested our view through a total of 12 experiments in which participants made JOLs for high- and low-frequency words. In a pilot study we tested if people identified word frequency when making JOLs for lists of words with different frequency intervals, but we found no effect of frequency on JOLs which indicated that the difference between high- and low-frequency words was not adequate. In Study 1, we tested if people could identify word frequency using extreme values of frequency and directly tested whether processing fluency was involved in the word frequency effect on JOLs. In Study 2, we presented (in)congruent prompts about the frequency of the studied words, to explore whether people would base their JOLs more on beliefs or on experience and also manipulated participants' beliefs regarding how word frequency affects memory. Finally, Study 3 explored whether word frequency would affect JOLs in combination with the additional cue font size. The results from this dissertation suggest that the word frequency effect on JOLs is better explained by experience-driven processes than theory-driven processes. Additionally, results suggest that beliefs about word frequency can be easily changed using a counter-belief, with an impact on JOLs. Finally, word frequency affected JOLs even when another salient cue (i.e., font size) was present, showing integration of different cues in JOLs. The present work helped clarifying the basis of the word frequency effect on JOLs, contributing to a better comprehension of metamemory.

Keywords: frequency, judgments of learning, memory, metamemory

Table of Contents

Abbreviations, Acronyms and Symbols

ANOVA – Analysis of Variance

APA – American Psychological Association

APS – Association for Psychological Science

BAWL-R – Berlin Affective Word List-Reloaded (a German lexical database)

C – a nonlinear component of the lens model, indicative of systematic covariation between JOLs and recall that cannot be attributed to a linear combination of the measured cues

CDSS – Center for Doctoral Studies in Social and Behavioral Sciences

dlex-DB – Deutsch Lexical Database (German)

EOL – Ease-of-learning

FOK – Feeling-of-knowing

fpm – frequency per million

G – matching index of the lens model, indicative of JOL resolution

Gamma – Goodman-Kruskal gamma correlation

HF – High-frequency

JCR – Journal Citation Reports

JOLs – Judgments of learning

LF – Low-frequency

P-PAL – Procura-PALavras (a Portuguese lexical database)

r_a – Pearson correlation

RCJ – Retrospective Confidence Judgments

R_{JOL} – the consistency with which the person weighs the cues when making JOLs (lens model)

R_{REC} – the linear predictability of memory performance on the basis of cues (lens model)

SJR – Scimago Journal & Country Rank

SNIP – Source Normalized Impact per Paper

SUBTLEX-PT – Subtitles Lexical Database (Portuguese)

TOT – Tip of the tongue

List of Figures

List of Tables

Thesis Overview

Metamemory – the thoughts we have about our memory and its functioning – is just one part of a broader concept that comprises the thoughts about our thoughts, defined as metacognition (Rhodes, 2016). The present work is focused on understanding the processes that guide people's judgments of learning – predictions of future memory performance (Nelson & Narens, 1990) – when studying common and rare words. This thesis is organized into three parts: an introductory theoretical part, a second one presenting the empirical studies conducted for this thesis, and a final part with conclusions and closing remarks about the present work.

The introductory part describes the relevance of the experimental study of metamemory and early research on the topic. In Chapter 1, we explore a theoretical framework and different theories on the understanding of metamemory and outline the different types of metamemory judgments explored in research. The following chapter is dedicated to judgments of learning, which are the core of the present work (Chapter 2). Then, we explain the state-of-the-art regarding the effects of word frequency, the variable manipulated in the empirical studies, on both memory and metamemory (Chapter 3). Finally, the motivation for the present work is clarified in Chapter 4, and a summary of the research presented in the empirical section is described in Chapter 5.

The second part of this dissertation contains all the 12 empirical experiments conducted for this thesis. First, an unpublished pilot experiment is presented in Chapter 6, followed by three published or submitted studies, the first two comprising four experiments each (Chapters 7 and 8), and the third comprising three experiments (Chapter 9). Each study has its own theoretical introduction, method, results, and discussion. Chapter 10 focuses on complementary data analyses which are relevant to the topic under study. The final part consists of a general discussion integrating results from all studies and a reflection about the contribution of the present work to understanding metamemory (Chapter 11), limitations and suggestions for future research (Chapter 12), and a brief final comment (Chapter 13).

PART I

INTRODUCTION

1. The Study of Metamemory

In the search for the foundation of knowledge through radical doubt, the philosopher René Descartes might have been one of the first to value the reflection about one's knowledge. In truth, he probably was relying on his metacognition, that is, the knowledge about cognitive phenomena (Flavell, 1979). In the self-reflective process of thinking about his own thoughts, Descartes reached the very well-known expression *cogito, ergo sum* (i.e., I think, therefore I am), which Metcalfe (2000) argued to be a misnomer, adding that "metacogito, ergo sum" (p. 197) would have been a more suitable conclusion.

Cognition involves different features such as memory, attention, or comprehension, so metacognition also comprises different aspects of such cognitive processes (e.g., metamemory, Flavell, 1971; meta-attention, Reisberg & McLean, 1985; metacomprehension, Maki & Berry, 1984). In this work, we will focus on the experimental study of metamemory, which started with the early studies of Joseph Hart (1965). Hart asked people to predict their subsequent memory performance on items that they could not recall at that moment, that is, he investigated feeling-of-knowing judgments. It was Flavell (1971), however, who later coined the term metamemory to describe the knowledge that children acquired on the functioning of memory. In fact, younger children are not so good to monitor the contents of their memory as older children or adults. Flavell and Wellman (1975) subsequent work on metamemory as a developmental domain of memory highlighted its interest and influenced later cognitive researchers (Dunlosky & Tauber, 2016; Dunlosky & Thiede, 2013).

After many years of research, the definition of metamemory has evolved, but the core meaning is still the same as first proposed by Flavell. According to Metcalfe and Dunlosky (2008), "metamemory refers to the processes and structures whereby people can examine the content of their memories, either prospectively or retrospectively, and make judgments or commentaries about them" (p. 149). This noteworthy ability is relevant in several domains of people's lives. We review some of these domains in the next section of this chapter.

1.1. Relevance of Studying Metacognitive Processes

Metacognition is a topic of interest not only on its own, but it also links different areas, such as decision-making and memory, or learning and motivation (Nelson & Narens, 1994). Metacognition can be applied to different areas, for example, in psychological therapy, to tackle the processes of maintaining problematic cognitions (see metacognitive therapy, Wells, 2011), in the field of decision-making and problem-solving, with implications for cyber-security, such as password selection (Luna, 2019) or phishing emails (Canfield et al., 2019), and implications for daily life tasks, such as medical and financial decisions (Dentakos et al., 2019). Additionally, findings from studies with neurological populations suggest that metacognition is related to higher cognitive processes, being linked to prefrontal and frontal cortical areas (Pannu & Kaszniak, 2005). However, most of the research on metacognition and metamemory is related to learning. Nelson and Narens (1994) said that metamemory could be considered the link between memory and learning. In fact, the importance of studying metamemory began with the interest in understanding how children learn and monitor their learning (Flavell, 1971), which may be relevant now more than ever because college admissions and job placements depend highly on success in the classroom (Soderstrom et al., 2016).

An example of the relevance of metamemory in the classroom is when a student is preparing for an exam. They must know how to monitor their learning to assess how well prepared (or not) they are. Metamemory is the tool students use to make that assessment. More importantly, metamemory guides their decisions during the learning process. Students may decide to focus on the topics they remember poorly, as to tackle their difficulties, or they may decide to dedicate their time to the topics they are already comfortable with, which may be necessary to pass the exam. Either way, metamemory research has important implications in the educational context (for a review, see Soderstrom et al., 2016). For example, students do not seem to engage in active learning strategies such as retrieval practice (Roediger & Karpicke, 2006). Instead, they report relying on other less effective strategies, such as rereading or highlighting (Dunlosky et al., 2013). These examples show that students seem to have beliefs about how memory and learning operate. Some of those beliefs may be inaccurate and can misguide their learning strategies, which ultimately could impair their learning. Nonetheless, metamemory has proven to be a useful tool that people may use to readjust strategies while learning. There is evidence that if people are given a chance to restudy some of the learned materials, their memory for those materials improves. For example, when making metamemory judgments and restudy choices for words presented in loud or quiet volume, people were more likely to choose to restudy words presented in quiet volume, which were those with lower metacognitive judgments (e.g., Rhodes & Castel, 2009). Memory improves for items that people choose themselves in comparison to items chosen randomly (e.g., Kornell & Metcalfe, 2006), reflecting the importance of monitoring ongoing learning in future memory performance.

Because of the relevance of metamemory in different domains, there has been an increase in metamemory research over the last decades (Kleitman & Narciss, 2019), which not only asserts its growing relevance but also illustrates the efforts in understanding metamemory processes and its functioning. Those efforts have produced empirical data and resulted in the development of different theoretical frameworks and models.

1.2. Theories of Metamemory

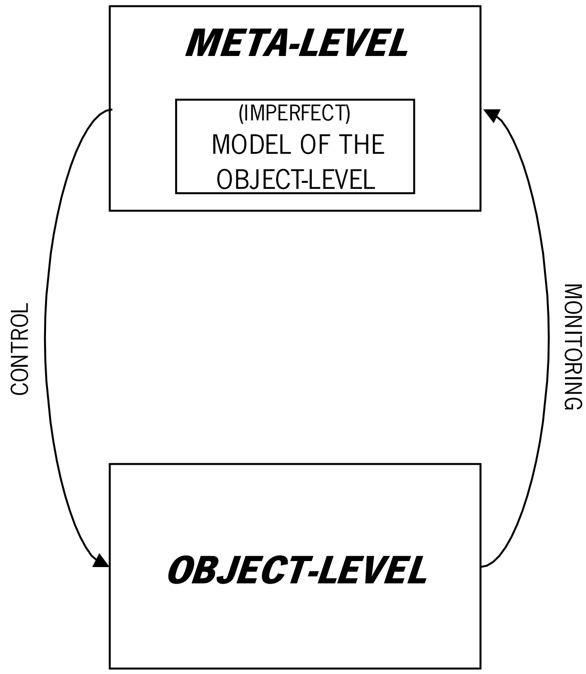
One of the first theories of metacognition was proposed by Flavell (1979). His model distinguished four components: Metacognitive knowledge, which are the beliefs or theories one could have about their cognition and abilities; metacognitive experiences, the conscious feeling of how well one would perform in a given task; goals (or tasks), which are the objectives of a given cognitive effort; and actions (or strategies), which are cognitions or behaviors employed to reach those goals. This pioneering model presented the key idea that metamemory directly influences the behaviors adopted during a given task, which was crucial in recognizing its importance. This is a notion that remains in more recent models.

Flavell's work increased the interest in the study of metacognition, and the subsequent body of research led to the formulation of the most influential model of its functioning. In 1990, Nelson and Narens presented a theoretical framework of metacognition based on three principles. The first states that cognitive processes can be organized in (at least) two interrelated levels: the meta-level and the object-level. The second principle defends that the meta-level contains a dynamic but imperfect model of the object-level. Finally, the third principle states that the flow of information between one level and another can be described as monitoring (i.e., the object-level informs the meta-level) and control (the meta-level modifies the object-level). Regarding metamemory, the object-level would comprise the memories themselves, while the meta-level would comprise the reflections and thoughts upon memories and ongoing learning. Figure 1 illustrates this model, which is still valid today. Nelson and Narens' (1990) theoretical framework was based on the research conducted until then, comprising several types of metamemory judgments. Indeed, these authors organized different metamemory judgments around the different memory processes (i.e., acquisition, retention, and retrieval), as depicted in Figure 2. For example, regarding acquisition, ease-of-learning judgments are requested before learning, while judgments of learning are requested during learning and are also related to retention of information. Feeling-of-knowing judgments monitor processes of both acquisition and retrieval, and confidence judgments are focused on the retrieval of information after it happened. These and other metamemory judgments will be discussed in the next section.

Considering this general framework of metamemory functioning, the processes behind different metamemory judgments are often explored considering two major theoretical approaches, which are

Figure 1

Illustration of Nelson and Narens' (1990) Metamemory Framework.

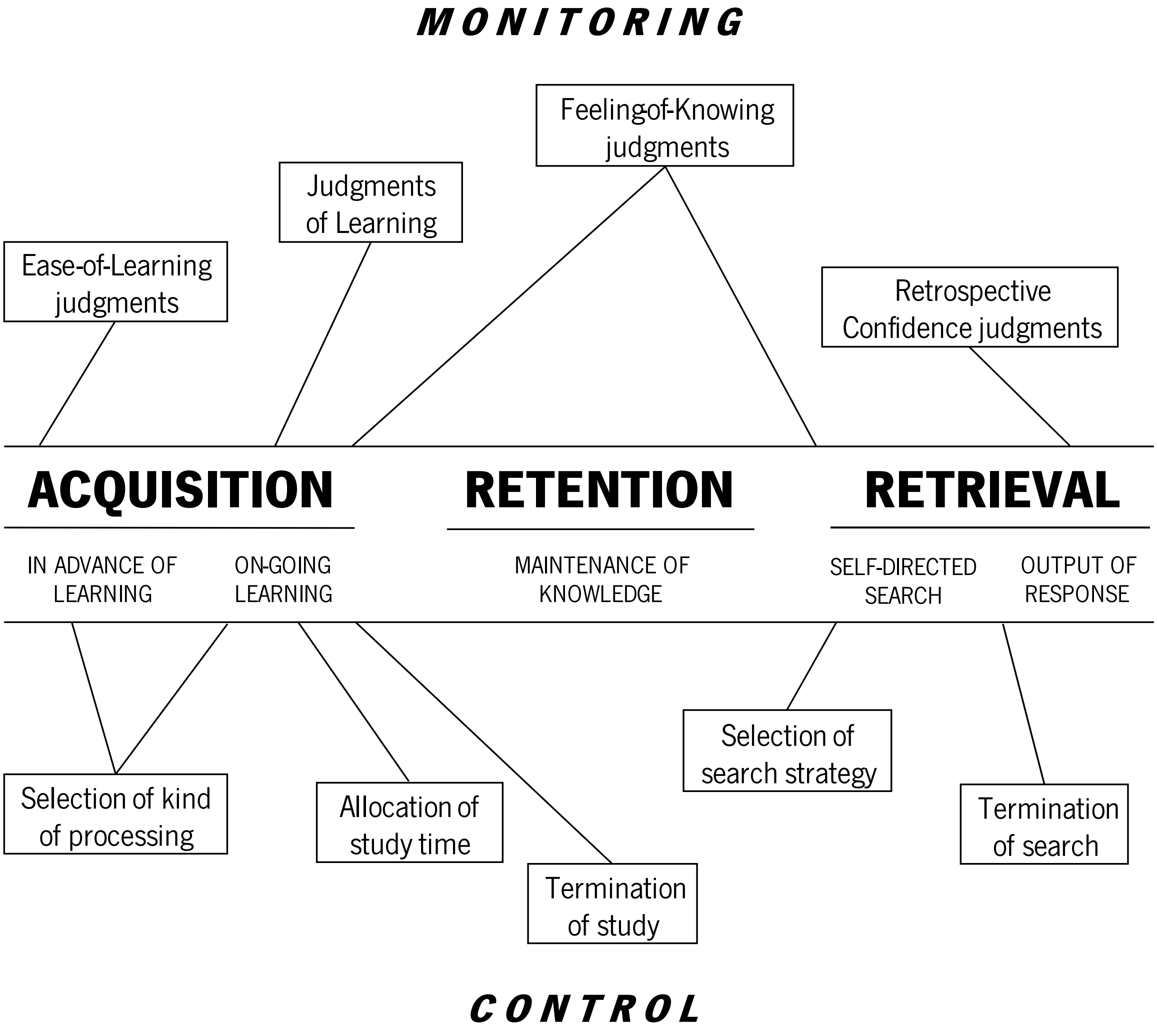


Notes: In Nelson and Narens' (1990) metamemory framework, information flows from the object-level to the meta-level through monitoring (e.g., changes in the state of the meta-level's model of the situation); the meta-level modifies the object-level through control processes (e.g., initiate or stop an action). Figure adapted from Nelson and Narens (1994).

not mutually exclusive: A direct-access approach and an inferential approach. The former implies that people have direct access to the strength of their memory traces and make metamemory judgments accordingly; the latter claims that people's metacognitive judgments are based on other sources of information related to the materials to be retrieved, but not their memory trace (Rhodes, 2016; Schwartz et al., 1997). Within these two major approaches, researchers have proposed specific theories that apply to the processes involved in different metamemory judgments. For example, consider theories first developed for the feeling-of-knowing judgments and tip-of-the-tongue states (see below for a brief review of these phenomena). On the one hand, the target retrievability hypothesis postulates that people can accurately predict their future knowledge based on how accessible a memory trace is (e.g., Burke et al., 1991), an idea that comes from earlier research (e.g., Hart, 1965) and which is based in the direct-access approach of metamemory. On the other hand, the cue familiarity hypothesis, which is based on a framework proposed by Reder (1987), states that metamemory judgments are made in accordance with how familiar a cue is to a person, with more familiar cues leading to the feeling that a given memory trace is stored in memory, which in turn

Figure 2

Relation of Different Metamemory Judgments and Behaviors in Relation to Stages of Memory Processes.



Notes: The figure depicts different metamemory judgments or monitoring components (top) and different decisions or control components (bottom) in relation to the different memory processes (middle). Figure adapted from Nelson and Narens (1994).

increases metamemory judgments (e.g., Metcalfe et al., 1993). From the inferential approach, the cue accessibility hypothesis proposes that judgments are inferences based on both the amount of information available (e.g., number of cues) and the intensity of the activated memory traces (e.g., how accessible the memory trace is) when the judgment is made (Koriat, 1993, 1995). Additionally, the competition hypothesis considers that metamemory judgments are inversely related to the number of competitors for a given memory trace (Schreiber, 1998). More recently, Koriat and Levy-Sador (2001) argued that the feeling-of-knowing judgments are based on both the cue familiarity and the cue accessibility hypotheses: Judgments would rely first on cue familiarity, and, if cue familiarity fails in providing enough information to make a judgment, then on cue accessibility.

Some of these theories were also tested with other metamemory judgments. For example, Koriat (1997) argued that judgments of learning (see Chapter 3) are also made through inferential processes based on available cues, which seems to be true for ease-of-learning judgments (see next section) as well (e.g., Jönsson & Linström, 2010). McGuire and Maki (2001) also explored the competitor and accessibility hypotheses' role in the judgments of learning. The multitude of paradigms and types of metamemory judgments that have been explored contributed to a better understanding of metamemory as a whole. Furthermore, evidence supports theories proposing that metamemory judgments are inferential in nature, and nowadays the direct-access view is not favored by most researchers (Rhodes, 2016). In fact, this inferential nature seems to be transversal to different metamemory judgments, as summarized in the next section.

1.3. Types of Metamemory Judgments

When assessing metamemory experimentally, people are usually asked different kinds of judgments that monitor different memory stages, such as the acquisition, retention, or retrieval of the study material (Nelson & Narens, 1990). Metamemory judgments include ease-of-learning (EOL) judgments (Underwood, 1966), feeling-of-knowing (FOK) judgments (Hart, 1965), tip-of-the-tongue (TOT) states (Brown & McNeill, 1966), retrospective confidence judgments (Lichtenstein et al., 1982), remember/know judgments (Tulving, 1985), source judgments (Schwartz, 1994), and judgments of learning (Arbuckle & Cuddy, 1969). A general overview of these judgments is presented below in the order of the memory processes to which they relate (for a review, see Metcalfe & Dunlosky, 2008; see also Nelson & Narens, 1990). However, judgments of learning (JOLs), which are predictions of future memory performance during the acquisition of information (e.g., Rhodes, 2016), are the focus of the present work and will be explained more extensively in Chapter 3.

1.3.1. Ease-of-Learning Judgments

EOL judgments consist of assessments of item difficulty before the acquisition of study materials (Nelson & Narens, 1990). Usually, people make EOL judgments for each item or set of items by rating how easy or difficult it will be for them to learn that item or items. These judgments are said to be accurate predictors of learning rate (Underwood, 1966) and are related to study time allocation (Nelson & Leonesio, 1988), demonstrating how perceived difficulty ratings can guide people's learning strategies and behaviors. More specifically, after judging which items are more challenging to learn, participants can decide which items they prefer to study more. There is evidence that the ease of processing an item (i.e., processing fluency) guides EOL judgments, but they can also be based on beliefs that people have regarding items' characteristics that make them easy or difficult to learn (Jemstedt et al., 2018).

1.3.2. Feeling-of-Knowing Judgments and Tip-of-the-Tongue States

FOK judgments were the first metacognitive judgments to be explored experimentally (Hart, 1965). In Hart's study, participants were asked general knowledge questions. When failing to remember the answer, participants should indicate if they had the feeling of knowing the correct answer even though it could not be retrieved at that moment. These judgments were predictive of memory performance in a subsequent recognition test, with participants being able to predict their performance with above-chance accuracy. Hart argued that FOK judgments were made based on direct access to the memory trace and worked as an indicator of its storage in memory. However, his view was not supported by subsequent research. For example, there is evidence that FOK judgments are affected by previous exposure to concepts, suggesting that cue familiarity influences these judgments (Metcalfe et al., 1993; Reder, 1987). Koriat (1993) found that when making FOK judgments, participants could often retrieve parts of studied items and proposed a partial target retrievability hypothesis instead. More recently, Koriat and Levy-Sadot (2001) proposed a cascade model in which FOK judgments are affected by cue familiarity first, with low familiarity leading to faster "no" FOK judgments; but when familiarity is enough to induce a positive FOK, an inferential process based on retrievability of the item will follow.

Feeling-of-knowing judgments are related to a commonly known phenomenon anyone can face in their daily lives: the tip-of-the-tongue (TOT) state. TOT states consist of the subjective feeling that the recall of an item is imminent, but a person is not able to produce an answer (Brown, 1991). While the feeling-of-knowing judgments are elicited by a question to assess the likelihood of recognizing an item that cannot be retrieved when the judgment is made, the TOT state is involuntary. Nonetheless, it is possible to elicit TOT judgments experimentally (for a review, see Brown, 1991). For example, in the first experimental study to explore TOT states (Brown & McNeill, 1966), participants heard definitions of rare words with the instruction to write the word if they knew it, write nothing if they did not, or indicate if they were in a tip-of-the-tongue state. When in a TOT state, participants should provide additional information about the target word, and participants were able to retrieve partial information (e.g., the first letter). Hence, explanations for the TOT states are in line with those presented for FOK judgments in terms of cue familiarity (Metcalfe et al., 1993), but also in terms of partial target accessibility (Koriat, 1993). There are, however, specific models developed to explain TOT states (e.g., Burke et al., 1991;

see also Metcalfe & Dunlosky, 2008) and authors supporting the dissociation between FOK and TOT judgments (e.g., Schwartz, 2006).

The study of FOK and TOT judgments allowed the development of theories about metamemory functioning, but the impact of both kinds of judgments on decision-making and retrieval is still unclear (Metcalfe & Dunlosky, 2008). There are other metamemory judgments that better evaluate how people monitor the correct retrieval of information, such as the remember/know judgments and retrospective confidence judgments (RCJ).

1.3.3. Remember/Know Judgments

According to Metcalfe and Dunlosky (2008), remember/know judgments consist of the distinction between what people can recollect, including the study material and the circumstances of having learned it (i.e., remember), and what people just know or are familiar with (i.e., know). While making remember/know judgments, people assess how confident they are that they remember learning an item, or rather just have the feeling that the item was presented before, which typically happens during recognition tasks (but see Mickes et al., 2013). These judgments are related to dual-process models of recognition, which argue that both recollection and familiarity influence recognition decisions (Metcalfe & Dunlosky, 2008). Alternatively, an explanation based on the signal-detection theory argues that remember judgments reflect stronger memory traces than know judgments (e.g., Wixted & Stretch, 2004), which has found some evidence in experimental psychology but little in the neuroscience literature (Wixted, 2009). Wixted and Mickes (2010) proposed a new signal detection model that does not deny the validity of dual-process approaches and instead tries to tackle the flaws of both dual-process and prior signal-detection theories. There is evidence that not only familiarity derives from inferential processes, but that recollection may also be partially based on inferential processes as well (Kurilla & Westerman, 2008).

Remember/know judgments were first introduced by Tulving (1985) when testing the role of consciousness on the distinction between episodic and semantic memory. Research using remember/know judgments is vast, and several factors seem to affect those judgments, including encoding or retrieval manipulations, but also different stimulus variables or special populations (for a review, see Yonelinas, 2002). For example, regarding encoding manipulations, Gardiner (1988) found a generation effect (i.e., better memory for items produced instead of read) for recollection (i.e., “remember” answers), but not for familiarity (i.e., “know” answers), results interpreted as consistent with dual-process theories of recognition. In another example, Hicks and Marsh (1999) compared remember/know judgments made after a typical old/new recognition test vs. made simultaneously in the recognition task (remember/know/new). The authors found that in the latter condition, participants were biased into labeling more items as remembered, suggesting that monitoring of learned materials was also affected by testing methodology. Research with remember/know paradigms helps to clarify the metacognitive ability to distinguish between remembering and knowing a memory, which is particularly important to understand the role of consciousness in memory (Metcalfe & Dunlosky, 2008).

1.3.4. Retrospective Confidence Judgments

RCJ are judgments that occur after retrieval and assess the confidence that the recalled information is correct (Nelson & Narens, 1990). These confidence ratings are thought to arise from the strength of the memory trace being retrieved (e.g., Roebers & Howie, 2003). Asking RCJ makes it possible to evaluate how well people monitor the correctness of what they recalled, indicating their accuracy. Initial studies focusing on these judgments showed that people tended to be overconfident (Lichtenstein et al., 1982), evaluating recalled information as correct more often than it was. In other words, people tend to overestimate how much they know, which can impair their judgment, making claims which may be wrong (see also the Dunning-Kruger effect, Kruger & Dunning, 1999). Examples of such claims include wrongly identifying a person as a perpetrator of a crime or advocating against vaccines.

There has been considerable research on RCJ. For example, RCJ have shown to be relevant in terms of evaluating metamemory development, with studies often comparing the metacognitive abilities of children and adults (e.g., Ghetti et al., 2008; Roebers, 2002; Roebers & Howie, 2003). Confidence ratings also guide future study strategies (e.g., Robey et al., 2017) and their role on memory was compared to other metamemory judgments such as JOLs, which also seem to impact memory (Dougherty et al., 2005). Additionally, RCJ have been important in exploring eyewitness accuracy (e.g., Deffenbacher, 1980; Luna & Martín-Luengo, 2010; Migueles & Garcia-Bajos, 1999). Accuracy of confidence ratings in eyewitness research seems to be lower than confidence for other types of information, suggesting that RCJ may depend on the material. For example, when comparing the accuracy of RCJ for general knowledge questions and eyewitness memory questions, Luna and Martín-Luengo (2012) found strong confidence–accuracy relationship with both general knowledge and eyewitness memory questions, but it was stronger with the former than the latter.

Theories for the basis of RCJ include dual-process theories, as remember/know judgments, with high confidence being linked to recollection and low confidence being linked to familiarity (Ghetti &

Angelini, 2008). There are also theories that defend that RCJ reflect linear changes in the strength of the memory trace (e.g., Macmillan & Creelman, 2005). Despite this, according to Ghetti et al. (2008), the tradition in metacognitive research is that the relation between RCJ and memory representations depends on different conditions, including the type of material or age of participants. Nevertheless, as with other metamemory judgments, RCJ seem to be based on inferential processes (Leippe & Eisenstadt, 2007).

Except for EOL judgments, all the aforementioned metamemory judgments evaluate, to some extent, the existence or absence of a memory trace for the studied materials. However, there is one kind of judgment that focuses on the origin of a memory rather than on the memory trace itself, which are the source judgments.

1.3.5. Source Judgments

In their daily lives, people need not only to monitor what they learn but also from where they learn it. Did someone learn that the prolonged use of surgical masks causes oxygen deficiency from their doctor or from social media? If they saw that information in some video on social media, but they think they heard it from a doctor, they may believe it more than would be advisable. The metacognitive judgments that refer to a thought or memory's origin, rather than the memory trace itself, are called source judgments (Johnson et al., 1993; Lindsay, 2008; Schwartz, 1994). The study of these judgments is useful to better understand situations where source monitoring fails, and the source of a memory is misattributed, such as unconscious plagiarism or interviewing crime eyewitnesses (Metcalfe & Dunlosky, 2008). Source judgments can be guided by memory traces which retain contextual details (e.g., where, when) or by previous knowledge and beliefs, such as stereotypes (e.g., Wulff & Scharf, 2020). For example, in one study participants who could not remember the source of a statement tended to guess the source based on age-stereotyped information (Kuhlmann et al., 2016). Such evidence supports that source judgments can be, to some extent, inferential. Research about source monitoring is extensive, but source judgments are rarely referred to as metacognitive judgments in empirical studies (for a detailed review, see Lindsay, 2008).

In sum, research has explored different metamemory judgments which monitor different memory processes. The present work is mostly interested in monitoring ongoing learning, which is commonly assessed through judgments of learning.

2. Monitoring Learning Through Judgments of Learning

According to Nelson and Narens (1990, 1994), judgments of learning (JOLs) are monitoring mechanisms that assess ongoing learning during the acquisition of information and knowledge maintenance during retention. In other words, JOLs are metacognitive judgments that assess future memory performance and are useful to monitor learning (Metcalfe & Dunlosky, 2008; Rhodes, 2016). In a typical JOL task, people are asked the likelihood of remembering specific information on a future memory task. Although studies have been using different scales to request JOLs, a popular alternative is the probability or percentage of later successful memory retrieval (Rhodes, 2016). Other differences between studies include the moment of requesting JOLs. Immediate JOLs are made directly after studying each stimulus (most commonly a word or a word pair), while delayed JOLs are requested after some time. Castel (2008) introduced pre-study JOLs, which are requested before presenting the stimulus by providing information about the nature of the to-be-studied item (e.g., the frequency of a word). Besides item-by-item requests, JOLs can also be requested in an aggregate/global manner by asking, for example, how many items, in total, one will remember (Rhodes, 2016). Because of their relevance to monitoring learning, one of JOL research's critical aspects is to assess their accuracy, that is, how well people can predict their memory performance (see Rhodes, 2016).

2.1. JOL Accuracy

One of the first studies on JOLs aimed to assess how accurately participants' memory predictions were when attending to different associative strengths of paired associates (Arbuckle & Cuddy, 1969). According to Rhodes (2016), there are two ways to assess the accuracy of a metamemory judgment. First, absolute accuracy, or calibration, refers to the overall correspondence between JOLs and memory performance. More specifically, if a person has an average JOL of 60%, their judgments would be entirely accurate only if they correctly remember 60% of the stimuli they learned. An average judgment above the actual memory performance would represent overconfidence, and an average judgment below actual memory performance would represent underconfidence (Rhodes, 2016). Second, relative accuracy, or resolution, describes whether JOLs can discriminate between items that will be remembered or not. Nelson (1984) compared different measures of resolution and recommended its calculation through the Kruskal-Goodman gamma correlation (Goodman & Kruskal, 1954). Gamma correlation varies from -1, when all remembered items received lower JOLs than all items that were not remembered, to +1, when all remembered items received higher JOLs than all items that were not remembered (Rhodes, 2016). A gamma correlation of approximately zero is interpreted as an absence of relationship between JOLs and memory performance. Thus, JOLs are accurate when gamma is closer to +1, suggesting good discrimination between which items will be remembered and those that will not.

Despite their recurrent use as an indicator of resolution, gamma correlations have received several criticisms, and other measures have been proposed (Bröder & Undorf, 2019). For example, Masson and Rotello (2009) found that gamma correlations varied with response bias and proposed signal detection measures as an alternative. Higham et al. (2016) also argued in favor of a signal detection framework, which allows analyzing both calibration and resolution of JOLs. Another alternative is the recent approach proposed by Bröder and Undorf (2019), who used Brunswik's (1952) lens model when multiple cues that may affect JOLs and their accuracy were manipulated. Although it was developed as a framework to describe how people make judgments in a perceptual environment, Brunswik's (1952) lens model could be applied to any judgmental task, including metacognitive judgments (for a review, see Karelaia & Hogarth, 2008). Typically, in perceptual studies, people make judgments about distal variables in the environment that are not directly available (e.g., the size of a distant object), so must be inferred. Thus, people rely on probabilistic cues, such as linear perspective or retinal size (see Hammond & Stewart, 2001) to make those judgments. Applying this rationale to metamemory, the environment to be predicted is memory performance and JOLs rely on the cues available at study (cf. Bröder & Undorf, 2019). According to Bröder and Undorf (2019), the lens model in metamemory research has the potential to account for cue redundancy and a large number of cues when exploring the variance of JOLs accuracy because of its regression-based analysis. These authors applied the model to research manipulating multiple cues and concluded that the lens model's matching index *G* was a measure more valid of JOL resolution than the gamma coefficient that is commonly used. For a more detailed explanation of the lens model applied to JOLs, please see Study 3 (Chapter 9).

Several factors affect both types of accuracy of JOLs. Rhodes (2016) stated that any factor influencing the magnitude of JOLs and/or the probability of remembering target information will influence absolute accuracy. For example, Rhodes and Castel (2008) found that words presented in larger font size (48 pt) were given higher JOLs than words presented in a smaller font size (18 pt) but found no effect of font size on memory. They also found an overall overconfidence of people's JOLs for larger font sizes. Relative accuracy is also affected by different factors (see Rhodes, 2016) such as practice across multiple study-test cycles or the timing of JOLs. For example, after a few study-test cycles, participants learn to give higher JOLs for first and last items, which reflects an increase of JOLs accuracy regarding the serial position effects on recall (Castel, 2008). Similarly, in a study with multiple study-test cycles, people tended to be overconfident on a first study-test cycle, but in the following cycles, they tended to become underconfident, an effect known as the underconfidence-with-practice effect (Koriat et al., 2002). In that study, that was true for absolute accuracy but not for relative accuracy, which generally improved across trials. Regarding the timing of JOLs, it is known that when compared to immediate JOLs, delayed JOLs tend to show higher accuracy (Nelson & Dunlosky, 1991). These authors suggested that the most likely explanation for this delayed-JOL effect is related to the fact that when JOLs are delayed, people attempt to retrieve information from long-term memory, but for immediate JOLs, the retrieval attempt is made from information in the short-term memory. Because the content from short-term memory is not diagnostic of future retrieval, accuracy for delayed JOLs is higher. For a review of other theories on the delayed JOL effect and the underconfidence-with-practice effect, see Metcalfe and Dunlosky (2008).

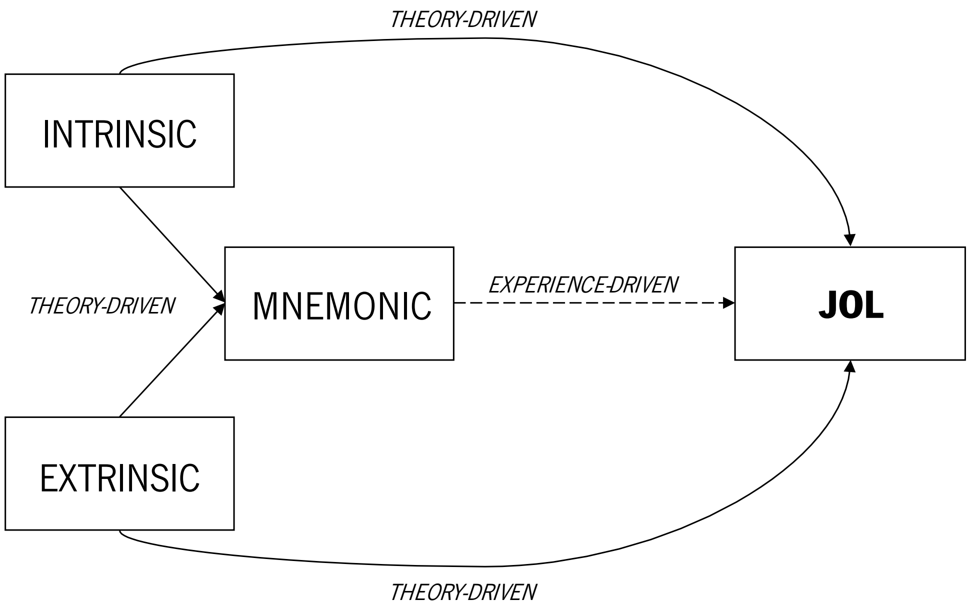
To better understand which variables influence JOLs' accuracy, it is also essential to understand how JOLs are made and the underlying mechanisms of these judgments. This has been another key aspect of JOL research.

2.2. Mechanisms Underlying JOLs

As with other metacognitive judgments, explanations of the processes underlying JOLs could be classified in terms of direct access to memory (e.g., Arbuckle & Cuddy, 1969) or in terms of inferential accounts (e.g., Koriat, 1997). Direct access explanations rarely found empirical support (Rhodes, 2016), so JOLs are mostly understood in the light of inferential processes. Inferential accounts propose that JOLs are based on the different information available during learning, namely, cues and heuristics (Rhodes, 2016). In his cue-utilization framework, Koriat (1997) argues that when making JOLs people do not directly monitor the strength of a memory trace but rather use the information from cues available at the time, which informs about the probability of retrieving the information later. The author classified these cues as intrinsic, extrinsic, and mnemonic. Intrinsic cues are those inherent to the study materials, such as word frequency (e.g., Begg et al., 1989) or the concreteness of a word (e.g., Witherby & Tauber, 2017). Extrinsic cues are those related to the study conditions or encoding operations of the learner, such as the number of presentations of a word (Kornell et al., 2011), presentation time (Mazzoni et al., 1990), or interactive imagery (Begg et al., 1989). Mnemonic cues are subjective internal factors that indicate if an item has been learned or not (e.g., subjective feelings of encoding fluency). Intrinsic and extrinsic cues can influence JOLs directly or indirectly through mnemonic cues, as depicted in Figure 3.

Figure 3

Koriat's (1997) Representation for the Cue-utilization Approach for Judgments of Learning (JOL).



Notes: Intrinsic (e.g., word frequency) and extrinsic (e.g., number of presentations) cues affect JOLs directly through theory-driven processes such as beliefs and knowledge, but also indirectly through mnemonic cues based on the direct experience with the study items. Figure adapted from Koriat (1997).

There are several cues affecting JOLs, which, among others, include word frequency (e.g., Begg et al., 1989), valence and arousal (e.g., Hourihan et al., 2017), font size (e.g., Rhodes & Castel, 2008), anticipated test format (e.g., Thiede, 1996), number of presentations (e.g., Undorf et al., 2018), concreteness (e.g., Witherby & Tauber, 2017), animacy (Li et al., 2016), the relatedness of word pairs (e.g., Mueller et al., 2013), and motoric fluency (Susser & Mulligan, 2015). However, not all these cues are equally diagnostic of memory performance. For example, emotional words are usually better remembered and are given higher JOLs accordingly (e.g., Hourihan et al., 2017). As previously mentioned, the font size has little to no impact on memory and still affects JOLs (Luna et al., 2018; Rhodes & Castel, 2008). There is also the case of retention interval: Memory declines with time, but JOLs remain stable for different retention intervals (Koriat et al., 2004). Basing JOLs on cues which have no impact on memory could account for JOLs' low relative accuracy and justify the existence of metamemory illusions such as the font-size effect (Rhodes, 2016). Indeed, after the reports of the font-size effect (Rhodes & Castel, 2008), it has become more relevant to understand how people use different cues to make JOLs.

The processes by which different cues, either intrinsic, extrinsic, or mnemonic, affect JOLs can be divided into theory-driven processes and experience-driven processes. Theory-driven processes are those that originate from an analytic inference of how different intrinsic and extrinsic cues could affect memory and consist of beliefs and knowledge; experience-driven processes are those arising from the direct experience with the items, resulting in a subjective feeling of how easy the processing of an item was, and are said to be automatic (Kelley & Jacoby, 1996; Koriat, 1997). These concepts are somewhat an update of Flavell's (1979) framework for metamemory, in which theory-driven processes could be considered akin to metacognitive knowledge, while experience-driven processes could be considered akin to metacognitive experiences.

Over the last decade, literature has been debating whether JOLs are based on experience-driven processes, on theory-driven processes, or both (Dunlosky et al., 2015). There is evidence from studies exploring processing fluency with different cues arguing that experience-driven processes affect JOLs. Auditory volume (Rhodes & Castel, 2009), semantic relatedness (Undorf & Erdfelder, 2015), or generation of images (Besken, 2016) are examples of such cues. A common way to assess differences in fluency is by measuring participants' response times in lexical decision tasks (e.g., Luna et al., 2019; Mueller et al., 2014; Undorf & Erdfelder, 2015; Undorf & Zimdahl, 2019), naming tasks (e.g., Susser et al., 2016), or study time during encoding (e.g., Jia et al., 2016; Undorf & Ackermann, 2017; Undorf & Erdfelder, 2015). Higher fluency in processing an item is translated in less time to study that item or in faster response times when naming or identifying it. Items processed more fluently, such as related word pairs or words presented in higher volume, would create a subjective feeling that the item has been learned and could be recalled in a future memory test (Koriat, 1997). For example, the higher JOLs attributed to words in larger font size was firstly attributed to differences in perceptual fluency between large and small font sizes because the effect disappeared when words were presented in aLtErNaTiNg format, which was less fluent (Rhodes & Castel, 2008).

Thus, fluency has proven to be an important cue to JOLs. There are, however, different ways to manipulate fluency. Alter and Oppenheimer (2009) reviewed different kinds of fluency that may require different measures to be assessed. Examples include perceptual fluency, linguistic fluency, and conceptual fluency. Perceptual fluency refers to the easiness of perceiving stimuli and is mostly assessed with visual manipulations, such as easy- vs. difficult-to-read font styles (e.g., Reber & Zupanek, 2002). Linguistic fluency is related to the processing of linguistic information and comprises phonological, lexical, syntactic, and orthographic fluency. These could be manipulated through pronounceability (e.g., Alter & Oppenheimer, 2006), word complexity (Oppenheimer, 2006), syntax complexity (Lowrey, 1998), or the use of written symbols to replace letters (Alter et al., 2007), respectively. Conceptual fluency is related to how concepts are associated or not and is often manipulated through semantic priming (e.g., Susser et al., 2016).

Alter and Oppenheimer (2009) reported that manipulating different types of fluency (e.g., perceptual fluency and conceptual fluency) in a similar way (e.g., clear vs. blurred texts) seem to converge on the same behavioral and cognitive outcomes (see also Graf et al., 2018). However, there is recent evidence suggesting that different types of fluency affect different types of judgments distinctively. For example, Vogel et al. (2020) found evidence supporting their fluency-specificity hypothesis, in which the effects of fluency on judgments are better explained when considering the fit between the judgment to be made (perceptual or conceptual judgment) and the process generated by the fluency experience (perceptual or conceptual processing). More specifically, results showed that judgments of truth were strongly affected by content repetition but not so much by visual contrast, and that aesthetic evaluations were strongly influenced by visual contrast but not by repetition of content (Vogel et al., 2020). Likewise, it is possible that different manipulations of fluency and how they are measured could also affect JOLs differently.

Another important idea is that fluency may also affect metacognitive judgments through the mediation of naïve theories that people have (Alter & Oppenheimer, 2009), which seem to account for different interpretations of how fluency affects different judgments (see also Schwartz, 2004). For example, Susser et al. (2016) found that although identity priming affected fluency (measured by naming latencies), it only affected participants' JOLs when primes were made evident (i.e., when the prime word was presented for 200 ms instead of 32 ms). These results suggest that fluency only affected JOLs when it was possible for participants to use beliefs of how it would affect memory. Besides, research with multiple study-test cycles indicates that JOLs may be updated from one cycle to the following (e.g., Benjamin, 2003; Castel, 2008), suggesting that experience could be incorporated into people's knowledge and beliefs, which would, in turn, affect subsequent JOLs.

In fact, there are several studies showing that JOLs are based on theory-driven processes, such as previous knowledge or beliefs of how a specific cue will affect memory performance (Jia et al., 2016; Koriat et al., 2004; Mueller et al., 2014; Susser et al., 2016; Undorf & Zimdahl, 2019). These beliefs can be assessed by asking participants to estimate memory performance in a hypothetical experiment or through pre-study JOLs (e.g., Castel, 2008; Mueller et al., 2014; Jia et al., 2016; Witherby & Tauber, 2017). In particular, pre-study JOLs proved to be a useful tool in support of this perspective. When people make pre-study JOLs, they do so before processing the items to be learned, which rules out any possible contribution of direct experience from processing the item. Even so, font size (e.g., Mueller et al., 2014), word frequency (e.g., Jia et al., 2016), or concreteness (e.g., Witherby & Tauber, 2017) still affected pre-study JOLs, which could be only due to beliefs about the effect of those variables on memory.

The analytic-processing theory (Dunlosky et al., 2015; Mueller et al., 2016; Mueller et al., 2013) is a recent approach based on Koriat's (1997) cue-utilization approach that favors theory-driven processes over experience-driven processes. This theory argues that theory-driven processes are the main basis of JOLs and adds that, besides having a priori beliefs, people may develop online beliefs which could be triggered by the task of making JOLs, causing people to search for variability in items and identify cues that could impact memory. For example, the authors argued that beliefs about fluency would guide its effects on JOLs. To explore this, Mueller and Dunlosky (2017) lead participants to believe that one color was processed more fluently than another. This manipulation carefully made no reference to how color affected memory, but instead focused only on the effects of processing fluency for different colors. When there was no suggestion that one color was more fluently processed than the other, color of presentation did not affect JOLs. However, when the instructions suggested that one color was more fluently processed participants' JOLs were higher for words presented in that color, despite it not having any impact on memory (Mueller & Dunlosky, 2017). These results suggest that people's beliefs about fluency (and not fluency per se) were the main basis of JOLs.

An additional finding supporting the effect of beliefs on JOLs is related to the introduction of a counter-belief, that is, information that goes in the opposite direction of the usually observed effect of a cue on JOLs. For example, while exploring the font-size effect on JOLs, Blake and Castel (2018) found that presenting a counter-belief (e.g., "memory is better for smaller font sizes") reduced the font size effect on JOLs, but presenting a congruent belief (e.g., "memory is better for larger font sizes") did not increase the effect size of the usually observed effect. It is not clear yet why presenting congruent beliefs do not affect JOLs, but the effect of counter-beliefs shows the relevance of beliefs and how easy they are to create and manipulate, in the direction of which information will be remembered later.

It is difficult to disentangle the contribution of experience- and theory-driven processes. The development of new techniques or creative methodologies to better understand the contribution of both types of processes is crucial to solve this issue. For example, Price and Harrison (2017) collected pre-study JOLs, immediate JOLs, or both while manipulating item relatedness and font size. They found that both pre-study and immediate JOLs were affected by both cues in all conditions. Additionally, they found that for participants that made both JOLs, immediate JOLs were higher in magnitude and relative accuracy than pre-study JOLs, suggesting an important role of direct experience with the items, which memory beliefs alone cannot explain (Price & Harrison, 2017). Indeed, recent research supports a joint contribution of theory-driven and experience-driven processes on the effects of different cues on JOLs (e.g., Blake & Castel, 2018; Frank & Kuhlmann, 2017; Hu et al., 2015; Koriat et al., 2004; Price & Harrison, 2017; Rhodes, 2016; Undorf et al., 2017; Yang et al., 2018). For example, the volume effect on JOLs was found to be based on both types of processes (Frank & Kuhlmann, 2017), as well as the font size effect (Undorf et al., 2017; Yang et al., 2018).

The contribution of experience- and theory-driven processes may diverge for different cues. For example, it has been proposed that beliefs may play a higher role in the font size effect (see Luna et al., 2019), but word frequency has been argued to affect JOLs exclusively through beliefs (Jia et al., 2016). However, research exploring lexical access provides strong evidence that high- and low-frequency words differ in processing fluency. The present work aims to better understand how word frequency influences JOLs, so we now review what is currently known about word frequency measures and word frequency effects on memory and JOLs.

3. Word Frequency: An Important but Puzzling Cue for JOLs

Research in experimental psychology often uses words as stimuli, which requires the control of word properties. This control is only possible thanks to the development of lexical databases that provide "information about the structural (attributes) and distributional (statistics) characteristics of words in a given language" (Soares et al., 2018, p. 1461). There are reports of lexical databases dating back to 1898 (for a review see Bontrager, 1991). However, the first development of these sorts of databases, especially for pedagogical purposes, is often attributed to Thorndike (1921), who manually counted the number of times a particular word occurred in an English corpus of 4.5 million words, ranking the most frequent 10,000 English words. Later, Thorndike's work was combined with Lorge's work in the book *Teacher's Word Book of 30,000 Words* (Thorndike & Lorge, 1944), which is still a reference for word frequency in the English language. Since then, lexical databases have greatly improved, mostly because of technological advances that automatically count words from different linguistic sources (Bontrager, 1991; Soares et al., 2018).

Despite these advances, it is often challenging to find lexical databases that comprise all the measures a researcher wants or needs to control in their studies, so it is likely for them to just rely on databases that are easily available (Brysbaert et al., 2011). In fact, lexical databases can be rare or limited in some measures, especially regarding subjective measures (Soares et al., 2017). There are several databases for languages like English, Dutch, or Spanish, but for European Portuguese, these are relatively scarce (Soares et al., 2018). To overcome this issue, recent work from Soares and colleagues has proved invaluable in the development of more complete lexical databases, such as Procura-PALavras (P-PAL, Soares et al., 2014, 2018), SUBTLEX-PT (Soares et al., 2015), and Minho Word Pool (Soares et al., 2017). One of the most important variables presented in lexical databases is word frequency, which is known to be central in the processing of words.

Word frequency is defined as the number of times a word occurs in a language and has been identified as a key aspect in word identification (Howes, 1957; Howes & Solomon, 1951). There are different ways to measure word frequency. Objective word frequency is the actual number of occurrences of the word in a particular corpus and is usually measured in terms of its occurrences per million of words, or frequency per million (fpm), a logarithmic standardized value that allows comparisons independent of corpus size (e.g., Brysbaert et al., 2011; Brysbaert et al., 2018). According to Brysbaert et al. (2018), low-frequency words are usually defined as having less than 5 fpm, while high-frequency words have more than 100 fpm. Still, this measure presents some issues. First, its quality depends on the size of the lexical corpus, with smaller corpora (i.e., with less than 20 million words), underestimating the number of times words occur in the language (Brysbaert et al., 2018; Soares et al., 2018). Moreover, the effects of frequency seem to be compressed in the fpm scale, which can be problematic for interpretation and can lead to wrong intuitions (Brysbaert et al., 2018). More specifically, people intuitively associate a measure of 1 occurrence with the lowest value, but more than half of the words in a frequency list have frequencies lower than the standardized 1 fpm (van Heuven et al., 2014). Besides, the effects of frequency on lexical decision times when comparing 0.1 fpm and 1 fpm are similar to those comparing 1 fpm and 10 fpm, which could be solved by means of a logarithmic transformation (van Heuven et al., 2014).

It seems, however, that words follow a systematic frequency distribution in which there are few high-frequency words and many low-frequency words, obeying a law known as Zipf's law (Piantadosi, 2014). It was precisely based on this law, which organizes frequency in ranks, that van Heuven et al. (2014) proposed an alternate measure to fpm. Zipf is, then, a logarithmic scale based on the frequency per billion words. Within this scale, a Zipf below 3 defines a low-frequency word, while a Zipf over 3 defines a high-frequency word (Brysbaert et al., 2018; Soares et al., 2018). In a review of the word frequency distribution, Piantadosi (2014) alerts that derivations of Zipf's law "do not account for any psychological processes of word production" (p. 1127), proceeding to speculate about the reasons *why* word frequencies obey Zipf's law. Additional problems with word frequency measures include the assumption that every encounter with a word has the same weight, which may not be the case (Brysbaert et al., 2018; see also *word prevalence*, Brysbaert et al., 2016).

In fact, in a recently updated review, Brysbaert et al. (2018, update of Brysbaert et al., 2011) alerted for the importance of individual differences and argued that word frequency effects might derive from the diversity of situations in which a person encounters a word rather than the number of encounters with that word. This relates to subjective word frequency, which is a subjective estimate of how many times a person encounters a word in its written or spoken form in their daily life (Balota et al., 2001; Soares et al., 2017). Subjective frequency is highly correlated to objective frequency, and although it has been used as an equivalent measure (Tryk, 1968), it has also been reasoned to capture something different (Carroll, 1971). For example, it is argued that subjective frequency accounts for more variability in word frequency effects than objective frequency (Brysbaert et al., 2018). Hence, databases based on corpora closer to the daily usage of the language are arguably more related to subjective frequency than databases based on formal written texts (e.g., scientific papers, newspapers). In other words, values of frequency taken from subtitles of movies and television shows are argued to be representative of the daily usage of words, particularly of young adults, who are the most predominant participants in this type of psychological experiments. One example comes from two important lexical databases of European Portuguese: P-PAL, which is based on a printed corpus (Soares et al., 2018), and SUBTLEX-PT, which is based on a corpus of movies and television shows subtitles (Soares et al., 2015). Word frequency in SUBTLEX-PT explains around 15% more of the variance in young adults' lexical decision performance than the frequency in the P-PAL database, and diversity measures included in SUBTLEX-PT account for an extra 2% of the variance (Soares et al., 2015).

Despite these facts, there is evidence that if objective word frequency is well counted, there might be no problem using objective measures over subjective measures (Brysbaert & Cortese, 2011). Most researchers studying word frequency effects on metamemory simply report the exact number of occurrences per million, the standardized fpm, or a standardized rank (see Fiacconi & Dollois, 2020), a decision that sometimes is unclear and seems arbitrary. Also, it seems that different authors categorize what is high or low frequency differently. For example, Susser and Mulligan (2015) considered high-frequency words those ranging between 100 and 500 fpm and low-frequency words those ranging between 1 and 22 fpm; Jia et al. (2016) considered high-frequency words those ranging between 186 and 603 fpm and low-frequency words those ranging between 8 and 9 fpm; regarding the rank unit (i.e., rank 1 would be the most common word in the corpus, rank 2, the second most common word in the corpus, etc.), Benjamin (2003) and Tullis and Benjamin (2012) considered high-frequency words those in the range 100-270 and low-frequency words those in the range 5000-5230. The following sections of this chapter will focus on the research that has been conducted to understand the effects of word frequency on human memory and metamemory, which has relied mostly on objective measures of frequency. Because of that, in the present work objective word frequency measures were also used.[1]

3.1. Word Frequency Effects on Memory

Our memory for words depends, among other variables, on their frequency. However, the impact of word frequency on our memory diverges across memory tasks. Namely, memory performance is usually better for low-frequency words (i.e., rare words) in recognition tests (e.g., Clark, 1992; Gorman, 1961; Malmberg & Murnane, 2002), but it is usually better for high-frequency words (i.e., common words) in free recall tests (e.g., Hall, 1954; Jia et al., 2016; Sumby, 1963). Lohnas and Kahana (2013) described this effect as the word frequency paradox. These authors stated that most

---

1 In all studies presented in this dissertation, we reported P-PAL frequency values, but classification of words being of high or low frequency did not differ when relying on frequency values from SUBTLEX-PT (see Appendixes).

studies using pure lists (i.e., lists with only one kind of words, either high- or low-frequency) find a robust advantage for common words in recall tests, but the same is not true for mixed lists (i.e., lists with both high- and low-frequency words). In fact, they explored recall and recognition for mixed lists across a continuous range of word frequency and found a U-pattern for recall, that is, higher correct recall for both high- and low-frequency words in comparison to mid-frequency words; but for recognition, they observed the typical benefit for low-frequency words over high-frequency words (Lohnas & Kahana, 2013).

There are several explanations for the typical higher recall for high frequency words. High-frequency words may be more accessible and easier to generate, facilitating their retrieval in free recall (e.g., Madan et al., 2010). Alternatively, they may be easier to rehearse during study in comparison to low-frequency words, which would strengthen their memory traces (e.g., Tan & Ward, 2000, Exp. 2). Additionally, there is a higher likelihood for high-frequency words to have more semantic associations with other high-frequency words, which could help to make associations when studying lists of words (e.g., Ozubko & Joordens, 2007). It is also important to consider that the more we encounter a word in different contexts, the more familiar it becomes (e.g., Brysbaert et al., 2018). More recently, it has been proposed that processing high-frequency words consumes less working memory resources than low-frequency words, which would allow for the maintenance and subsequent storage of more high-frequency words than low-frequency words (for a review and a recent theory, see Popov & Reder, 2020).

Alternatively, low-frequency words are better recognized, with more hits (i.e., saying "old" for old items) and fewer false alarms (i.e., saying "old" for new items) than high-frequency words (e.g., Clark, 1992). This result posed a challenge for early theories that explained word frequency effects on memory through item strength. The source of activation confusion model (Reder et al., 2000) is a dual-process model stating that recognition is based on the recollection following the episodic activation of a node (context), and when an item cannot be recollected, it is based on a general familiarity involving the semantic activation of a node (concept; Reder et al., 2000). According to this model, high-frequency words have more robust conceptual representations and more contextual associations. In a recognition test, high-frequency words would activate more related nodes, which could account for lower hits; similarly, if there is no recollection of an item, and the decision is based on familiarity, it is more likely that high-frequency words lead to more false alarms (Reder et al. 2000). Similarly, the recent model of Popov and Reder (2020) also argues that, in a recognition test, memory for high-frequency words suffers from higher contextual competitors, which masks the encoding advantage of high-frequency words.

Additional support for high-frequency words to be more accessible and easily processed than low-frequency words comes from lexical decision or naming tasks (e.g., Balota & Chumbley, 1984; Brysbaert et al., 2018; Gao et al., 2016; Schilling et al., 1998): People are quicker to visually recognize and name common words over rare words when they are presented among pronounceable sequences of letters without meaning (i.e., pseudowords). As noted before, these tasks are commonly used to measure the processing fluency of different stimuli in metamemory research (e.g., Mueller et al., 2014).

3.2. Word Frequency Effects on JOLs

Word frequency effects on memory have been abundantly studied, but the same is not true for its effects on metamemory (Fiacconi & Dollois, 2020). A recent meta-analytic review investigated a total of 17 experiments regarding word frequency effects on people's JOLs and found a small but robust effect (Fiacconi & Dollois, 2020). Overall, people tend to judge high-frequency words as more memorable than low-frequency words (e.g., Benjamin, 2003; Jia et al., 2016; for exceptions, see Susser & Mulligan, 2015; and Susser et al., 2017). Importantly, this effect on JOLs seems to be, in general, independent of the memory test (for an exception, see Tullis & Benjamin, 2012).

In the first study conducted on word frequency effects in people's metacognitive judgments, Begg et al. (1989) found that high-frequency words were judged as more memorable, easier to understand, easier to pronounce, easier to imagine, and easier to study than low-frequency words. Hence, the effects of word frequency on JOLs were attributed to processing fluency. However, several years later, Benjamin (2003) reported that despite giving higher JOLs for high-frequency words before a recognition test, the same was not true for a second study-test cycle. In that study, young adults adjusted their judgments after the recognition test, and, in the second study-test cycle, JOLs were higher for low-frequency words. These findings suggest that participants updated their metacognitive knowledge, which could have been previously based on the belief that common words are better remembered and thus challenged the previous idea that frequency effects on JOLs were based exclusively on fluency. With experience, that belief might have been updated, with a noticeable impact on JOLs. In a replication of that study, Tullis and Benjamin (2012) found that this ability to update JOLs was also observed in older adults. However, contrary to Benjamin (2003), young adults predicted that they would better recognize low-frequency words on the first study-test cycle (Tullis & Benjamin, 2012), which further suggests that previous beliefs are the basis for the effects of word frequency on JOLs.

Strong evidence supporting the explanation of the effect in terms of beliefs, but not fluency, was found by Jia et al. (2016). In their study, fluency was measured through study time (Experiment 2a) and by using easy- vs. difficult-to-read font-styles (Experiment 2b). They found no differences in processing fluency between high- and low-frequency words, but JOLs and correct recall were still higher for high-frequency words. Additionally, in Experiments 3a and 3b, Jia et al. (2016) evaluated the role of beliefs in conditions where participants were not exposed to the words while making JOLs (i.e., no direct experience with the items). In Experiment 3a, participants were asked to judge how many high- and low-frequency words hypothetical participants from another experiment would have remembered. In Experiment 3b, the authors collected pre-study JOLs. In both experiments, results showed that JOLs were higher for high-frequency words, and the authors proposed a belief-only explanation for word frequency effects on JOLs. However, there are good theoretical reasons to challenge that belief-only explanation of the word-frequency effect on JOLs, which are presented in the next chapter.

4. Arguments Against a Belief-only Explanation of the Word Frequency Effect on JOLs

There has been a shift of evidence turning from fluency to beliefs in terms of explanations of effects on JOLs of different variables, such as font size or auditory volume. Similarly, from Begg et al. (1989) to Jia et al. (2006), this shift was also observed for word frequency. However, most recent research suggests that both experience-based and theory-based factors are the basis of JOLs regarding font size (e.g., Undorf et al., 2017) and auditory volume (e.g., Frank & Kuhlmann, 2017), being yet unclear how much one factor contributes more than the other. It is possible that the word frequency effect on JOLs could also be explained by a combination of both types of processes rather than only theory-driven processes. Here we present arguments challenging an explanation of the word frequency effect on JOLs based exclusively on beliefs, which motivate the present work.

First, because word frequency effects on memory are task-dependent, it may be difficult for people to rely on a consistent belief about the actual impact of frequency on memory to make their JOLs. This is especially true if they do not know the research on word frequency and human memory or do not know what kind of test they will experience. Contrasting results from Tullis and Benjamin (2012) and Benjamin (2003) with young adults suggest the existence of two opposing beliefs in the face of a recognition test: that memory is better either for low-frequency words or for high-frequency words, respectively. Nevertheless, Fiacconi and Dollois' (2020) recent meta-analysis showed that more studies report higher JOLs to high-frequency words, indicating that it is more likely for people to have a general belief that high-frequency words are more memorable.

Second, the contribution of experience-based factors, in particular processing fluency, in the effect of word frequency on JOLs may not have been measured most appropriately. Indeed, there is one critical issue with findings from Jia et al. (2016), who argued in favor of a belief-only explanation of the effects of frequency on JOLs. These authors' manipulations of fluency (difficulty of font style) probably altered perceptual fluency, but not lexical fluency, which may be more related to a linguistic frequency manipulation. Remember that different types of fluency are assessed differently (Alter & Oppenheimer, 2009): Perceptual fluency is related to how visually clear one item is and can be manipulated through font style (Reber & Zupanek, 2002), and lexical fluency relates to the utilization and familiarity of words and can be manipulated through word complexity (Oppenheimer, 2006).[2] This means that each fluency manipulation may require a correspondent measuring method. In other words, different types of fluency require adequate tasks that are sensitive to differences in that specific type of fluency. For example, font

2 In Oppenheimer's (2006) study, texts with more obscure and less familiar words were harder to process.

size is an extrinsic cue[3] that differs perceptually (words presented in 48 pt are perceptually different from words presented in 18 pt) and manipulating font-style may disrupt perceptual processing (e.g., Rhodes & Castel, 2008). While differences in processing fluency between words presented in large and small font sizes have been detected by means of a lexical decision task with extreme manipulations (e.g., Luna et al., 2019; Undorf & Zimdahl, 2019), there is evidence suggesting that using a lexical decision task to detect differences in perceptual fluency may be inadequate (see Yang et al., 2018).

In contrast, frequency is an intrinsic cue that is inextricable from each word, and high- and low-frequency words differ in semantic access and familiarity. The consistent finding that high-frequency words are more easily recognized in lexical decision tasks (Brysbaert et al., 2018) strongly suggests that high- and low-frequency words differ in fluency, most likely lexical fluency and not perceptual fluency. However, Jia et al. (2016, Exp. 2b) manipulated perceptual fluency through font-style (and not lexical fluency) and failed to find an interaction between word frequency and font-style. That result led to the conclusion that fluency does not mediate word frequency effects on JOLs. Additionally, semantic access may be unaffected by the manipulation of the font-style, which the authors acknowledge. Lexical decision and naming tasks may be more sensitive to evaluate differences in lexical fluency, as well as other components of linguistic fluency or even conceptual fluency. Hence, high- and low-frequency words may differ not in perceptual fluency but in lexical fluency, which would be more fittingly measured through lexical decision tasks.

Third, if we assume that a belief-only explanation of the word frequency on JOLs is correct, it is likely that JOLs are made considering the belief of how word frequency affects memory. For that to be true, participants must activate their beliefs somehow. The belief could be that common words are better remembered for being more familiar, for example, or that rare words are better remembered for being more distinctive, which could justify giving higher JOLs to either high-frequency words or low-frequency words, respectively. Also, one could think that memory is independent of word frequency and give similar JOLs to common and rare words. How could someone activate their beliefs? One possibility would be informing them that the to-be-studied words will differ in frequency. This information could activate a belief before studying the words, as with pre-study JOLs. Alternatively, participants could detect during the study that words differ in frequency and then incorporate their belief into their JOLs. This would be rather easy when referring to a cue manipulated perceptively, such as font size (i.e., it is

3 Not all authors consider font size as an extrinsic cue, but we follow the rationale presented by Undorf et al., (2018).

clear that some words are presented in large font size and others are presented in small font size). In contrast, an intrinsic cue such as word frequency could be more difficult to identify in mixed lists.

Evidence supporting this argument comes from a recent study in which participants were asked to identify the categories manipulated across different lists (Barbosa & Albuquerque, 2020). In one experiment, participants studied several lists comprising 20 words from different perceptive (e.g., presented in 60 pt vs. 22 pt; presented in blue vs. green) and semantic (e.g., high- vs. low-frequency; concrete vs. abstract) categories. Their task was to name the categories manipulated for each list as soon as they had identified them. Results showed a high identifiability rate for perceptive lists, with a mean accuracy of 89%, and all categories were identified correctly by at least 83% of participants (font size was correctly identified by 92% of participants). However, identifiability of semantic categories varied between 22% (for word frequency), and 97% (Portuguese vs. English words), with a significantly lower mean identifiability rate of 57%. Additionally, on average, participants needed to study significantly fewer words to identify perceptive categories (approx. 6 words) than they needed to identify semantic categories (approx. 10 words). These results suggest that variations in font size are more easily perceived than variations in word frequency. Despite the low identifiability of word frequency, in a second experiment participants still gave higher JOLs for high-frequency words than for low-frequency words (Barbosa & Albuquerque, 2020). Nonetheless, most studies use words that fall on the extremes of frequency, which could make word frequency a cue easier to identify as relevant for memorability when making JOLs than when it is manipulated through non-extreme intervals of frequency. If participants can identify word frequency, it is reasonable to assume that they can develop beliefs about how it affects memory and use those beliefs to guide their judgments. However, this does not seem to be the case because words used by Barbosa and Albuquerque (2020) fell upon the extremes of word frequency (M_{LF} = 0.84 fpm, SD_{LF} = 0.79; M_{HF} = 381.58 fpm, SD_{HF} = 175.57), and still the category was not easily identifiable. This may challenge the use of beliefs to guide JOLs when studying high- and low-frequency words.

Finally, in real life, people make metacognitive judgments when multiple cues are available, so recently there has been a growing interest in understanding how JOLs are made in such circumstances (e.g., Tatz & Peynircioğlu, 2020; Undorf et al., 2018). When several cues are available, it might be harder to identify all cues that could affect memory. Thus, activating beliefs that could guide JOLs may become more difficult. In fact, there is evidence that word frequency may not affect JOLs when other relevant cues are available. For example, Susser and Mulligan (2015, Exp. 2) explored the effects of motoric fluency on JOLs by asking participants to write high- and low-frequency words with their dominant (vs. non-dominant) hand. While they found effects of motoric fluency (measured through writing time), there was no effect of frequency on JOLs. These results were replicated later (Susser et al., 2017). The absence of the word frequency effect on JOLs in those studies may occur because motoric information was more salient than frequency, which limited participants' reliance on beliefs about frequency. However, Hourihan et al. (2017, Experiment 3) manipulated word frequency simultaneously with valence and arousal, finding that only word frequency affected JOLs despite the other variables affecting JOLs in previous experiments. Even when basing JOLs on beliefs of how frequency impacts memory might have been difficult due to the presence of other relevant cues, word frequency still affected JOLs. While there is evidence for cue integration in JOLs, with participants relying on up to four different cues (e.g., Undorf et al., 2018), those contrasting findings regarding word frequency leave the question of whether word frequency affects JOLs in combination with other cues unsolved. One can speculate that if word frequency and an additional cue affect JOLs together through beliefs, either people have the sensitivity to detect both cues to activate their beliefs, or beliefs alone may not be enough to explain the effect on JOLs.

5. The Present Work

This dissertation aimed to better understand the mechanisms behind the word frequency effect on JOLs. The main theoretical rationale adopted was the cue-utilization approach (Koriat, 1997), based on an inferential process in which word frequency is used by participants as a cue to make their JOLs. More specifically, this thesis aimed to explore if, in addition to theory-driven processes, experience-driven processes also serve as the basis for the effects of word frequency on JOLs. To that end, a total of 12 experiments were conducted and were organized into four different studies: An unpublished pilot study with one experiment, and three published or submitted studies comprising four (Study 1), four (Study 2), and three experiments (Study 3), respectively. These are summarized below. For specific objectives, hypotheses, methodology, and results, please see the relevant study (Chapters 7 to 9). The work for this dissertation received the ethical approval of the Ethics' Subcommittee for Human and Social Sciences of theUniversity of Minho (see Appendix E).

The pilot study aimed to test if people were able to identify word frequency as a cue while studying high- and low-frequency words. If participants can identify word frequency as a cue, they can develop beliefs about its influence on memory. However, if they cannot, they must rely on other factors, either fluency that arises from direct experience with the items, beliefs about fluency, or other cues considered relevant by participants. Across four study-test cycles, participants studied mixed lists of words, each with different frequency intervals, and made immediate JOLs. Then, they completed a free recall test. Overall, results showed no effect of frequency on JOLs or recall. This was especially troubling in the list that used extreme values of word frequency (M_{LF} = 7.57 fpm, M_{HF} = 112.62 fpm). Post-hoc inspection of the specific words selected lead to the conclusion that high- and low-frequency words needed to be more extreme for the effects of frequency to occur, which lead to the following studies.

Study 1 followed the same rationale of the pilot study. The aim was to test if people could identify word frequency as a cue when making JOLs for high- and low-frequency words. In this study, we used more extreme values of frequency. In Experiments 1 and 2, participants made immediate JOLs for high- and low-frequency words and answered some questions to explore cue identification and cue use on JOLs, followed by a recall test. Besides, in Experiment 3, fluency for high- and low-frequency words was measured by response times in a lexical decision task. Finally, Experiment 4 explored if response times in a lexical decision task (i.e., fluency) mediated the effect of word frequency on JOLs. In general, the results showed higher JOLs for high-frequency words than for low-frequency words, and this pattern was observed even for participants who could not identify word frequency as a cue and reported basing their JOLs on other cues. Additionally, response times in the lexical decision task were lower for high-frequency words, suggesting these to be more fluently processed than low-frequency words. Response times in the lexical decision task mediated effects of frequency on JOLs, strongly suggesting that experience-based processes also guide the effect.

Study 2 aimed to explore which factor contributes the most to word frequency effects on immediate JOLs: Direct experience with the items or beliefs. To address this question, a paradigm that requested a combination of pre-study and immediate JOLs was used. Critically, false/misleading information about word frequency was introduced in prompts presented before the actual to-be-studied words (e.g., “You will see a high/low-frequency word.”). Half of the prompts were incongruent with the actual word frequency of the presented words. The main goal was to detect changes from pre-study to immediate JOLs when the prompt presented (e.g., "You will see a low-frequency word") did not match the actual frequency of the to-be-studied word (e.g., the high-frequency word "time"). When making pre-study JOLs, participants can only rely on beliefs, but immediate JOLs may incorporate both fluency and beliefs. If experience with the items drives the effect, immediate JOLs should be based on actual word frequency and not on the information given by the prompt. Results showed that when prompts were presented before studying the word (Experiment 1a), pre-study JOLs were based on the information given by the prompt (higher for prompts of high frequency), but immediate JOLs were based on actual word frequency (higher for high-frequency words). This latter finding was also observed when participants only made immediate JOLs, but prompts were presented after studying the word (Experiment 1b). Additionally, and with a similar procedure (i.e., asking both pre-study and immediate JOLs), participants' beliefs were manipulated through instructions. Instructions stating that memory is generally better for low-frequency words (i.e., a counter-belief) was presented for all participants (Experiment 2), or for half of the participants whereas the remaining half read that high-frequency words are better remembered (Experiment 3). When a counter-belief was introduced, previously observed effects of frequency on pre-study JOLs disappeared, but immediate JOLs were still higher for high-frequency words; with a same-direction belief, pre-study JOLs were higher for prompts of high frequency, and immediate JOLs were higher for high-frequency words regardless of previous prompt. Together, these results provide evidence that experience-driven effects seem to be the main basis of word frequency effects on JOLs, with a smaller role of beliefs.

Finally, Study 3 aimed to explore if word frequency would affect JOLs when an additional cue (i.e., font size) was simultaneously manipulated. It is possible that people are unable to integrate word frequency and another cue in their JOLs, so maybe only one cue would affect JOLs in such situations.

More specifically, it could be more challenging for people to rely on theory-driven processes to make JOLs while attending to more than one cue. A secondary aim of this study was to explore if JOLs are sensitive to subjective frequency. Experiment 1 tested if both word frequency and font size would affect immediate JOLs. We expected higher JOLs for high-frequency words and higher JOLs for larger font sizes because of the robustness of both effects and based on previous evidence for combined effects of more than one cue on JOLs. Experiment 2 aimed to extend findings from Experiment 1 to a scenario in which word frequency was manipulated in a continuum, which might make it less salient to be used as a cue. If word frequency effects on JOLs are robust, JOLs should increase with word frequency. In both experiments, JOLs increased with font size and word frequency, with the majority of participants relying on both cues. In those scenarios relying on theory-driven processes may have been more difficult, suggesting that participants' JOLs were also based on experience. Experiment 3 explored the role of subjective frequency. Besides standard high- and low-frequency words, this experiment included two additional pools of words: related to math vs. related to psychology. Psychology students and students from math-based curricula (e.g., Business Economics) made immediate JOLs for high-frequency words, low-frequency words, psychology-related words, and math-related words. It was assumed that subjective frequency between both sets of students was the same for standard high- and low-frequency words, but that it differed for the curricula-related words. If the subjective frequency has an impact on JOLs, we anticipated a pattern of results similar to those observed with objective frequency, with more familiar words receiving higher JOLs. More specifically, psychology students were expected to give higher JOLs to psychology-related words than math-related words, and vice versa for math students. Results from Experiment 3 showed the expected word frequency effects on JOLs (i.e., higher for standard high-frequency than standard low-frequency), but also higher JOLs for math words than standard low-frequency words among math students, and higher JOLs for psychology words than standard low-frequency words. These findings suggest that both objective and subjective frequency affect JOLs. Through all experiments of Study 3, we used the aforementioned Brunswik's (1952) lens model to evaluate JOL accuracy. Results provided support for the lens model measure of accuracy as a valid resolution measure. Findings from these three experiments further suggest that the word frequency affects JOLs in a scenario where it may be harder to rely on beliefs. In other words, when a more salient cue such as font size is available, people may rely on experience by giving higher JOLs to high-frequency words.

Following this brief description of all the experiments conducted for this thesis, the next part will present each study individually. The published studies will be presented as they were submitted to the scientific journals in which they were published, so each will have its own theoretical introduction and discussion. For each study, we provide details about the scientific quality of the journals (i.e., JCR and Scopus impact factors and their quartile). In addition, we also indicate the communications in scientific events that included data reported in each study. Finally, a global discussion on the overall contribution of this dissertation will be presented in the third part.

References

Alter, A. L., & Oppenheimer, D. M. (2006). Predicting short-term stock fluctuations by using processing fluency. *Proceedings of the National Academy of Sciences, 103*(24), 9369–9372. https://doi.org/10.1073/pnas.0601071103

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

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

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. https://doi.org/10.1037/0096-1523.10.3.340

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29*(4), 639–647. https://doi.org/10.3758/BF03200465

Barbosa, B., & Albuquerque, P. B. (2020). *Influência da identificação das categorias de listas de palavras na memória e metamemória: Estudo com evocação livre e julgamentos de aprendizagem* [Master's thesis, Universidade do Minho]. RepositoriUM. http://hdl.handle.net/1822/67518

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. https://doi.org/10.1016/0749-596X(89)90016-8

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31*(2), 297–305. https://doi.org/10.3758/BF03194388

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. https://doi.org/10.1037/xlm0000246

Blake, A. B., & Castel, A. D. (2018). On belief and fluency in the construction of judgments of learning: Assessing and altering the direct effects of belief. *Acta Psychologica, 186*, 27–38. https://doi.org/10.1016/j.actpsy.2018.04.004

Bontrager, T. (1991). The development of word frequency lists prior to the 1944 Thorndike-Lorge list. *Reading Psychology, 12*(2), 91–116. https://doi.org/10.1080/0270271910120201

Bröder, A., & Undorf, M. (2019). Metamemory viewed through the judgment lens. *Acta Psychologica, 197*, 153–165. https://doi.org/10.1016/j.actpsy.2019.04.011

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109*(2), 204–223. https://doi.org/10.1037/0033-2909.109.2.204

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*(4), 325–337. https://doi.org/10.1016/S0022-5371(66)80040-3

Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago University Press.

Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology, 64*(3), 545–559. https://doi.org/10.1080/17470218.2010.503374

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology, 58*(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance, 42*(3), 441–458. https://doi.org/10.1037/xhp0000159

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30*(5), 542–579. https://doi.org/10.1016/0749-596X(91)90026-G

Canfield, C., Fischhoff, B., & Davis, A. (2019). Better beware: comparing metacognition for phishing and legitimate emails. *Metacognition and Learning, 14*(3), 343–362. https://doi.org/10.1007/s11409-019-09197-5

Carroll, J. B. (1971). Measurement properties of subjective magnitude estimates of word frequency. *Journal of Verbal Learning and Verbal Behavior, 10*(6), 722–729. https://doi.org/10.1016/S0022-5371(71)80081-6

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition, 36*(2), 429–437. https://doi.org/10.3758/MC.36.2.429

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition, 20*(3), 231–243. https://doi.org/10.3758/BF03199660

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior, 4*(4), 243–260. https://doi.org/10.1007/BF01040617

Dentakos, S., Saoud, W., Ackerman, R., & Toplak, M. E. (2019). Does domain matter? Monitoring accuracy across domains. *Metacognition and Learning, 14*(3), 413–436. https://doi.org/10.1007/s11409-019-09198-4

Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory & Cognition, 33*(6), 1096–1115. https://doi.org/10.3758/BF03193216

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational Psychology. *Psychological Science in the Public Interest, 14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Dunlosky, J., & Tauber, S. K. (2016). A brief history of metamemory research and handbook overview. In J. Dunlosky & S. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 7–21). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.13

Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction, 24*, 58–61. https://doi.org/10.1016/j.learninstruc.2012.05.002

Dunlosky, J., Mueller, M. L., & Tauber, S. K. (2015). The contribution of processing fluency (and beliefs) to people's judgments of learning. In D. S. Lindsay (Ed.) & C. M. Kelley (Trans.) & A. P. Yonelinas, H. L. Roediger III (Eds.), *Psychology Press festschrift series. Remembering:*

Attributions, processes, and control in human memory: Essays in honor of Larry Jacoby (pp. 46–64). Psychology Press. https://psycnet.apa.org/record/2014-56350-004

Fiacconi, C. M., & Dollois, M. A. (2020). Does word frequency influence judgments of learning (JOLs)? A meta-analytic review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale.* Advance online publication. https://doi.org/10.1037/cep0000206

Flavell, J. H. (1971). First discussant's comments: What is memory development the development of? *Human Development, 14*(4), 272–278. https://doi.org/10.1159/000271221

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Flavell, J. H., & Wellman, H. M. (1975, August). *Metamemory* [Paper presentation]. 83rd Annual Meeting of the American Psychological Association, Chicago, Illinois. https://files.eric.ed.gov/fulltext/ED115405.pdf

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(5), 680–693. https://doi.org/10.1037/xlm0000332

Gao, X.Y., Li, M. F., Chou, T.L, & Wu, J. T. (2016). Comparing the frequency effect between the lexical decision and naming tasks in Chinese. *Journal of Visualized Experiments, 110*, e53815, https://doi.org/10.3791/53815

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

Ghetti, S., & Angelini, L. (2008). The development of recollection and familiarity in childhood and adolescence: Evidence from the dual-process signal detection model. *Child Development, 79*(2), 339–358. https://doi.org/10.1111/j.1467-8624.2007.01129.x

Ghetti, S., Lyons, K. E., Lazzarin, F., & Cornoldi, C. (2008). The development of metamemory monitoring during retrieval: The case of memory strength and memory absence. *Journal of Experimental Child Psychology, 99*(3), 157–181. https://doi.org/10.1016/j.jecp.2007.11.001

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association, 49*(268), 732–764. https://doi.org/10.2307/2281536

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology, 61*(1), 23–29. https://doi.org/10.1037/h0040561

Graf, L. K., Mayer, S., & Landwehr, J. R. (2018). Measuring processing fluency: One versus five items. *Journal of Consumer Psychology, 28(3)*, 393–411. http://doi.org/10.1002/jcpy.1021

Hall, J. F. (1954). Learning as a function of word-frequency. *The American Journal of Psychology, 67*(1), 138–140. https://doi.org/10.2307/1418080

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*(4), 208–216. https://doi.org/10.1037/h0022263

Hicks, J. L., & Marsh, R. L. (1999). Remember-know judgments can depend on how memory is tested. *Psychonomic Bulletin & Review, 6*(1), 117–122. https://doi.org/10.3758/BF03210818

Higham, P. A., Zawadzka, K., & Hanczakowski, M. (2016). Internal mapping and its impact on measures of absolute and relative metacognitive accuracy. In J. Dunlosky, & S. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 39–61). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.15

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2017). The influences of valence and arousal on judgments of learning and on recall. *Memory & Cognition, 45*(1), 121–136. https://doi.org/10.3758/S13421-016-0646-3

Howes, D. (1957). On the relation between the probability of a word as an association and in general linguistic usage. *The Journal of Abnormal and Social Psychology, 54*(1), 75–85. https://doi.org/10.1037/h0043830

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology, 41*(6), 401–410. https://doi.org/10.1037/h0056020

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). https://doi.org/10.1371/journal.pone.0142351

Jemstedt, A., Schwartz, B. L., & Jönsson, F. U. (2018). Ease-of-learning judgments are based on both processing fluency and beliefs. *Memory, 26*(6), 807–815. https://doi.org/10.1080/09658211.2017.1410849

Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2016). The effect of word frequency on judgments of learning: Contributions of beliefs and processing fluency. *Frontiers in Psychology, 6:*1995. https://doi.org/10.3389/fpsyg.2015.01995

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

Jönsson, F. U., & Lindström, B. R. (2010). Using a multidimensional scaling approach to investigate the underlying basis of ease of learning judgments. *Scandinavian Journal of Psychology, 51*(2), 103–108. https://doi.org/10.1111/j.1467-9450.2009.00762.x

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134*(3), 404–426. https://doi.org/10.1037/0033-2909.134.3.404

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. https://doi.org/10.1006/jmla.1996.0009

Kleitman, S., & Narciss, S. (2019). Introduction to the special issue "Applied metacognition: Real-world applications beyond learning". *Metacognition and Learning, 14*(3), 335–342. https://doi.org/10.1007/s11409-019-09214-7

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*(4), 609–639. https://doi.org/10.1037/0033-295X.100.4.609

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*(3), 311–333. https://doi.org/10.1037/0096-3445.124.3.311

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. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 34–53. https://doi.org/10.1037/0278-7393.27.1.34

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 Psychology: General, 133*(4), 643–656. https://doi.org/10.1037/0096-3445.133.4.643

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*(2), 147–162. https://doi.org/10.1037/0096-3445.131.2.147

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 609–622. https://doi.org/10.1037/0278-7393.32.3.609

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. https://doi.org/10.1177/0956797611407929

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Kuhlmann, B. G., Bayen, U. J., Meuser, K., & Kornadt, A. E. (2016). The impact of age stereotypes on source monitoring in younger and older adults. *Psychology and Aging, 31*(8), 875–889. https://doi.org/10.1037/pag0000140

Kurilla, B. P., & Westerman, D. L. (2008). Processing fluency affects subjective claims of recollection. *Memory & Cognition, 36*(1), 82–92. https://doi.org/10.3758/MC.36.1.82

Leippe, M. R., & Eisenstadt, D. (2007). *Eyewitness confidence and the confidence-accuracy relationship in memory for people.* In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. 2. Memory for people* (pp. 377–425). Lawrence Erlbaum Associates Publishers. https://psycnet.apa.org/record/2007-04817-016

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

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). New York: Cambridge University Press.

Lindsay, D. S. (2008). Source monitoring. In H.L. Roediger, III (Ed.), *Learning and Memory: A comprehensive reference* (Vol. 2, pp. 325–347). Oxford: Elsevier. https://doi.org/10.1016/B978-012370509-9.00175-3

Lohnas, L. J., & Kahana, M. J. (2013). Parametric effects of word frequency in memory for mixed frequency lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1943–1946. https://doi.org/10.1037/a0033669

Lowrey, T. M. (1998). The effects of syntactic complexity on advertising persuasiveness. *Journal of Consumer Psychology, 7*(2), 187–206. https://doi.org/10.1207/s15327663jcp0702_04

Luna, K. (2019). If it is easy to remember, then it is not secure: Metacognitive beliefs affect password selection. *Applied Cognitive Psychology, 33*(5), 744–758. https://doi.org/10.1002/acp.3516

Luna, K., & Martín-Luengo, B. (2010). New advances in the study of the confidence-accuracy relationship in the memory for events. *The European Journal of Psychology Applied to a Legal Context*, 55–71.

Luna, K., & Martín-Luengo, B. (2012). Confidence-accuracy calibration with general knowledge and eyewitness memory cued recall questions. *Applied Cognitive Psychology, 26*(2), 289–295. https://doi.org/10.1002/acp.1822

Luna, K., Martín-Luengo, B., & Albuquerque, P. B. (2018). Do delayed judgements of learning reduce metamemory illusions? A meta-analysis. *Quarterly Journal of Experimental Psychology, 71*(7), 1626–1636. https://doi.org/10.1080/17470218.2017.1343362

Luna, K., Nogueira, M., & Albuquerque, P. B. (2019). Words in larger font are perceived as more important: Explaining the belief that font size affects memory. *Memory, 27*(4), 555–560. https://doi.org/10.1080/09658211.2018.1529797

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates Publishers.

Madan, C. R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language, 63*(1), 46–63. https://doi.org/10.1016/j.jml.2010.03.001

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 663–679. https://doi.org/10.1037/0278-7393.10.4.663

Malmberg, K. J., & Murnane, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 616–630. https://doi.org/10.1037/0278-7393.28.4.616

Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 509–527. https://doi.org/10.1037/a0014876

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

McGuire, M. J., & Maki, R. H. (2001). When knowing more means less: The effects of fan on metamemory judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1172–1179. https://doi.org/10.1037/0278-7393.27.5.1172

Metcalfe, J. & Dunlosky, J. (2008). Metamemory. In H.L. Roediger, III (Ed.), *Learning and Memory: A comprehensive reference* (Vol 2., pp. 349-362). Academic Press. https://doi.org/10.1016/B978-012370509-9.00159-5

Metcalfe, J. (2000). Feelings and judgments of knowing: Is there a special noetic state? *Consciousness and Cognition, 9*(2), 178–186. https://doi.org/10.1006/ccog.2000.0451

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(4), 851–861. https://doi.org/10.1037/0278-7393.19.4.851

Mickes, L., Seale-Carlisle, T. M., & Wixted, J. T. (2013). Rethinking familiarity: Remember/Know judgments in free recall. *Journal of Memory and Language, 68*(4), 333–349. https://doi.org/10.1016/j.jml.2013.01.001

Migueles, M., & García-Bajos, E. (1999). Recall, recognition, and confidence patterns in eyewitness testimony. *Applied Cognitive Psychology, 13*(3), 257–268. https://doi.org/10.1002/(SICI)1099-0720(199906)13:3<257::AID-ACP566>3.0.CO;2-7

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? *Quarterly Journal of Experimental Psychology, 69*(4), 781–799. https://doi.org/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. https://doi.org/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. https://doi.org/10.3758/s13423-012-0343-6

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. https://doi.org/10.1016/j.jml.2016.10.008

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95(*1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect.". *Psychological Science, 2*(4), 267–271. https://doi.org/10.1111/j.1467-9280.1991.tb00147.x

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect.". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 676–686. https://doi.org/10.1037/0278-7393.14.4.676

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.). *The Psychology of learning and motivation* (Vol. 26, pp. 125–141). Academic Press.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). The MIT Press.

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology, 20*(2), 139–156. https://doi.org/10.1002/acp.1178

Ozubko, J. D., & Joordens, S. (2007). The mixed truth about frequency effects on free recall: Effects of study list composition. *Psychonomic Bulletin & Review, 14*(5), 871–876. https://doi.org/10.3758/BF03194114

Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology Review, 15*(3), 105–130. https://doi.org/10.1007/s11065-005-7091-6

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review, 21*(5), 1112–1130. https://doi.org/10.3758/s13423-014-0585-6

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review, 127*(1), 1–46. http://doi.org/10.1037/rev0000161

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. https://doi.org/10.1016/j.jml.2016.12.003

Reber, R., & Zupanek, N. (2002). Effects of processing fluency on estimates of probability and frequency. In P. Sedlmeier & T. Betsch (Eds.), *Frequency processing and cognition* (pp. 175-188). Oxford University Press.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19*(1), 90–138. https://doi.org/10.1016/0010-0285(87)90005-3

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(2), 294–320. https://doi.org/10.1037/0278-7393.26.2.294

Reisberg, D., & McLean, J. (1985). Meta-attention: Do we know when we are being distracted? *Journal of General Psychology, 112*(3), 291–306. https://doi.org/10.1080/00221309.1985.9711015

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky, & S. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 90–117). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.4

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. https://doi.org/10.1037/a0013684

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*(3), 550–554. https://doi.org/10.3758/PBR.16.3.550

Robey, A. M., Dougherty, M. R., & Buttaccio, D. R. (2017). Making retrospective confidence judgments improves learners' ability to decide what not to study. *Psychological Science, 28*(11), 1683–1693. https://doi.org/10.1177/0956797617718800

Roebers, C. M. (2002). Confidence judgments in children's and adult's event recall and suggestibility. *Developmental Psychology, 38*(6), 1052–1067. https://doi.org/10.1037/0012-1649.38.6.1052

Roebers, C. M., & Howie, P. (2003). Confidence judgments in event recall: Developmental progression in the impact of question format. *Journal of Experimental Child Psychology, 85*(4), 352–371. https://doi.org/10.1016/S0022-0965(03)00076-6

Roediger, H. L. III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

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. http://doi.org/10.3758/BF03201199

Schreiber, T. A. (1998). Effects of target set size on feelings of knowing and cued recall: Implications for the cue effectiveness and partial-retrieval hypotheses. *Memory & Cognition, 26*(3), 553–571. https://doi.org/10.3758/BF03201162

Schulman, A. I. (1974). Memory for words recently classified. *Memory & Cognition, 2(1)*, 47–52. https://doi.org/10.3758/BF03197491

Schwartz, B. L. (2006). Tip-of-the-tongue states as metacognition. *Metacognition and Learning, 1*(2), 149–158. https://doi.org/10.1007/s11409-006-9583-z

Schwartz, B. L., Benjamin, A. S., & Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science, 6*(5), 132–137. https://doi.org/10.1111/1467-8721.ep10772899

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review, 1*(3), 357–375. https://doi.org/10.3758/BF03213977

Soares, A. P., Iriarte, Á., Almeida, J. J., Simões, A., Costa, A., França, P., Machado, J., & Comesaña, M. (2014). Procura-PALavras (P-PAL): Uma nova medida de frequência lexical do Português Europeu contemporâneo. *Psicologia: Reflexão e Crítica, 27*(1), 110–123. https://doi.org/10.1590/s0102-79722014000100013

Soares, A. P., Iriarte, Á., Almeida, J. J., Simões, A., Costa, A., Machado, J., França, P., Comesaña, M., Rauber, A., Rato, A., & Perea, M. (2018). Procura-PALavras (P-PAL): A web-based interface for a new European Portuguese lexical database. *Behavior Research Methods, 50*(4), 1461–1481. https://doi.org/10.3758/s13428-018-1058-z

Soares, A. P., Machado, J., Costa, A., Iriarte, Á., Simões, A., Almeida, J. J., Comesaña, M., & Perea, M. (2015). On the advantages of word frequency and contextual diversity measures extracted from subtitles: The case of Portuguese. *Quarterly Journal of Experimental Psychology, 68*(4), 680–696. https://doi.org/10.1080/17470218.2014.964271

Soares, A. P., Costa, A. S., Machado, J., Comesaña, M., & Oliveira, H. M. (2017). The Minho Word Pool: Norms for imageability, concreteness, and subjective frequency for 3,800 Portuguese words. *Behavior Research Methods, 49*(3), 1065–1081. https://doi.org/10.3758/s13428-016-0767-4

Soderstrom, N. C., Yue, C. L., & Bjork, E. L. (2016). Metamemory and education. In J. Dunlosky, & S. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 197–215). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.6

Sumby, W. H. (1963). Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior, 1*(6), 443–450. https://doi.org/10.1016/S0022-5371(63)80030-4

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review, 22*(4), 1014–1019. https://doi.org/10.3758/s13423-014-0768-1

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. https://doi.org/10.1037/xlm0000189

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. https://doi.org/10.1016/j.jml.2017.03.002

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1589–1625. https://doi.org/10.1037/0278-7393.26.6.1589

Tatz, J. R., & Peynircioğlu, Z. F. (2020). Judgments of learning in context: Backgrounds can both reduce and produce metamemory illusions. *Memory & Cognition, 48*(4), 581–595. https://doi.org/10.3758/s13421-019-00991-9

Thiede, K. W. (1996). The relative importance of anticipated test format and anticipated test difficulty on performance. *The Quarterly Journal of Experimental Psychology, 49*(4), 901–918. https://doi.org/10.1080/713755673

Thorndike, E. L. (1921). *The teacher's word book*. Teachers College, Columbia University.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. Bureau of Publications, Teachers Co.

Tryk, H. E. (1968). Subjective scaling of word frequency. *The American Journal of Psychology, 81*(2), 170–177. https://doi.org/10.2307/1421261

Tullis, J. G., & Benjamin, A. S. (2012). The effectiveness of updating metacognitive knowledge in the elderly: Evidence from metamnemonic judgments of word frequency. *Psychology and Aging, 27*(3), 683–690. https://doi.org/10.1037/a0025838

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

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

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition, 43*(4), 647–658. https://doi.org/10.3758/s13421-014-0479-x

Undorf, M., & Ackerman, R. (2017). The puzzle of study time allocation for the most challenging items. *Psychonomic Bulletin & Review, 24*(6), 2003-2011. https://doi.org/10.3758/s13423-017-1261-4

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. https://doi.org/10.1037/xlm0000571

Undorf, M., & Bröder, A. (2020). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology, 73*(4), 629–642. https://doi.org/10.1177/1747021819882308

Undorf, M., Söllner, A. & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory & Cognition, 46*(4), 507–519. https://doi.org/10.3758/s13421-017-0780-6

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. https://doi.org/10.1016/j.jml.2016.07.003

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Vogel, T., Silva, R. R., Thomas, A., & Wänke, M. (2020). Truth is in the mind, but beauty is in the eye: Fluency effects are moderated by a match between fluency source and judgment dimension. *Journal of Experimental Psychology: General, 149*(8), 1587–1596. https://doi.org/10.1037/xge0000731

Wells, A. (2011). *Metacognitive therapy for anxiety and depression*. The Guilford Press

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. https://doi.org/10.3758/s13421-016-0681-0

Wixted, J. T. (2009). Remember/Know judgments in cognitive neuroscience: An illustration of the underrepresented point of view. *Learning & Memory, 16(7)*, 406–412. https://doi.org/10.1101/lm.1312809

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117*(4), 1025–1054. https://doi.org/10.1037/a0020874

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

Wulff, L., & Scharf, S. E. (2020). Unpacking stereotype influences on source-monitoring processes: What mouse tracking can tell us. *Journal of Experimental Social Psychology, 87*, 103917. https://doi.org/10.1016/j.jesp.2019.103917

Yang, C., Huang, T. S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language, 99*, 99–110. https://doi.org/10.1016/j.jml.2017.11.005

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. https://doi.org/10.1006/jmla.2002.2864

PART II

EMPIRICAL STUDIES

6. Pilot Study

Exploring the Effects of Different Intervals of Frequency on JOLs

Although this empirical work has not been published in a scientific journal, it was presented in scientific events. Namely, at the 2018 International Meeting of the Psychonomic Society (Amsterdam, The Netherlands), at the 2018 Summer School of the Swiss Graduate School for Cognition, Learning, and Memory, organized by the University of Bern (Weggis, Switzerland), and at the 2019 Third Biennial International Convention of Psychological Science (APS, Paris, France). The abstract submitted for the 2018 International Meeting of the Psychonomic Society was bestowed an Accommodation Award (cf. https://www.psychonomic.org/page/accommodationawards). Additionally, this study was also presented at the workshop for The Center for Doctoral Studies in Social and Behavioral Sciences (CDSS) from the School of Social Sciences at the University of Mannheim, while a guest doctoral student there. The instructions and materials used in this study can be found in Appendix A.

6.1. Introduction

This experiment aimed to replicate the effect of word frequency on JOLs and recall found by previous research (e.g., Begg et al., 1989; Benjamin, 2003; Jia et al., 2016), but across different intervals of word frequency. According to Brysbaert et al. (2018), low-frequency words usually have a frequency per million words (fpm) of less than 5, and high-frequency words have more than 100 fpm. However, as mentioned in the introduction of this dissertation, studies exploring the effects of word frequency on JOLs seem to use different criteria for choosing values of word frequency (Fiacconi & Dollois, 2020). For example, Jia et al. (2016) used low-frequency words of 8 to 9 fpm, and high-frequency words of 185 to 603 fpm. Relying on extreme values of frequency could make word frequency an easily identifiable cue that participants use to make JOLs. In the present study, we manipulated word frequency through different contrasting intervals (the difference between high- and low-frequency words) and explored if participants could identify word frequency as an observable cue.

If beliefs explain the effects of word frequency on JOLs, these effects should disappear in conditions where the difference between high- and low-frequency words is less obvious (with a narrower contrasting interval). Considering the word frequency effect on recall tasks, participants should correctly recall more high-frequency words than low-frequency words. However, when the contrasting interval between high- and low-frequency words is very narrow rather than very wide, effects might disappear.

6.2. Method

6.2.1. Participants

Jia et al. (2016, Exp.1) obtained effect-sizes on recall and JOLs of $d = 1.11$ and $d = .87$, respectively (with a sample of 30 participants). Using the software G*Power (v 3.1), with a $d = .87$, a sample of 20 participants would be enough to reach statistical power of .80. Twenty-eight college students (5 male) with an age average of 23.82 years-old ($SD = 3.90$) participated in this experiment in exchange for course credits.

6.2.2. Design

A 4 (Contrasting interval between high- and low-frequency words: very wide, wide, narrow, very narrow) x 2 (Word frequency: high, low) design was used and both variables were manipulated within-subjects. We measured JOLs, correct recall, and identification of the lists' characteristics.

6.2.3. Materials

Words used in this pilot study were extracted from the Minho Word Pool (Soares et al., 2017), and values for absolute frequency were extracted from the Procura-PALavras (P-PAL, Soares et al., 2014). Four lists of 40 nouns each with different values of frequency (20 high-frequency words, 20 low-frequency words) were created. Analysis of variance showed that lists did not differ significantly in terms of concreteness ($p = .34$), number of letters ($p = .41$) and number of syllables ($p = .53$). Table 1 depicts values of frequency for each list of words. High-frequency and low-frequency words differed significantly in terms of word frequency values within and between each condition (all $p < .001$). Note that designation of "high-frequency" and "low-frequency" words was adopted for comparison purposes but truly words in wide, narrow and very narrow contrasting intervals could be considered mid-frequency words.

Table 1

Mean (Standard Deviation) of Word Frequency (fpm) for Words Used in the Pilot Study.

	Contrasting interval			
	Very wide	**Wide**	**Narrow**	**Very narrow**
High frequency (range)	**95 – 130** 112.64 (10.34)	**80 – 95** 86.17 (4.24)	**65 – 80** 72.34 (4.71)	**50 – 65** 56.66 (4.50)
Low frequency (range)	**5 – 10** 7.57 (1.48)	**10 – 20** 13.39 (3.21)	**20 – 30** 24.12 (3.28)	**30 – 40** 34.87 (3.29)

6.2.4. Procedure

Participants completed the experiment individually on a computer, inside soundproof cabins. After giving informed consent, participants were informed that they would see lists of words which they would later have to remember and that, for each word, they should rate the probability of later remembering it on a scale of 0 (sure that they would not remember the word) to 100 (sure that they would remember the word). Four cycles were then completed, one for each contrasting interval condition. Words were randomized within each contrasting interval condition and order of condition was counterbalanced across participants. Each cycle started with the study phase. First, a fixation cross was presented in the middle of the screen for 500ms before each word. Then, a word was presented for 2 seconds in the center of the screen and followed by a new screen with a blank space where participants made self-paced JOLs for the previous word by writing down any number between 0 and 100. Primacy and recency buffers (2 high-frequency words + 2 low-frequency words) were presented for each list, but not analyzed. After the 40 words, participants completed a 3-minutes filler task (arithmetic calculus), and then a 4-minutes free recall task. This cycle was repeated four times, one for each contrasting interval condition. At the end of the four cycles, participants were asked if they could identify any characteristic of the lists ("Were you able to identify any characteristic of the lists you have studied? If so, please identify for each list"). Participants answered this last question in a blank space referring to each list by its order of presentation. There was no explicit information about the frequency of the words presented.

6.3. Results

For the analyses conducted, the alpha level was set to .05 and all effect sizes are reported in terms of partial eta squared for ANOVAs, and Cohen's *d* for *t*-tests. Main descriptive statistics for JOLs and proportion of correct recall are presented in Table 2.

6.3.1. Judgments of Learning

A 4 (Contrasting interval) x 2 (Word frequency) repeated-measures ANOVA was conducted for mean JOLs. This analysis revealed no effects of contrasting interval, $F(3, 25) = 1.06$, $p = .372$, $\eta_p^2 = .04$, no effects of word frequency, $F(1, 27) = .25$, $p = .622$, $\eta_p^2 = .01$, and no significant interaction, $F(3, 25) = 2.01$, $p = .119$, $\eta_p^2 = .07$. Considering their theoretical interest, we conducted pairwise comparisons to explore differences between JOLs for high- and low-frequency words within each contrasting interval condition. Applying a Bonferroni correction, the alpha was set to .013 (four comparisons). There was no significant difference between participants' JOLs for high- and low-frequency words within each contrasting interval condition. These results showed that, overall, participants gave similar JOLs to high- and low-frequency words across all contrasting interval conditions.

6.3.2. Recall

A 4 (Contrasting interval) x 2 (Word frequency) repeated-measures ANOVA was conducted on mean proportion of correct recall. This analysis showed no significant effect of contrasting interval, $F(3, 25) = 2.58$, $p = .06$, $\eta_p^2 = .09$. However, there was a main effect of word frequency, $F(1, 27) = 4.87$, $p = .04$, $\eta_p^2 = .15$. Correct recall was higher for low-frequency words ($M = .27$, $SD = .16$) than high-frequency words ($M = .24$, $SD = .14$). The interaction was not significant, $F(3, 25) = .45$, $p = .72$, $\eta_p^2 = .02$. Four pairwise comparisons (corrected alpha set to .013) showed no differences between correct recall of high- and low-frequency words within each contrasting interval condition.

Table 2

Mean (Standard Deviation) for Judgments of Learning and Correct Recall in the Pilot Study.

	Very wide	**Wide**	**Narrow**	**Very narrow**
Judgments of Learning				
High Frequency	33.79 (15.00)	34.90 (12.65)	33.42 (16.42)	32.91 (15.95)
Low Frequency	33.49 (13.47)	36.32 (12.80)	30.61 (15.87)	33.09 (15.66)
Correct Recall				
High Frequency	.24 (.15)	.23 (.14)	.26 (.14)	.24 (.13)
Low Frequency	.23 (.16)	.26 (.15)	.30 (.16)	.26 (.16)

6.3.3. Identification of Word Frequency as an Observable Cue

Most participants (75%) answered "no" to the question on cue identification. Only seven (25%) participants identified any characteristic in the studied words. Four participants mentioned categories of words (e.g., "words related to holidays"); one participant mentioned the theme of the list; one participant identified concreteness as a characteristic; and one participant mentioned word length. No participant reported word frequency as a characteristic of the studied lists, suggesting word frequency was not an observable cue across lists.

6.4. Discussion

In this pilot study, we found no effect of word frequency on JOLs, and no participant identified word frequency as an observable cue. While previous research also failed to find effects of word frequency on JOLs (e.g., Susser & Mulligan, 2015), the absence of the effect here was probably because the word lists were not contrasting enough to produce effects on JOLs. Also, the fact that no participant identified word frequency could suggest word frequency to be difficult to identify, but the simplicity of the question used poses many limitations. For example, after four study-test cycles, participants could simply answer "no" as to be done with the experiment.

Regarding recall performance, our results revealed an overall higher recall performance for low-frequency words, which was unexpected. While some studies have found better recall performance for low-frequency than high-frequency words (e.g., Ozubko & Joordens, 2007), usually recall is either better for high-frequency words (Popov & Reder, 2020) or follows a U-pattern (Lohnas & Kahana, 2013). In Lohnas and Kahana's (2013) study words ranging from 37 to 1575 fpm were categorized as mid-frequency and had worse recall than words with both lower and higher frequencies. Most of the words used here could be in truth mid-frequency words, which could justify the similar recall performance across conditions and between our high- and low-frequency words. However, most research on metamemory has found effects of word frequency on recall performance for intervals of frequency like those used in the very high contrasting interval. In fact, Jia et al. (2016) used high-frequency words ranging from 185 to 603 fpm, obtaining higher correct recall for high-frequency words. In that sense, it is possible that effects of word frequency in recall may only arise when using extreme values of word frequency. The fact that even in the very high contrasting interval we did not find any benefit for high-frequency words further suggests that differences between high- and low-frequency words were not substantial to produce an effect.

Critically for interpreting the results of this study, at the moment it was conducted we did not consider measures of subjective frequency or Zipf. Post-hoc calculations of these values showed that although Zipf values between all conditions differed significantly within each condition, all of them were higher than 3 for every condition (cf. Appendix A). Low-frequency words should have a Zipf between 1 and 3 (Brysbaert et al., 2018), which could be the reason why there was no effect of word frequency on memory and metamemory.

References

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. https://doi.org/10.1016/0749-596X(89)90016-8

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31*(2), 297–305. https://doi.org/10.3758/BF03194388

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Fiacconi, C. M., & Dollois, M. A. (2020). Does word frequency influence judgments of learning (JOLs)? A meta-analytic review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale.* Advance online publication. https://doi.org/10.1037/cep0000206

Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2016). The effect of word frequency on judgments of learning: Contributions of beliefs and processing fluency. *Frontiers in Psychology, 6*:1995. https://doi.org/10.3389/fpsyg.2015.01995

Lohnas, L. J., & Kahana, M. J. (2013). Parametric effects of word frequency in memory for mixed frequency lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1943–1946. https://doi.org/10.1037/a0033669

Ozubko, J. D., & Joordens, S. (2007). The mixed truth about frequency effects on free recall: Effects of study list composition. *Psychonomic Bulletin & Review, 14*(5), 871–876. https://doi.org/10.3758/BF03194114

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review, 127*(1), 1–46. http://doi.org/10.1037/rev0000161

Soares, A. P., Iriarte, Á., Almeida, J. J., Simões, A., Costa, A., França, P., Machado, J., & Comesaña, M. (2014). Procura-PALavras (P-PAL): Uma nova medida de frequência lexical do Português Europeu contemporâneo. *Psicologia: Reflexão e Crítica, 27*(1), 110–123. https://doi.org/10.1590/s0102-79722014000100013

Soares, A. P., Costa, A. S., Machado, J., Comesaña, M., & Oliveira, H. M. (2017). The Minho Word Pool: Norms for imageability, concreteness, and subjective frequency for 3,800 Portuguese words. *Behavior Research Methods, 49*(3), 1065–1081. https://doi.org/10.3758/s13428-016-0767-4

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review, 22*(4), 1014–1019. https://doi.org/10.3758/s13423-014-0768-1

7. Study 1

Word Frequency Effects on Judgments of Learning: More Than Just Beliefs

This empirical work was published in the Journal of General Psychology. Note that it was published under the 6th Edition of the APA style but here it was updated to the 7th Edition. The Journal of General Psychology is an international publication from Taylor & Francis devoted to research in experimental psychology. According to Scopus, as of 2019 it had a CiteScore of 1.0, a Scimago Journal & Country Rank (SJR) of 0.298 and a Source Normalized Impact per Paper (SNIP) of 0.467. Regarding quartiles, according to SJR website it was ranked in Q4 for the topic of Experimental and Cognitive Psychology, and Q2 for Arts and Humanities (miscellaneous). According to Clarivate InCites Journal Citation Reports (JCR), in 2019 this journal was ranked in Q4 for Psychology, Multidisciplinary and its Impact Factor was of 0.409.

Reference:

Mendes, P. S., Luna, K., & Albuquerque, P. B. (2019). Word frequency effects on judgments of learning: More than just beliefs. *The Journal of General Psychology*. Advance online publication. https://doi.org/10.1080/00221309.2019.1706073

In addition, results from this empirical work were partially or fully presented at two scientific events: At the 2018 Summer School of the Swiss Graduate School for Cognition, Learning, and Memory, organized by the University of Bern (Weggis, Switzerland), and at the 2019 Third Biennial International Convention of Psychological Science (Paris, France). Additionally, results from this study were also discussed at the CDSS workshop from the School of Social Sciences at the University of Mannheim, while a guest doctoral student there. The instructions and materials used in this study, as well as its Supplemental Materials can be found in Appendix B.

7.1. Abstract

Judgments of learning (JOLs) are usually higher for high-frequency words than for low-frequency words, which has been attributed to beliefs about how word frequency affects memory. The main goal of the present study was to explore if identifying word frequency as a relevant cue is necessary for it to affect JOLs. The idea is that for one to base judgments in beliefs of how a variable affects memory, one must first consider that variable. In Experiments 1 and 2, participants studied a list of high- and low-frequency words, made immediate JOLs, and answered questions aimed at identifying the cues used to make those JOLs. The results showed that identifying word frequency as a cue was not necessary for effects on JOLs to occur, suggesting that some participants could not have used beliefs about how word frequency affects memory when making JOLs. In Experiment 3, we measured processing fluency of high- and low-frequency words through a lexical decision task. Participants identified high-frequency words quicker than low-frequency words, suggesting the former to be more fluently processed. In Experiment 4, we explored if response times in a lexical decision task mediated the effect of word frequency on JOLs. Results showed a significant mediation of 8% to 13%, depending on the analysis technique. We argue that theory-driven processes do not fully account for word frequency effects on JOLs.

Keywords: judgments of learning, metamemory, word frequency, beliefs, fluency

7.2. Introduction

Our metamemory consists of the structures and processes through which we examine the contents of our memories, making judgments about them (Metcalfe & Dunlosky, 2008). One popular way to measure metamemory is through judgments of learning (JOLs), which are predictions of later performance in a memory test (Nelson & Narens, 1990; Soderstrom et al., 2016; Rhodes, 2016). Usually, participants are asked how likely they will later remember a specific stimulus and can be requested to do so either before being exposed to the item (pre-study JOL) or after (post-study JOLs or simply JOLs). According to Nelson and Narens (1990), these judgments of future performance are a way of monitoring one's learning processes, which can guide subsequent study strategies, with a definite impact on actual learning. For example, a student may adapt her study strategies for an upcoming exam based on her thoughts on how well she will remember the studied material. Erroneous judgments about memory could misguide people's subsequent learning strategies, impairing retention of information. Therefore, it is essential to understand which cues guide people's JOLs and the processes behind the act of making JOLs.

JOLs can be based on theory-driven processes (e.g., beliefs) that originate from an analytic inference of how a certain cue will affect memory, but also from nonanalytic, experience-driven processes that derive from direct experiences with items (e.g., processing fluency; Kelley & Jacoby, 1996; Koriat, 1997). Evidence for experience-driven effects on JOLs comes from studies exploring processing fluency on different variables, such as font size (Rhodes & Castel, 2008), auditory volume (Rhodes & Castel, 2009), semantic relatedness (Undorf & Erdfelder, 2015), or generation of images (Besken, 2016). A common way to assess differences in fluency is by measuring participants' response times in lexical decision tasks (e.g., Luna et al., 2019; Mueller et al., 2014; Undorf & Erdfelder, 2015; Undorf & Zimdahl, 2019) or study time during encoding (e.g., Jia et al., 2016; Undorf & Ackerman, 2017; Undorf & Erdfelder, 2015). Items that require either less time to be studied or have quicker response times are considered to be processed more fluently.

However, some studies have suggested that JOLs are mainly theory driven, for example, based on beliefs about the functioning of memory (Koriat et al., 2004; Jia et al., 2016; Mueller et al., 2014; Susser et al., 2016; Undorf & Zimdahl, 2019). Strong evidence in support of this perspective originates from pre-study JOLs. Even when participants make judgments *before* processing the material to be learned, effects on JOLs still occur (see Mueller et al., 2014), showing that direct experience with an item is not necessary to affect JOLs. In this line, the analytic-processing theory (Dunlosky et al., 2015;

Mueller et al., 2016; Mueller et al., 2013) states that theory-driven processes are the main bases of JOLs, and adds that, besides having a priori beliefs, people may develop online beliefs through an analytic search of variability in items that could help identify cues that could impact memory. In fact, results from Mueller and Dunlosky (2017) showed that participants who were led to believe that one color was more fluently processed than the other gave higher JOLs to words presented in that color, which suggests that people's beliefs about fluency (and not fluency per se) play an important role in making JOLs.

Recent research has also shown that both theory-driven and experience-driven processes may contribute together to the effects of different cues on JOLs (e.g., Blake & Castel, 2018; Frank & Kuhlmann, 2017; Hu et al., 2015; Koriat et al., 2004; Price & Harrison, 2017; Rhodes, 2016; Undorf, Zimdahl, & Bernstein, 2017; Yang et al., 2018). For example, Frank and Kuhlmann (2017) explored the role of fluency and beliefs in volume effects on JOLs. They concluded that both theory-based and experience-based factors contributed to the effect. Similar conclusions were reached by Undorf et al. (2017) and Yang et al. (2018) for the effect of font size on JOLs. Critically, it is possible that the contribution of both experience- and theory-driven processes may vary for different cues. Thus, although volume and font size effects on JOLs seem to be better explained by a combination of fluency and beliefs, with a higher role of beliefs (see Luna et al., 2019), other cues, such as word frequency, could be affected exclusively by beliefs, by fluency, or by a mix of both. Our objective with the present study was to extend the discussion of the factors affecting JOLs to another cue, namely word frequency.

Word frequency affects people's JOLs (e.g., Begg et al., 1989), but the mechanisms behind the effect have not been fully explained yet. Although word frequency effects on memory depend on the type of test (Lohnas & Kahana, 2013), participants usually attribute higher JOLs to high-frequency words than to low-frequency words (Begg et al., 1989; Benjamin, 2003; Hourihan et al., 2017, Experiment 3; Jia et al., 2016; but for an exception see Susser & Mulligan, 2015).

On the one hand, some research suggested that the effect of word frequency on JOLs may be based on experience factors. For example, people are quicker to name or identify (in lexical decision tasks and semantic decision tasks) high-frequency words than low-frequency words (Balota & Chumbley, 1984; Brysbaert et al., 2018; Gao et al., 2016; Schilling et al., 1998), which suggests that high-frequency words are more fluently processed. In fact, Begg et al. (1989, Experiment 1) concluded that JOLs were based on the subjective experience of fluency, since participants considered high-frequency words as more easily processed, more easily imagined, and as more memorable than low-frequency words.

On the other hand, Benjamin (2003) found that although predictions of future memory were higher for high-frequency words before a recognition test, participants gave higher predictions of future memory performance for low-frequency words in a second study-test cycle, arguing that participants' metacognitive knowledge was based in deliberated analytic processes. In fact, Jia et al. (2016) found support for theory-driven factors and evidence against experience-driven factors in the word frequency effect on JOLs. The authors manipulated word frequency and measured processing fluency either through a self-paced study procedure (Experiment 2a) or by presenting words in easy- or difficult-to-read font style (Experiment 2b). Results showed no difference in either measure, even though the effects of word frequency on JOLs and recall were still found. These results suggest no role of fluency for the word frequency effect on JOLs and provide evidence against experience-driven factors. In addition, the authors also explored the role of beliefs through a questionnaire (Experiment 3a) and pre-study JOLs (Experiment 3b). In the questionnaire, participants estimated how many high- and low-frequency words hypothetical participants in another experiment would have remembered. In the pre-study JOLs, participants were informed about the frequency of the word they would study immediately afterward and were asked to make a JOL before actually seeing the word. Critically, in both experiments, participants were not exposed to the words while making JOLs and, thus, processing fluency could not affect the results. The word frequency effects were still observable, suggesting that beliefs play a strong role in the word frequency effect on JOLs.

Based on the research above, here we tested the idea that if word frequency effects on JOLs are solely based on theory-driven processes, namely on beliefs about how word frequency affects memory, then it would be expected that participants would identify word frequency as a cue while making JOLs. Identifying word frequency as a cue would allow participants to make JOLs based on theory-driven analytic processes of how word frequency affects memory (e.g., belief that common words are better remembered). Nonetheless, if word frequency is not identified, participants cannot develop or access beliefs about word frequency effects on memory. Rather, they would base their JOLs on processing fluency, beliefs about fluency, or even other cues besides word frequency.

Thus, the present research aimed to explore if identifying word frequency as a cue is necessary for word frequency effects on JOLs to occur. If effects of word frequency on JOLs are based on theory-driven processes, namely beliefs about how word frequency affects memory, one must know that the studied words differ in frequency. To do so, we manipulated word frequency in four experiments. In Experiment 1, we asked participants to make JOLs for high- and low- frequency words, and to identify any cue that might have helped them make their JOLs. If participants could not identify word frequency as an observable cue, they could not have developed or access beliefs about its impact on memory. Hence, any effect of word frequency on JOLs could not be explained by a belief regarding how word frequency may affect memory. In Experiment 2, we improved the method from the previous experiment to gather a more complete information about the basis of participants' JOLs. In Experiment 3, we tested differences in processing fluency between high- and low-frequency words through a lexical decision task. Finally, in Experiment 4, we explored if response times in a lexical decision task mediated the effect of word frequency on JOLs.

7.3. Experiment 1

The present experiment aimed to explore if participants could identify word frequency as a relevant cue when making JOLs for high- and low-frequency words. The idea is that participants might identify word frequency as a relevant cue, which could explain variations in their JOLs. However, participants might not be able to detect word frequency as a cue and base their JOLs on other cues (hence, not on beliefs about how word frequency affects memory). We expected higher JOLs and higher correct recall for high-frequency words than for low-frequency words.

7.3.1. Method

7.3.1.1. Participants. The sample size was estimated using G*Power software (Faul et al., 2007). Based on Jia et al. (2016) effect size of 0.87 for word frequency effect on JOLs (Experiment 1), a total of 20 participants would be enough to reach a power of .95. In this and all the experiments reported, we aimed for approximately thirty participants.

Thirty-six college students (33 female; 3 male) with an average age of 20.75 years old (SD = 2.79) participated in this experiment in exchange for course credits. All the experiments were conducted in Portugal, and all participants included in the analyses were native European Portuguese speakers.

7.3.1.2. Design and Materials. The only independent variable was word frequency, which was manipulated within-participants. JOLs, correct recall, and correct identification of word frequency as cue were measured.

In a pilot study, we found no differences in JOLs for high-frequency words (range [95, 130] and low-frequency words (range [5, 10]), which lead us to further differentiate values of frequency in the present study. A list of 40 nouns (20 high-frequency, 20 low-frequency) was created (see Appendix B or Supplemental Materials at https://osf.io/uyh9z). Words were extracted from the Minho Word Pool (Soares et al., 2017), with absolute frequency per million extracted from the P-PAL online tool (Soares et al., 2014). Frequency for high-frequency words (e.g., *newspaper*, *jornal* in the original Portuguese) was higher (M = 303.41, SD = 146.74, range [160, 758]), than for low-frequency words (e.g., *clupea*, *arenque* in Portuguese; M = .27, SD = .25, range [0.009, 0.81]), $t(19)$ = 9.25, p < .001, Cohen's d = 2.07. Both set of words did not differ on concreteness (M_{HF} = 4.00, M_{LF} = 4.10; scale 1 to 7) , $t(19)$ = .43, p = .67, Cohen's d = 0.10, number of letters (M_{HF} = 6.40, M_{LF} = 6.40), $t(19)$ = 0, p = 1, Cohen's d = 0, nor number of syllables (M_{HF} = 2.55, M_{LF} = 2.65), $t(19)$ = .57, p = .58, Cohen's d = 0.13.

7.3.1.3. Procedure. Participants completed the experiment individually on a computer, inside soundproof cabins. After giving informed consent, participants read instructions stating they would see a list of words that they would later have to remember and, for each word, they should rate the probability of later remembering it on a scale of 0 (sure that they will not remember the word) to 100 (sure that they will remember the word). First, a fixation cross was presented in the middle of the screen for 500ms before each word. Then, a word was presented for 2 seconds in the center of the screen, followed by a new screen with a blank space where participants made self-paced JOLs for the previous word by writing down any number between 0 and 100. Primacy and recency buffers (2 high-frequency words + 2 low-frequency words) were presented, but not analyzed. The order of presentation was pseudo-randomized. After the 44 words, participants completed a 3-minutes filler task (arithmetic task) and then a 4-minutes free recall task. At the end of the experiment, participants were asked if they could identify any characteristic of the lists ("Were you able to identify any characteristic of the list of words you have studied? If so, please specify"). Participants answered this last question in a blank space, and if they were able to identify any characteristic from the list, they were presented with the words that they recalled and asked to categorize the words according to that characteristic. Across the experiment, there was no explicit information about the frequency of the words presented. Overall, the procedure took about 20 minutes.

7.3.2. Results

For all the analyses conducted on the four experiments, the alpha level was set to .05, and all effect sizes are reported in terms of Cohen's d for t-tests, and partial eta squared for ANOVAs. Paired t-

tests were conducted for both participants' mean JOLs and the mean proportion of correct recall. For completeness, gamma correlations were calculated and are reported in the Supplemental Materials (available at https://osf.io/uyh9z, and in Appendix B.3)[4], but these results will not be discussed.

7.3.2.1. Judgments of Learning. Participants gave higher JOLs for high-frequency words (M = 50.96, SD = 14.69) than for low-frequency words (M = 21.55, SD = 10.90), $t(35) = 11.08$, $p <$ .001, Cohen's d = 1.85.

7.3.2.2. Recall. Proportion of correct recall was significantly higher for high-frequency words (M = .33, SD = .12) than for low-frequency words (M = .13, SD = .06), $t(35) = 9.53$, $p < .001$, d = 1.59.

7.3.2.3. Identification of Word Frequency as a Cue. Answers were classified by two researchers through discussion and classified after reaching an agreement. Seven out of the 36 participants (19.44%) correctly identified word frequency or mentioned one characteristic that could be considered comparable to the concept of *word frequency*: "frequency", "known and unknown words", "common and uncommon words", "daily usage of words", "difficulty" (one participant each), and "words and non-words" (two participants). As the presented low-frequency words were really uncommon and could be unknown to some participants, we accepted the identification of "words and non-words" as a categorization of word frequency. Together, these seven participants correctly recalled a total of 58 words and, when asked to categorize the recalled words into the categories they mentioned (e.g., "known and unknown words"), they correctly categorized 53 of them (91.4%). This correct categorization corresponds to correctly identifying, for example, "known words" as high-frequency words, and "unknown" as low-frequency words. Finally, three participants (8.3%) mentioned other cues (e.g., "words ending with -ão"), and the remaining participants (26) did not produce any characteristic (i.e., answered "no" or similar).

Seven participants were able to identify word frequency as a cue while making JOLs. However, the remaining 29 participants were not able to identify word frequency as a cue and, thus, could not have developed a belief about word frequency. We repeated the analyses on JOLs with these 29 participants, and the word frequency effect on JOLs was still observable. Participants gave higher JOLs for high-frequency words (M = 51.85, SD = 15.11) than for low-frequency words (M = 22.02, SD = 11.22), $t(28) = 9.67$, $p < .001$, Cohen's d = 1.80.

4 These results are included in this dissertation in Chapter 10, in which JOL accuracy was examined.

7.3.3. Discussion

In Experiment 1, participants made JOLs for high- and low-frequency words followed by a recall task, and results showed higher JOLs and higher correct recall for high-frequency words. Only a few participants identified word frequency as a cue, and participants who did not identify it still rated high-frequency words with higher JOLs.

Jia et al. (2016) attributed the effects of word frequency on JOLs exclusively to beliefs about how word frequency may affect memory. However, this experiment suggested that beliefs are not needed for the effect to occur, because even when removing participants who identified word frequency as a cue, the magnitude of the effect was similar.

Experiment 1 had some limitations. First, we asked one open question not directed at JOLs cues but at word characteristics, and thus participants could have based their answers, for example, on the words retrieved at test. Although that would be informative about the chances to identify the frequency manipulation, that would be only indirectly linked with the chances that it was detected during study. And second, the question required an answer, but participants could have answered that they identified nothing just to finish the experiment earlier and without an actual effort to search and provide a cue. This may have artificially increased the number of participants who did not mention frequency as a cue. To solve these limitations, we conducted Experiment 2.

7.4. Experiment 2

In Experiment 2, the objective was the same as in Experiment 1, but we asked more questions regarding participant's identification of the characteristics of the list of words to get a more nuanced information about the use of word frequency as a cue that informs participants' JOLs.

7.4.1. Method

7.4.1.1. Participants. Thirty-three college students participated in this experiment in exchange for course credits. Three were excluded for not being native speakers of European Portuguese. A final sample of 30 participants (27 female; 3 male) with an average age of 21.77 years old (*SD* = 5.30) who had not participated in Experiment 1 was considered for analysis.

7.4.1.2. Design and Materials. Materials and design were the same as in Experiment 1 except for the final questions (see below).

7.4.1.3. Procedure. The procedure was the same as in Experiment 1 with two modifications. First, between study and recall phases, participants were questioned about their JOLs through three questions, which are presented in Table 3. All questions aimed to capture the use of word frequency as a basis of participants' JOLs, with different levels of precision and bias. We started with an open-ended question that would give us an unbiased answer for the bases of participants' JOLs, then moved to present a hint that could foster identification of word frequency. Finally, we addressed directly the use of word frequency by explicitly stating that word frequency was manipulated. Answers were recorded and later transcribed to be categorized. After these questions, the experiment proceeded as in Experiment 1 with an arithmetic task for two minutes, and a four-minute free-recall test. Overall, the procedure took about 20 minutes.

Table 3

Questions Used in Experiment 2 to Evaluate if Participants Identified Word Frequency as a Cue.

Level of analysis	Question	Categorization of answers
General	1. Your judgments of learning for the words (these evaluations you made) varied throughout the experiment. What did you base your judgements on?	No (does not mention word frequency) Yes (mentions word frequency)
Specific	2. There is a characteristic of the words from which you can divide the list in two groups. Can you tell which characteristic that is?	No (does not mention word frequency) Yes (mentions word frequency)
Very specific	3. In the list you studied, there were high- and low-frequency words (common and rare words). Did you base your judgments on the frequency of the words?	No; Yes.

7.4.2. Results

7.4.2.1. Judgments of Learning. Participants gave higher JOLs for high-frequency words (M = 54.37, SD = 16.03) than for low-frequency words (M = 27.33, SD = 11.91), $t(29) = 9.49$, $p < .001$, Cohen's d = 1.73.

7.4.2.2. Recall. The proportion of correct recall was significantly higher for high-frequency words (M = .24, SD = .14) than for low-frequency words (M = .15, SD = .09), $t(29) = 4.24$, $p < .001$, Cohen's d = 0.78.

7.4.2.3. Identification of Word Frequency as a Cue. The answers were classified by two researchers (inter-rater agreement of 94%). When the classification differed, the items were discussed with a third researcher and classified after reaching an agreement. For Question 1, only two participants unambiguously mentioned word frequency as a base for their JOLs (e.g., mentioning common and rare words), but others used different ideas of word frequency. For this reason, we classified participants' answers as "word frequency" when participants mentioned subjective frequency (e.g., "words I often use"), the idea of knowing/not knowing the words (e.g., "strange words", "words I did not know or that were weird"), and familiarity (e.g., "words that seemed more familiar").

Table 4

Number of Participants who Mentioned Frequency as Basis for JOLs in Experiment 2 (N = 30).

	Question 1	**Question 2**	**Question 3**
Used/detected word frequency	19 (63%) *	11 (37%)	29 (97%)
Did not use/detect word frequency	11 (37%)	19 (63%)	1 (3%)

* In Question 1, six participants (20%) mentioned using word frequency as the only base to make their JOLs, and 13 participants (43%) mentioned using word frequency and also other cues (e.g., extension of the words or associations they could make).

Percentages of participants who identified or did not identify word frequency as a cue for all questions are presented in Table 4. The majority of participants stated that they based their JOLs on word frequency, but more than one-third of participants did not (Question 1). In contrast, in Question 2, the majority of participants did not mention word frequency. In Question 3, only one participant did not state to have used word frequency while making JOLs. We also looked at patterns of response by participant across the three questions. Nine participants (30%) answered consistently that they were using word frequency as a cue in all three questions, and eight participants (26.67%) only agreed that they identified/used frequency in Question 3 but not in previous questions. Ten participants (33,33%) mentioned word frequency in Questions 1 and 3 but not in Question 2. Finally, two participants (6.67%) mentioned word frequency in Questions 2 and 3 (but not in Question 1), and only one participant (3.33%) did not mention word frequency in all three questions.

To explore if effects of word frequency on JOLs would still be found even for participants who did not base their JOLs on word frequency in Question 1, a paired *t*-test comparing JOLs for high- and low-frequency words was conducted for the 11 participants who did not identify word frequency. Preliminary tests of normality and homogeneity of variance showed non-compliance with the assumptions underlying the use of parametric tests. For this reason, considering the suggestion of Fife-Schaw (2006), we performed the parametric and nonparametric tests. Since both types of tests revealed similar results, we report the results of the parametric tests. Participants who did not base their JOLs on word frequency still gave higher JOLs for high-frequency words (M = 56.41, SD = 16.73) than for low-frequency words (M = 27.31, SD = 14.61), t(10) = 4.51, p < .001, Cohen's d = 1.36.

In Experiment 1, we asked participants to identify a cue from the lists of words they had studied, and Question 2 in the present experiment was the one most similar to that (i.e., just identifying a cue). Thus, we conducted another analysis comparing JOLs of participants who did not mention word frequency in Question 2. Again, preliminary tests showed non-compliance with the assumptions underlying the use of parametric tests. Parametric and nonparametric tests revealed similar results, and we report those from the parametric tests. For participants who did not identify word frequency as a characteristic to divide the list in two categories, JOLs for high-frequency words (M = 54.92, SD = 12.19) were higher than for low-frequency words (M = 24.17, SD = 9.48), t(18) = 10.81 p < .001, Cohen's d = 2.48.

7.4.3. Discussion

Results from Experiment 2 showed that word frequency affected JOLs and recall, with participants giving higher JOLs and correct recall for high-frequency words. The focus of this experiment was to understand if participants base their JOLs in word frequency through three different questions. The majority of participants seemed to base their JOLs on word frequency, but some of them (n = 13) also reported using other cues besides word frequency. These results suggest that making JOLs may be a complex issue and that participants may integrate different cues while making JOLs (Undorf et al., 2018). Even so, more than one-third of the participants did not base their JOLs on word frequency and, still, gave higher JOLs for high-frequency words. Similarly, results from Question 2 showed a similar pattern to that of Experiment 1, in which the majority of participants were not able to identify a characteristic from the studied list even when a hint about two categories of items was provided. Finally, when asked directly if they used word frequency, 97% of participants said they used it to make their JOLs.

Note, however, that Question 1 would give us an unbiased response for the bases of participants' JOLs, while Question 2 introduced a hint, and Question 3 explicitly informed participants that word frequency was manipulated. Because of this, there is the possibility that Question 3 was suggestive, and participants may have answered "Yes" to comply with the researcher, to sound smart, or to make a good impression. Moreover, the fact that Question 3 followed other two questions, there might have been carry-over effects from previous questions. Nonetheless, word frequency may be a difficult cue to put into words, but if noted by the researcher, it may become obvious to participants that the studied words were very common or very rare. We argue that data from Question 1 is probably more informative than from Question 3. In each case, at least for some participants, JOLs could not have been based on beliefs about the effects of word frequency on memory, because they only mentioned that they were using word frequency when we explicitly told them that word frequency was manipulated (Question 3).

7.5. Experiment 3

Results from Experiments 1 and 2 suggest that beliefs may not lie at the core of the effect of word frequency on JOLs. We conducted Experiment 3 to test the effect of experience-based factors, specifically processing fluency, through a lexical decision task. Jia et al. (2016) found no effect of fluency in participants' JOLs for high- and low-frequency words measured through self-paced study and manipulating the type of font (easy or difficult to read). Attending to different kinds of processing fluency, such as perceptual fluency and lexical fluency (see Alter & Oppenheimer, 2009), it may be that the tasks used by Jia et al. measured perceptual fluency. However, word frequency is basically the number of times that a word may appear in a text, and most if not all the times in which we encounter a word in a text lexical accessing is required. It is possible that high-and low-frequency may not differ in perceptual fluency, but rather in lexical or semantic fluency, meaning that a task that involves lexical decision may be more sensitive to measure fluency effects for linguistic frequency.

7.5.1. Method

7.5.1.1. Participants. Thirty college students (24 female; 6 male) with an average age of 19.90 years old (*SD* = 3.30) participated in this experiment in exchange for course credits. No participant took part in the previous experiments.

7.5.1.2. Design and Materials. A 2 (word-frequency: high, low) x 2 (type of word: word, nonword) was used. Stimuli were 40 target words (from Experiments 1 and 2) and another 40 nonwords that matched each word in length and orthography, which were created by changing the order of the syllables or by changing one or two letters from the original high- and low-frequency words. Four extra items were created for training, and another four primacy (two items) and recency (two items) buffers. Nonwords have no associated frequency, but we named "high-frequency nonwords" those created based on high-frequency words and "low-frequency nonwords" those created based on low-frequency words.

7.5.1.3. Procedure. The procedure followed Mueller et al. (2014, Experiment 1) and Luna et al. (2019, Experiment 3). Participants were asked to decide, as quickly as possible, whether the item on the screen was a word or not by pressing the keys "z" and "m" (counterbalanced). After four training trials, the 84 items were presented. For each trial, an initial screen was displayed with a two-second countdown, asking participants to get ready. Then, a fixation point (+) was presented for 500ms, followed by the item. The item remained on the screen until the participant pressed a key. If a key other than "z" or "m" was pressed, a pop-up window reminded participants about the correct keys and moved to the next trial. Items were presented in a fixed-random order, with the only constraint that no more than two items of the same type (either word or non-word or high- or low-frequency word) were presented consecutively.

7.5.2. Results

As suggested by Leys et al. (2013), outliers were defined as response times above and below three times the median absolute deviation. Five percent of the data were identified as outliers and removed. Response times are presented in milliseconds. Incorrect answers (error rate of about 11%) were removed from the analysis of response times. Mean response times and accuracy per condition are presented in Table 5.

Table 5

Mean (Standard Deviation) Results From the Lexical Decision Task in Experiment 3.

	Words	**Nonwords**	**Total**
Response Times			
High Frequency	828 (124)	1184 (339)	1006 (266)
Low Frequency	1004 (212)	1160 (293)	1082 (339)
Total	916 (194)	1172 (314)	
Accuracy			
High Frequency	.99 (.02)	.96 (.06)	.98 (.05)
Low Frequency	.66 (.18)	.95 (.05)	.80 (.20)
Total	.82 (.21)	.96 (.05)	

7.5.2.1. Response times. We conducted a 2 (word-frequency: high, low) x 2 (type of word: word, nonword) ANOVA. Results showed that participants were quicker for high-frequency items than for low-frequency items, $F(1, 29) = 31.76$, $p < .001$, $\eta_p^2 = .52$. Participants were quicker for words than for nonwords, $F(1, 29) = 13.98$, $p < .001$, $\eta_p^2 = .33$. There was also a significant interaction, $F(1, 29) = 34.75$, $p < .001$, $\eta_p^2 = .55$. Pairwise comparisons showed that participants were quicker for high-frequency words than low-frequency words, $t(29) = 5.56$, $p < .001$, Cohen's $d = 1.02$, but that there was no difference between high-frequency nonwords and low-frequency nonwords, $t(29) = 1.20$, $p = .24$, Cohen's $d = 0.22$.

7.5.2.2. Accuracy. A 2 (word-frequency: high, low) x 2 (type of word: word, nonword) ANOVA showed that participants' accuracy was higher for high-frequency items than for low-frequency items, $F(1, 29) = 105.48$, $p < .001$, $\eta_p^2 = .78$. Accuracy was lower for words than for nonwords, $F(1, 29) = 41.47$, $p < .001$, $\eta_p^2 = .59$. There was also a significant interaction, $F(1, 29) = 77.63$, $p < .001$, $\eta_p^2 = .73$. Pairwise comparisons showed that accuracy was significantly lower for low-frequency words than for high-frequency words, $t(29) = 10.00$, $p < .001$, Cohen's $d = 1.83$, but did not differ between high-frequency nonwords and low-frequency nonwords, $t(29) = .90$, $p = .38$, Cohen's $d = 0.17$.

7.5.3. Discussion

Results from Experiment 3 showed that participants were quicker to identify high-frequency words than low-frequency words, suggesting that fluency is higher for high-frequency words. This contrasts with the results of Jia et al. (2016), who failed to find differences in fluency between high- and low-frequency words by measuring self-paced study and manipulating font readability. We will come back to the differences between our results and Jia et al. (2016) in the General Discussion.

It is important to note that participants' accuracy for low-frequency words was very low compared to the other conditions, which suggests that some words have values of frequency so low that participants may have considered them as nonwords. This suggests the interesting idea that the effect of word frequency on memory and metamemory may be confounded with the distinction between words and nonwords. However, this idea is not directly related to our objectives here and, thus, will be left open for future research.

7.6. Experiment 4

While results from Experiment 3 showed that the high- and low-frequency words differ in fluency, it is still unclear if fluency affected JOLs. Experiment 4 aimed to test if fluency, measured as response times in a lexical decision task, mediates participants' JOLs for high- and low-frequency words. To this end, we followed the procedure used by Undorf and Zimdahl (2019, Experiments 1, 2, and 3), in which participants completed a study with a JOL-recall phase and then completed a lexical decision task. We took advantage of the experiment to try and replicate results from Experiment 2 regarding participants' bases of JOLs, but by making only one question to each participant. Because of this, we increased the sample size in order to have 20 participants answering to one unbiased general question (such as Question 1 in Experiment 2) and 20 participants answering a direct question about using word frequency as a cue to make JOLs (such as Question 3 in Experiment 2).

7.6.1. Method

7.6.1.1. Participants. We collected data from 46 participants, but six were excluded either for incomplete data or because they had participated in the previous experiments. A final sample of forty college students (7 male, 33 female) with an average age of 19.05 years old ($SD = 2.71$) participated in this experiment in exchange for course credits.

7.6.1.2. Design, Materials, and Procedure. Experiment 4 was divided into three phases: 1) a study-JOLs-recall phase; 2) a lexical decision task; and 3) questions. Design, materials, and procedure for the study-JOL-recall phase were the same as in Experiment 1, and for the lexical decision task were the same as in Experiment 3.

In the third and last phase, the question about cue identification was manipulated between participants. Participants were randomly assigned to answer one of two questions: A) "During the first phase of the experiment, what did you base your judgments on for each word?"; B) "During the first phase of the experiment (during judgments), did you notice that words differed in frequency? Did you use frequency as a cue to make your judgments?".

7.6.2. Results

7.6.2.1. JOLs and Recall. JOLs were higher for high-frequency words ($M = 52.76$, $SD = 20.72$) than for low-frequency words ($M = 28.60$, $SD = 16.14$), $t(39) = 9.38$, $p < .001$, Cohen's $d = 1.48$. Participants correctly recalled more high-frequency words ($M = .29$, SD = .17) than low-frequency words ($M = .17$, $SD = .13$), $t(39) = 5.13$, $p < .001$, Cohen's $d = 0.81$.

7.6.2.2. Lexical Decision Task – Response Times. Results replicated those from Experiment 3 and are presented in Table 6. Incorrect answers (about 12%) and outliers (about 5%) were removed. A 2 (word-frequency: high, low) x 2 (type of word: word, nonword) ANOVA showed that participants were quicker for high-frequency items than for low-frequency items, $F(1, 39) = 30.73$, $p < .001$, $\eta_p^2 = .44$, that they were quicker for words than for nonwords, $F(1, 39) = 35.42$, $p < .001$, $\eta_p^2 = .48$, and a significant interaction, $F(1, 39) = 18.73$, $p < .001$, $\eta_p^2 = .32$. Pairwise comparisons showed that participants were quicker for high-frequency words than low-frequency words, $t(29) = 5.72$, $p < .001$, Cohen's $d = 0.91$, but the difference was smaller between high-frequency nonwords and low-frequency nonwords, $t(29) = 3.21$, $p = .003$, Cohen's $d = 0.51$.

Table 6

Mean (Standard Deviation) Results From the Lexical Decision Task in Experiment 4.

	Words	**Nonwords**	**Total**
Response Times			
High Frequency	912 (164)	1395 (635)	1154 (521)
Low Frequency	1241 (459)	1481 (676)	1361 (587)
Total	1076 (381)	1438 (653)	
Accuracy			
High Frequency	.99 (.02)	.94 (.08)	.97 (.06)
Low Frequency	.83 (.14)	.92 (.09)	.88 (.12)
Total	.91 (.13)	.93 (.08)	

7.6.2.3. Lexical Decision Task – Accuracy. Results also replicated those from Experiment 3 and are presented in Table 6. A 2 (word-frequency: high, low) x 2 (type of word: word, nonword) ANOVA showed that participants' accuracy was higher for high-frequency items than for low-frequency items, $F(1, 39) = 63.26$, $p < .001$, $\eta_p^2 = .62$, that accuracy was lower for words than for nonwords, $F(1, 39) = 1.76$, $p < .001$, $\eta_p^2 = .04$, and a significant interaction, $F(1, 39) = 32.80$, $p < .001$, $\eta_p^2 = .46$. Accuracy was lower for low-frequency words than for high-frequency words, $t(39) = 7.49$, $p < .001$, Cohen's $d = 1.19$, but did not differ between high-frequency nonwords and low-frequency nonwords, $t(39) = 1.550$, $p = .13$, Cohen's $d = 0.25$.

7.6.2.4. Mediation Analysis. We conducted a mediation analyses to investigate whether the effect of word frequency on JOLs was mediated by fluency, as measured by response time in the lexical decision task, using two different methods in R software (R Core Team, 2019). We first performed multilevel regression, a technique used before to explore mediation effects on JOLs (Undorf & Erdfelder, 2015, Undorf et al., 2017), using the R packages lme4(Bates et al., 2015), lmerTest (Kuznetsova et al., 2016), and mediation (Tingley et al., 2014).[5] We also performed a Bayesian multilevel mediation analysis using bmlm package (Vuorre, 2017), also previously used to explore mediation effects on JOLs (Luna et al., 2019; Yang et al., 2018).

For the multilevel regression analysis, response times were log-transformed to decrease skewness, moving their distribution closer to normality. Then, two separate multilevel regression models (level 1: items, level 2: participants) were performed, in which participants were treated as random effects, and word frequency and response times were treated as fixed effects. The unstandardized regression coefficients (all significant, $p < .001$) were of -.25 for the direct effects of word frequency on response times; 19.14 for the direct effect of word frequency on JOLs; and of -11.06 for the direct effect of response times on JOLs. The indirect effect of word frequency on JOLs mediated by response times was estimated with 10 000 simulations and was found to be significantly different from zero (2.82, 95% CI [1.71, 3.96], $p < .001$). This means that high-frequency words increased JOLs indirectly by decreasing response times. In this analysis the percentage mediated was of 13%, 95% CI [8%, 18%], $p < .001$.

The Bayesian analysis estimates the mediated effect through four Markov Chain Monte Carlo (MCMC) chains and 10 000 iterations per chain. We entered word frequency as the manipulated variable, response times in the lexical decision task as the mediator, and JOLs as the outcome. The results showed that response times explained 8%, CI [3%, 14%], of the effect of word frequency on JOLs. In sum, results from both methods are consistent and show a mediation of response times in the word frequency effect on JOLs.

7.6.2.4.5. Identification of Word Frequency as a Cue. Answers were classified by two researchers, with an inter-rater agreement of 82.5%. The answers that differed in classification were discussed with a third researcher and classified after reaching an agreement. All participants gave an answer, that is, mentioned at least one cue as a base for their JOLs. Four participants out of 20 on each condition did not mention word frequency. From the 32 participants across conditions who used word frequency, ten (25%) also mentioned additional cues (e.g., serial position, personal associations). Interestingly, four of the participants who stated to have used word frequency expected to remember low-frequency words better. As in Experiments 1 and 2, we repeated analysis on JOLs of participants

---

5. We thank Dr. Monika Undorf for helping us with this analysis.

who did not identify word frequency as a cue. These participants gave higher JOLs to high-frequency words (M = 46.71, SD = 25.17) than low-frequency words (M = 24.08, SD = 10.38), $t(7) = 3.39$, p = .01, Cohen's d = 1.20.[6]

7.6.3. Discussion

In Experiment 4, we replicated the results from previous experiments. The mediation analyses with two different methods consistently showed a mediation effect of response times on word frequency effects on JOLs, suggesting the influence of fluency on participants' JOLs for high- and low-frequency words.

7.7. General Discussion

In the present study, we investigated the mechanisms underlying the word frequency effect on JOLs and, in particular, whether participants were able to identify word frequency as a relevant cue that they could use to make JOLs. To that end, we manipulated word frequency and measured JOLs and recall. We will focus this Discussion on the mechanism by which word frequency may affect JOLs and, thus, we will only briefly discuss the effects of frequency on recall. The main conclusion from this research is that beliefs do not fully account for word frequency effects on JOLs and that processing fluency as measured by response times in a lexical decision task mediates, at least in part, this effect. We will address the key findings by arguing that at least some participants could not have used beliefs about how word frequency affects memory to make JOLs, and then by presenting alternative explanations for our results.

In the present study, results from three experiments (Experiments 1, 2, and 4) consistently replicated past research of higher correct recall for high-frequency words (e.g., Jia et al., 2016; Lohnas & Kahana, 2013). Also, considering the low accuracy in the lexical decision task for low-frequency words, our results support the notion that high-frequency words are known to more people, which could explain easier semantic encoding, and easier access when retrieving information during recall (see Brysbaert et al., 2018).

Experiments 1, 2, and 4 showed a clear effect of word frequency on JOLs, with higher JOLs to high-frequency words, as previously found in the literature (Begg et al., 1989; Benjamin, 2003; Hourihan et al., 2017; Jia et al., 2016). Our goal, however, was not to replicate those effects but to

6. Nonparametric tests revealed similar results

explore if participants used beliefs about word frequency to make JOLs. Previous research by Jia et al. (2016) showed that the effects of word frequency on JOLs might depend on beliefs about how frequency affects memory performance. People can have pre-existing beliefs about word frequency, or develop them during the experimental procedure, as suggested by the analytic-processing theory (Dunlosky et al., 2015; Mueller et al., 2016). In any case, to rely on such beliefs to make JOLs, it seems necessary to identify a specific variable as a cue through deliberate, analytic processes. Our results showed that this does not seem to be the case with word frequency.

Critically, in the present study, identifying word frequency as a cue did not seem to be necessary for word frequency to influence JOLs. Indeed, the majority of participants in Experiment 1 did not identify word frequency as a cue to make JOLs. Additionally, in Experiment 2, when asked about the basis of their JOLs (Question 1), only a minority of participants mentioned that they based their JOLs on word frequency alone, with the majority relying on other cues besides word frequency, and about a third of participants not mentioning word frequency at all. In Experiment 4, the majority of participants reported using word frequency as the base of their JOLs. Even so, in all three experiments, word frequency affected JOLs from participants who did not report noticing or using word frequency as a cue. While it has been established that people may have beliefs about how word frequency affects memory, and these beliefs may affect participants' JOLs (Jia et al., 2016), our results suggest that those beliefs may not fully account for effects on JOLs.

Since beliefs about word frequency may not fully account for effects on JOLs, other mechanisms must be at play, such as differences in processing fluency (e.g., Rhodes & Castel, 2008). We found that high-frequency words are more fluently processed than low-frequency words. These differences may be the basis for participants’ JOLs, at least for those who did not identify word frequency as a cue. Additional support for this argument comes from the lower accuracy of low-frequency words when compared to high-frequency words in Experiment 3. This finding suggests that participants had trouble accessing lexical information of low-frequency words. Attending to the concept of lexical fluency (Alter & Oppenheimer, 2009), it is possible that some participants could not access the meaning of those words, perceiving them as more difficult, or even as nonwords, as suggested by some participants’ answers to questions in Experiments 1 and 2. This could explain giving lower JOLs to low-frequency words.

In addition, in Experiment 4, we found that two different methods of mediation analysis consistently showed that response times mediated the word frequency effect on JOLs. Although mediation was significant, the percentage of mediation was not very high, and it does not explain the effect entirely. This low mediation could reflect the fact that we collected the response times apart from the JOLs rating. Response times in a lexical decision task is a good indicator of fluency, but it does not have to be perfectly correlated with response times and fluency during JOLs. For this reason, our mediation analyses likely underestimated the actual mediation of fluency on the word frequency effect.

Our results contrast with those from Jia et al. (2016), which we speculate may be due to the different tasks used to measure fluency. Jia et al. used self-paced study time and manipulated the font style (easy or difficult to read), which may measure perceptual fluency. There may not be good reasons why high- and low-frequency words should differ in perceptual fluency because the perceptual characteristics of both types of words are similar (e.g., both are strings of letters that are pronounceable). However, high- and low- frequency words are more clearly different concerning their lexical or semantic fluency. Thus, a lexical decision task such as that used in Experiments 3 and 4, may have been more appropriate to the objective of measuring the type of fluency directly linked with word frequency. Language differences could additionally be at play, because Chinese has a logographic writing system (see Hoosain, 1992), and Portuguese has an alphabetic writing system. It is not clear how the writing system could have affected, but future research could try to shed light on this issue.

It is possible that participants who did not identify frequency may still base their JOLs on beliefs. However, those beliefs would not be related to how word frequency per se affects memory but instead would reflect the effects of other cues on memory, such as those detected after a subjective experience with the items which might even be correlated with word frequency. In other words, participants may use beliefs about processing fluency, since high-frequency words are processed quicker than low-frequency words. If this is the case, word frequency effects on JOLs could rely on both analytic theory-driven and subjective experience-driven processes, as shown with other metamemory cues (e.g., Blake & Castel, 2018; Frank & Kuhlman, 2017; Hu et al., 2015; Koriat et al., 2004; Price & Harrison, 2017; Rhodes, 2016; Undorf et al., 2017; Yang, et al., 2018). Future research should address whether participants have or develop beliefs about processing fluency or about cues that come from subjective experience with the items (e.g., familiarity) while making JOLs for high- and low-frequency words.

The present research contributes to the current discussion in the metamemory literature regarding the explanation of the effects of different cues on JOLs. It seems that in recent years there has been a tendency to attribute effects of different cues on JOLs to beliefs (Jia et al., 2016; Li et al., 2016; Mueller et al., 2014, 2016; Witherby & Tauber, 2017). For example, after considerable theoretical and empirical discussion between fluency and beliefs to explain the effect of font size on JOLs, the coin seems to have dropped by the side of beliefs (Luna et al., 2019; Undorf & Zimdahl, 2019). However, this does not mean that the effect of other cues should also be attributed exclusively to beliefs. Frequency is a good example. Past research has attributed its effects on JOLs exclusively to beliefs, but this research shows a more nuanced picture in which experience-based factors provide an alternative explanation.

In sum, our findings support an approach that considers experience-based factors on the process of making JOLs instead of just beliefs. Namely, our results show that beliefs about how a certain variable (i.e., word frequency) affects memory may not be necessary for it to affect JOLs. In addition, they also provide evidence that differences in processing fluency between high- and low-frequency words mediated the effect of frequency on JOLs. Word frequency might be just one of many cues affecting people's JOLs. In practical terms, understanding how this and other cues influence the process of learning monitoring could be relevant for educational contexts, where it is ideal that students correctly monitor their own learning.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235. https://doi.org/10.1177/1088868309341564

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. https://doi.org/10.1037/0096-1523.10.3.340

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/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. https://doi.org/10.1016/0749-596X(89)90016-8

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31*(2), 297–305. https://doi.org/10.3758/BF03194388

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. https://doi.org/10.1037/xlm0000246

Blake, A. B., & Castel, A. D. (2018). On belief and fluency in the construction of judgments of learning: Assessing and altering the direct effects of belief. *Acta Psychologica, 186*, 27–38. https://doi.org/10.1016/j.actpsy.2018.04.004

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Dunlosky, J., Mueller, M. L., & Tauber, S. K. (2015). The contribution of processing fluency (and beliefs) to people's judgments of learning. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas & H. L. Roediger III (Eds.), *Remembering: Attributions, Processes, and Control in Human Memory* (Part I, pp. 46–64). Psychology Press. https://doi.org/10.4324/9781315752808

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fife-Schaw, C. (2006). Principles of statistical inference tests. In G. M. Breakwell, S. Hammond, C. Fife-Schaw & J.A. Smith (Eds.), *Research methods in psychology* (pp. 388–413). Sage.

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(5), 680–693. https://doi.org/10.1037/xlm0000332

Gao, X. Y., Li, M. F., Chou, T. L, & Wu, J. T. (2016). Comparing the frequency effect between the lexical decision and naming tasks in Chinese. *Journal of Visualized Experiments*, *110*. https://doi.org/10.3791/53815

Hoosain, R. (1992). Psychological reality of the word in Chinese. *Advances in psychology, 90*, 111–130. https://doi.org/10.1016/S0166-4115(08)61889-0

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2017). The influences of valence and arousal on judgments of learning and on recall. *Memory & Cognition, 45*(1), 121–136. https://doi.org/10.3758/S13421-016-0646-3

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. https://doi.org/10.1371/journal.pone.0142351

Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2016). The effect of word frequency on judgements of learning: Contributions of beliefs and processing fluency. *Frontiers in Psychology, 6*:1995. https://doi.org/10.3389/fpsyg.2015.01995

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. https://doi.org/10.1006/jmla.1996.0009

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. https://doi.org/10.1037/0096-3445.126.4.349

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 Psychology: General, 133*(4), 643–656. https://doi.org/10.1037/0096-3445.133.4.643

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models (R package version 2.0–33). [Computer Software]. Retrieved from http://CRAN.Rproject.org/package/lmerTest

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*, 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

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

Lohnas, L. J., & Kahana, M. J. (2013). Parametric effects of word frequency in memory for mixed frequency lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1943–1946. https://doi.org/10.1037/a0033669

Luna, K., Nogueira, M., & Albuquerque, P. B. (2019). Words in larger font are perceived as more important: explaining the belief that font size affects memory. *Memory, 27*(4), 555–560. https://doi.org/10.1080/09658211.2018.1529797

Metcalfe, J., & Dunlosky, J. (2008). Metamemory. In H.L. Roediger, III (Ed.), *Learning and Memory: A comprehensive reference* (pp. 349–362). Elsevier.

Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people's metamemory judgments: What are contributions of processing fluency and beliefs about memory? *Quarterly Journal of Experimental Psychology, 69*(4), 781–799. https://doi.org/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. https://doi.org/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. https://doi.org/10.3758/s13423-012-0343-6

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. https://doi.org/10.1016/j.jml.2016.10.008

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 26, pp. 125–141). Academic Press.

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. https://doi.org/10.1016/j.jml.2016.12.003

R Core Team. (2019). *R: A language and environment for statistical computing [Computer Software]*. R Foundation for Statistical Computing.

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 65–80). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.4

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. https://doi.org/10.1037/a0013684

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review, 16*(3), 550–554. https://doi.org/10.3758/PBR.16.3.550

Schilling, H. E. H., 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. https://doi.org/10.3758/BF03201199

Soares, A. P., Iriarte, A., Almeida, J. J., Simões, A., Costa, A., França, P., Machado, J., & Comesaña, M. (2014). Procura-PALavras (P-PAL): Uma nova medida de frequência lexical do Português Europeu contemporâneo. *Psicologia: Reflexão e Crítica, 27*(1), 110–123. https://doi.org/10.1590/s0102-79722014000100013

Soares, A. P., Costa, A., Machado, J., Comesaña, M., & Oliveira, H. M. (2017). The Minho Word Pool: Norms for imageability, concreteness, and subjective frequency for 3,800 Portuguese words. *Behavior Research Methods, 49*(3), 1065–1081. https://doi.org/10.3758/s13428-016-0767-4

Soderstrom, N. C., Yue, C. L., & Bjork, E. L. (2016). Metamemory and education. In J. Dunlosky & S. Tauber (Eds.), *The Oxford Handbook of Metamemory*. https://doi.org/10.1093/oxfordhb/9780199336746.013.6

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review, 22*(4), 1014–1019. https://doi.org/10.3758/s13423-014-0768-1

Susser, J. A., Jin, A., & Mulligan, N. W. (2016). Identify 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. https://doi.org/10.1037/xlm0000189

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014) mediation: R Package for causal mediation analysis. *Journal of Statistical Software, 59*(5), 1–38. https://doi.org/10.18637/jss.v059.i05

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition, 43*(4), 647–658. https://doi.org/10.3758/s13421-014-0479-x

Undorf, M., & Ackermann, R. (2017). The puzzle of study time allocation for the most challenging items. *Psychonomic Bulletin & Review, 24*(6), 2003–2011. https://doi.org/10.3758/s13423-017-1261-4

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. https://doi.org/10.1037/xlm0000571

Undorf, M., Söllner, A. & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory & Cognition, 46*(4), 507–519. https://doi.org/10.3758/s13421-017-0780-6

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. https://doi.org/10.1016/j.jml.2016.07.003

Vuorre, M. (2017). bmlm: Bayesian multilevel mediation. R package version 1.3.4. Retrieved from https://cran.r-project.org/package=bmlm

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. https://doi.org/10.3758/s13421-016-0681-0

Yang, C., Huang, T. S.-T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language, 99*, 99–110. https://doi.org/10.1016/j.jml.2017.11.005

8. Study 2

Experience Matters: Effects of (In)Congruent Prompts About Word Frequency on Judgments of Learning

This empirical work was published in the journal Zeitschrift für Psychologie. Note that it was published under the 6th Edition of the APA style but here it was updated to the 7th Edition. The Zeitschrift für Psychologie was founded in 1890 by Hermann Ebbinghaus and Arthur König, being the oldest psychology journal in Europe and the second oldest in the world. It publishes research from all branches of empirical psychology, doing so in four topical issues per year. It is an international publication from Hogrefe Publishing and publishes in English since 2007. This article was published in the topical issue New Directions in Metamemory Research. According to Scopus, as of 2019 it had a CiteScore of 3.7, SJR of 0.731 and a SNIP of 0.841. Regarding quartiles, according to SJR website it was ranked in Q2 for the topic Psychology (miscellaneous) and Q1 for Arts and Humanities (miscellaneous). According to Clarivate InCites JCR, in 2019 this journal was ranked in Q2 for Psychology, Multidisciplinary, and its Impact Factor was of 1.528.

Reference:

Mendes, P. S., Luna, K., & Albuquerque, P. B. (2020). Experience matters: Effects of (in)congruent prompts about word frequency on judgments of learning. *Zeitschrift für Psychologie*, *228*(4), 254–263. https://doi.org/10.1027/2151-2604/a000423

In addition, results from this empirical work were partially or fully presented at several scientific events. In 2018, it was presented as a conference poster at the 13° Encontro Nacional da Associação Portuguesa de Psicologia Experimental (Braga, Portugal) and at the International Meeting of the Psychonomic Society (Amsterdam, The Netherlands) where it was bestowed an Accommodation Award, and as a conference presentation at the Summer School of the Swiss Graduate School for Cognition, Learning, and Memory, organized by the University of Bern (Weggis, Switzerland). In 2019, it was presented as a conference presentation at the 14° Encontro Nacional da Associação Portuguesa de Psicologia Experimental (Évora, Portugal). In 2020, this work would have been presented as part of a symposium on metamemory at the 62nd Conference of Experimental Psychologists (TeaP, Jena, Germany). Because of COVID-19 this event was canceled, but its abstract book was still published. Additionally, results from this study were also discussed at the CDSS workshop from the School of Social Sciences at the University of Mannheim, while a guest doctoral student there. The instructions and materials used in this study, as well as its Supplemental Materials can be found in Appendix C.

8.1. Abstract

The present study tested if word frequency effects on judgments of learning (JOLs) are exclusively due to beliefs or if the direct experience with the items also plays a role. Across four experiments, participants read prompts about the frequency of the words (high/low), which could be congruent/incongruent with the words' actual frequency. They made pre-study JOLs (except Experiment 1b), immediate JOLs, and completed a recall test. If experience drives the effect, JOLs should be based on actual word frequency rather than the prompts. Results showed higher pre-study JOLs for prompts of high frequency, but higher immediate JOLs for high-frequency words regardless of the prompt, suggesting an effect of direct experience with the words. In Experiments 2 and 3, we manipulated participants' beliefs, finding a small effect of beliefs on JOLs. We conclude that, regarding word frequency, direct experience with the items seems more relevant than beliefs when making immediate JOLs.

Keywords: judgments of learning, metamemory, word frequency, beliefs, direct experience

8.2. Introduction

When we reflect on how good or bad our memory is on a specific task (e.g., remembering a word list), we are making a metamemory judgment. Metamemory is the ability to know and understand how learning and memory operate and it can be assessed through judgments of learning (JOLs). JOLs are estimates of the future performance on a memory test (Rhodes, 2016) and they are made when someone is asked the probability of remembering a specific item, either before study (pre-study JOLs), immediately after study (immediate JOLs), or after some time (delayed JOLs). If JOLs are wrong or inaccurate, people's learning strategies can be misguided and impair the retention of the information to be learned (Soderstrom et al., 2016). Thus, it is relevant to understand how JOLs are made and which cues guide these judgments.

According to research, people make JOLs by attending to different cues (Koriat, 1997), and the processes by which one uses these cues have been organized around two categories. On the one hand, metacognitive judgments could be based on experience-driven processes (e.g., processing fluency, Besken, 2016; Undorf & Erdfelder, 2015). On the other hand, JOLs could be based on theory-driven processes that arise from an analytic examination of cues that may affect memory (e.g., beliefs, Koriat et al., 2004; Mueller et al., 2014; Mueller & Dunlosky, 2017; Undorf & Zimdahl, 2019). However, JOLs can be based on both types of processes since there is evidence for a joint contribution of fluency and beliefs on metamemory judgments (e.g., Blake & Castel, 2018; Frank & Kuhlman, 2017; Hu et al., 2015; Koriat et al., 2004; Price & Harrison, 2017; Undorf et al., 2017; Yang et al., 2018).

Additionally, research has identified cases in which theoretical explanations shifted from one factor to another. For example, the effect of font size on JOLs was primarily explained by feelings of fluency when processing the items (e.g., experience-driven factors, Rhodes & Castel, 2008). Subsequent studies, however, found that two different measures of fluency, self-paced study time and lexical decision time, did not differ between words presented in large or in small font size (Mueller et al., 2014). Instead, people seem to believe that words in larger font sizes are more memorable or more important (Luna et al., 2019; Mueller et al., 2014; Undorf & Zimdahl, 2019), which could potentially explain why they tend to give higher JOLs to words presented in larger font sizes.

Another factor known to affect JOLs is word frequency, that is, the number of times a word occurs in a certain language. Word frequency is usually measured as frequency per million (fpm, Brysbaert et al., 2018). Studies manipulating word frequency have shown that high-frequency words are usually rated with higher JOLs when compared to low-frequency words (e.g., Begg et al., 1989;

Benjamin, 2003; Hourihan et al., 2017; Jia et al., 2016; Mendes et al., 2019; for an exception, see Susser & Mulligan, 2015). Similar to the font size effect, the effect of word frequency on JOLs was first attributed to processing fluency (Begg et al., 1989), but more recently it has been argued that it is better explained by beliefs (Benjamin, 2003; Jia et al., 2016). While this shift from fluency to beliefs has been extensively explored regarding the font size effect on JOLs, the same is not true for word frequency (Fiacconi & Dollois, 2020). Thus, it remains unclear if the word frequency effect on JOLs is based on experience, on theory, or on both types of processes.

Despite the consistent findings that JOLs are higher for high-frequency words, researchers have proposed different explanations as to why that is the case. Responses to high-frequency words in lexical decision and naming tasks are faster (Mendes et al., 2019; Schilling et al., 1998), suggesting that they are more fluently processed than low-frequency words. In addition, Begg et al. (1989) found that participants rated high-frequency words as easier to imagine, easier to understand, easier to pronounce, and easier to study than low-frequency words. These studies hint that direct experience with the item and feelings of fluency when processing high-frequency words could explain why they are rated with higher JOLs.

Nonetheless, Jia et al. (2016) measured the effect of direct experience and fluency through study time and by using easy and difficult font styles and found no differences in fluency between high- and low-frequency words. Despite that, JOLs and correct recall were still higher for high-frequency words. These results challenge the explanation of the effect of word frequency on JOLs in terms of fluency. Moreover, Jia et al. (2016) assessed the role of beliefs in two experiments in which participants were not exposed to the words while making JOLs and, hence, there could be no effects of direct experience with the items. In one experiment, participants read a description of a hypothetical study about word frequency and requested participants' expectations for JOLs and recall. In another experiment, participants were presented with a prompt about the frequency of the following word (e.g., "You will see a high-frequency word") and made pre-study JOLs before seeing the actual items. In both experiments, Jia et al. (2016) found higher JOLs for high-frequency words and concluded that beliefs about how word frequency affects memory play a substantial role in the effect. However, it is possible that the effect of word frequency on JOLs might not be exclusively based on beliefs. In support of this idea, Mendes et al. (2019) recently reported higher fluency for high-frequency than low-frequency words, suggesting that fluency may also play a role in the effect of word frequency on JOLs.

Data against a belief-only explanation of the word frequency effect on JOLs requires more evidence to clarify the contribution of experience-driven and theory-driven processes. The present study aimed to understand the factors and processes underlying the effect of word frequency on JOLs. To that extent, we requested a combination of both pre-study JOLs and immediate JOLs followed by a recall test. Comparing both JOLs provides invaluable information about the weight of direct experience versus beliefs. For example, Price and Harrison (2017) explored the effect of fluency and beliefs on the font-size effect and requested either pre-study JOLs, immediate JOLs, or a combination of both. In the combination condition, pre-study JOLs were higher for words in large font, supporting the relevance of beliefs, but immediate JOLs were higher in magnitude and accuracy than pre-study JOLs, showing that direct experience with the items also affects metamemory ratings. With a similar procedure, Undorf and Bröder (2020) concluded that immediate JOLs are made using a strategic cue integration based on both beliefs and information gained from the experience with the items. In the present research, we also requested pre-study and immediate JOLs to investigate the role of direct experience and beliefs in the word frequency effect on JOLs.

It is reasonable to assume that pre-study JOLs are based on beliefs since there is no direct experience with the items. However, it is still unclear if immediate JOLs are predominantly based on beliefs or direct experience. In this research, we explored this issue by introducing incongruent information in prompts of word frequency. In four experiments, participants were presented with prompts that could be either congruent or incongruent with the actual word frequency of the studied word. Presenting incongruent prompts may lead participants to a state of uncertainty in which to make immediate JOLs they have to rely on information given by the prompt (based on beliefs), or on information from the studied item (based on experience). If participants rely only on beliefs, we would expect immediate JOLs to be higher for words preceded by high-frequency prompts regardless of whether the prompt is congruent or incongruent with the word. However, if immediate JOLs are based mostly on the actual frequency of the word through direct experience with the items, JOLs would be higher for high-frequency words regardless of the previous prompt.

To address these issues, in Experiment 1a we presented prompts of word frequency (congruent or incongruent with the actual word frequency) and collected pre-study JOLs. Then, participants studied high- and low-frequency words and made immediate JOLs. In Experiment 1b, we presented the (in)congruent information after participants had studied the word and during the immediate JOLs. In Experiment 2, we introduced the counter-belief that low-frequency words are usually better remembered. We expected this manipulation to affect their judgments if based on beliefs, but not if based on direct experience. Finally, in Experiment 3, to compare the effect of directing participants' beliefs in different directions, two groups of participants were instructed that either high-frequency words or low-frequency words are usually better remembered. In all the experiments, after the study-JOLs phase, participants completed a filler task followed by a free recall test. We expected a higher proportion of correct recall for high-frequency words than low-frequency words regardless of the prompt and instructions. Congruity between prompt and actual frequency might help memory (see congruity effects, Bein et al., 2015; Schulman, 1974), but we did not expect this because the simple prompts used here may not be strong enough to affect memory performance.

8.3. Experiment 1a

This study aimed to replicate the effects of word frequency on JOLs and recall (e.g., Jia et al. 2016) and explore the effect of (in)congruent prompts about word frequency on pre-study and immediate JOLs. We expected higher pre-study JOLs for prompts of high frequency than for prompts of low frequency. As per immediate JOLs, if the word frequency effect is mainly based on beliefs, then immediate JOLs should be higher for words preceded by prompts of high frequency. However, if immediate JOLs are mainly based on the direct experience with the items, then we expected higher JOLs for high-frequency words regardless of the prompt. We expected higher correct recall for high-frequency words than for low-frequency words, regardless of the prompt.

8.3.1. Method

8.3.1.1. Participants. The sample size for all experiments was estimated using G*Power 3 (Faul et al., 2007). For Experiments 1a and 1b it was based on previous research that found a strong word frequency effect on JOLs (Cohen's $d = 1.85$, Mendes et al., 2019). Computations showed that 23 participants were needed to obtain a statistical power of .95 and to detect a large effect size ($f = .40$). We aimed for a sample size of approximately 30 participants. All participants included in the analysis were native speakers of European Portuguese and all data were collected in Portugal. In this experiment we excluded one participant that was not European Portuguese native speaker. A final sample of 33 university students (1 male) with a mean age of 19.45 years-old ($SD = 1.62$) took part in this study in exchange for course credits.

8.3.1.2. Design. We presented two types of prompts indicating that participants would see a high-frequency or a low-frequency word. However, for clarity in the exposition and to avoid confusion with the actual frequency of the words, we converted the two levels of prompt to whether they were congruent or incongruent with the actual frequency of the word that followed the prompt. A prompt indicating that the upcoming word would be a high-frequency word was congruent when participants would indeed see a high-frequency word, it was incongruent when they would indeed see a low-frequency word, and vice versa for prompts about low-frequency words. A 2 (prompt about word frequency: congruent, incongruent) x 2 (word frequency: high, low) within-subjects design was used.

8.3.1.3. Materials. Forty Portuguese nouns (20 high-frequency, 20 low-frequency) were extracted from the Minho Word Pool (Soares et al., 2017), and absolute frequencies were obtained through Procura-PALavras (P-PAL, Soares et al., 2014). Frequency of high-frequency words, e.g., *tempo* (time) was higher ($M = 302.90$, $SD = 147.21$, range [160, 758]), than frequency of low-frequency words, e.g., *arenque* (clupea); $M = .36$, $SD = .25$, range [0.009, 0.81]), $t(19) = 9.21$, $p < .001$, Cohen's $d = 2.06$. Both sets of words did not differ on concreteness ($M_{HF} = 3.87$, $M_{LF} = 3.79$; 1-7 scale), number of letters ($M_{HF} = 6.40$, $M_{LF} = 6.40$), nor number of syllables ($M_{HF} = 2.55$, $M_{LF} = 2.60$), all $p > .511$.

8.3.1.4. Procedure. Participants completed the experiment on a computer in an individual soundproof booth. After giving informed consent, participants read instructions indicating that they would study a list with high- and low-frequency words and that they should remember them for a later memory test. Instructions also stated they should rate the probability of remembering each word later, both before and after reading the word, on a scale from 0 (sure that they will not remember the word) to 100 (sure that they will remember the word). After starting the experiment, participants saw a prompt (either congruent or incongruent) about the frequency of the word ("You will see a high/low-frequency word") and made a self-paced pre-study JOL ("How likely are you to remember the word later?") on a 6-point Likert scale (0-20-40-60-80-100). After a 500ms fixation cross, the word appeared at the center of the screen for 5s. The order of presentation of words and congruency of the prompt was pseudo-randomized (i.e., randomized once, but the same for all participants). After each word, on a new screen, participants were asked an immediate JOL ("How likely are you to remember later the word you just studied?" on the same scale). After studying all 44 words (two primacy and two recency buffers were included but not analyzed), participants completed a filler task (arithmetic calculus) for three minutes and, finally, a free recall task for five minutes.

8.3.2. Results

Table 7 presents the main descriptive statistics for immediate JOLs and recall. For all experiments, additional analysis and data are presented in the Supplemental Materials (available at https://osf.io/8qvay) and in Appendix C.3.

Table 7

Mean (Standard Deviation) for Immediate JOLs and Correct Recall (proportion) per Condition across All Experiments.

		Immediate JOLs		**Correct recall**	
	Actual word frequency	Congruent prompt	Incongruent prompt	Congruent prompt	Incongruent prompt
Exp. 1a	HF	54.30 (17.48)	54.30 (18.54)	.33 (.17)	.26 (.15)
	LF	27.88 (12.80)	26.73 (10.31)	.21 (.13)	.15 (.12)
Exp. 1b	HF	58.27 (21.45)	54.07 (18.41)	.41 (.18)	.30 (.15)
	LF	25.20 (16.37)	27.47 (14.08)	.20 (.13)	.20 (.14)
Exp. 2	HF	55.03 (23.24)	57.52 (19.51)	.35 (.20)	.21 (.14)
	LF	41.94 (21.84)	39.39 (21.29)	.16 (.12)	.14 (.12)
Exp. 3 (better-for-high-instruction)	HF	58.90 (21.08)	59.60 (19.35)	.28 (.17)	.22 (.17)
	LF	31.80 (17.29)	36.30 (16.06)	.14 (.15)	.20 (.16)
Exp. 3 (better-for-low-instruction)	HF	54.40 (17.69)	56.30 (17.49)	.29 (.19)	.34 (.13)
	LF	47.60 (16.42)	43.00 (19.99)	.22 (.16)	.20 (.12)

Notes: HF – high-frequency; LF – low-frequency; Congruent prompts matched the actual word frequency of presented word; incongruent prompts were of the opposite frequency of the presented word.

8.3.2.1. Pre-study JOLs. A paired-samples *t*-test showed higher pre-study JOLs to prompts of high frequency (M = 53.55, SD = 12.76) than prompts of low frequency (M = 28.18, SD = 12.63), $t(32) = 8.09$, $p < .001$, Cohen's $d = 1.41$.

8.3.2.2. Immediate JOLs. A 2 (frequency) x 2 (prompt) ANOVA showed higher JOLs to high-frequency words (M = 54.30, SD = 17.88) than low-frequency words (M = 27.30, SD = 11.54), $F(1, 32) = 109.16$, $p < .001$, $\eta_p^2 = .77$. There was no significant effect of prompt, $F(1, 32) = 0.43$, $p = .519$, $\eta_p^2 = .01$, and no significant interaction, $F(1, 32) = 0.21$, $p = .652$, $\eta_p^2 < .01$.

8.3.2.3. Recall. A 2 (frequency) x 2 (prompt) ANOVA showed higher recall for high-frequency words ($M = .29$, $SD = .16$) than low-frequency words ($M = .18$, $SD = .13$), $F(1, 32) = 31.78$, $p < .001$, $\eta_p^2 = .50$, and for words preceded by a congruent prompt ($M = .27$, $SD = .16$) than preceded by an incongruent prompt ($M = .20$, $SD = .15$), $F(1, 32) = 12.68$, $p = .001$, $\eta_p^2 = .28$. The interaction was not significant, $F(1, 32) = 0.14$, $p = .713$, $\eta_p^2 < .01$.

8.3.3. Discussion

In this experiment, participants gave higher pre-study JOLs to prompts of high frequency, which could only be based on beliefs, but our results also showed that for immediate JOLs, they relied on actual word frequency. These results suggest that word frequency can affect JOLs through experience-driven factors. Participants recalled high-frequency words better than low-frequency words and we also found the unexpected result that our simple prompt was enough as to affect recall, consistent with the congruity effect.

8.4. Experiment 1b

In Experiment 1a, immediate JOLs were based on the actual frequency of the word and not on the prompt. One explanation for participants not using the prompt as a cue for JOLs is that presenting the prompt before the word made it harder for participants to integrate the two cues in their immediate JOLs. Experiment 1b addressed this issue by presenting the prompt after the studied word and during the immediate JOL. If participants' immediate JOLs are based mostly on beliefs, they should be higher for words followed by a prompt of high frequency. However, if participants rely more on item characteristics, their immediate JOLs should be higher for high-frequency words regardless of the prompt.

8.4.1. Method

8.4.1.1. Participants. Thirty university students (6 male) with a mean age of 23 years-old ($SD = 7.77$) took part in this study in exchange for course credits.

8.4.1.2. Design and Materials. Design and materials were the same as in Experiment 1a.

8.4.1.3. Procedure. The procedure was similar to that of Experiment 1a, with two differences: pre-study JOLs were not requested, and prompts about word frequency were presented *after* the words and when making the immediate JOL.

8.4.2. Results

Main descriptive statistics are presented in Table 7.

8.4.2.1. Immediate JOLs. A 2 (frequency) x 2 (prompt) ANOVA showed higher JOLs for high-frequency words (M = 56.17, SD = 19.93) than for low-frequency words (M = 26.33, SD = 15.18), $F(1, 29) = 75.45$, $p < .001$, $\eta_p^2 = .72$, no effect of prompt, $F(1, 29) = 0.67$, $p = .420$, $\eta_p^2 = .02$, and no significant interaction, $F(1, 29) = 3.61$, $p = .067$, $\eta_p^2 = .11$.

8.4.2.2. Recall. A 2 (frequency) x 2 (prompt) ANOVA showed higher recall for high-frequency words (M = .36, SD = .17) than low-frequency words (M = .20, SD = .13), $F(1, 29) = 21.31$, $p < .001$, $\eta_p^2 = .42$, higher recall for words followed by a congruent prompt (M = .31, SD = .18) than for words followed by an incongruent prompt (M = .25, SD = .15), $F(1, 29) = 5.38$, $p = .028$, $\eta_p^2 = .16$. Also, the interaction was significant, $F(1, 29) = 9.53$, $p = .004$, $\eta_p^2 = .25$. Pairwise comparisons showed that the effect of prompt was only significant for high-frequency words, $t(29) = 3.40$, $p = .002$, Cohen's d = 0.62.

8.4.3. Discussion

In Experiment 1b, participants gave higher immediate JOLs for high-frequency words regardless of the prompt, replicating results from Experiment 1a, even when incongruent information was presented after the studied word. This suggests that participants relied mostly on actual word frequency.[7]

Overall, memory performance was better for high-frequency words. A congruity effect on recall was also found, but only for high-frequency words. This finding may be due to the recall of low-frequency words being already very low.

8.5. Experiment 2

Results from Experiments 1a and 1b suggest that experience-driven processes contribute to the word frequency effect on immediate JOLs. However, participants may also rely on the belief that high-frequency words are better remembered. To test this idea, in Experiment 2 we introduced a counter-belief suggesting that low-frequency words are better remembered and requested pre-study and

7 Participants could learn that the prompts were unreliable. Data from pre-study JOLs across the study phase suggest that this is not the case (cf. Supplemental Materials, available in Appendix C.3).

immediate JOLs.[8] Changing participants' beliefs should affect their pre-study JOLs, which should be higher for prompts of low frequency. Additionally, if immediate JOLs are mostly based on beliefs, they should also be higher for words preceded by prompts of low frequency. However, if immediate JOLs rely mostly on information obtained through direct experience, then their judgments should still be higher for high-frequency words despite the counter-belief.

8.5.1. Method

8.5.1.1. Participants. The sample size for Experiment 2 was based on that of Experiment 1a. Thirty-three university students (5 male) with a mean age of 20.12 years-old (*SD* = 3.35) took part in this study in exchange for course credits.

8.5.1.2. Design and Materials. Design and materials were the same as in Experiment 1a.

8.5.1.3. Procedure. The procedure was the same as in Experiment 1a. The only difference was that the following information was presented before the instructions (translated from the original Portuguese):

"There are more than 200 000 words in the Portuguese language. By analyzing written texts and oral speech entries, it is possible to verify that some words appear very frequently, while others only appear sometimes. High-frequency words are those that appear many times (more than 100 times per million words), for example, 'house'. Low-frequency words are those which rarely appear (less than 1 time per million words), for example, 'reef'.

Curiously, our memory is very good for low-frequency words because, since they are so rare, they capture our attention and become very salient among others. This could be a reason why, when we learn an uncommon word, we start seeing it everywhere, because it has become more salient".

8.5.2. Results

Table 7 presents the main descriptive statistics for immediate JOLs and recall.

8.5.2.1. Pre-study JOLs. A paired-samples *t*-test showed that participants gave similar pre-study JOLs to prompts of high frequency (M = 54.30, SD = 19.44) and to prompts of low frequency (M = 53.85, SD = 17.19), $t(32) = 0.14$, $p = .892$, Cohen's $d = 0.02$.[9]

---

8 By presenting a counter-belief, Blake and Castel (2018) successfully manipulated participants' beliefs about the font size effect on memory, with an impact on JOLs.
9 Individual data showed that most participants gave higher pre-study JOLs to prompts of low frequency, suggesting the effectiveness of the belief manipulation (cf. Supplemental Materials).

8.5.2.2. Immediate JOLs. A 2 (frequency) x 2 (prompt) ANOVA showed higher JOLs for high-frequency words (*M* = 56.27, *SD* = 21.32) than low-frequency words (*M* = 40.67, *SD* = 11.54), $F(1, 32) = 13.42$, $p < .001$, $\eta_p^2 = .30$, no effect of prompt, $F(1, 32) = 0.001$, $p = .973$, $\eta_p^2 < .001$, and a marginal interaction, $F(1, 32) = 4.15$, $p = .050$, $\eta_p^2 = .12$. The interaction showed that presenting a low-frequency prompt increased immediate JOLs for both high-frequency words (incongruent prompt) and for low-frequency words (congruent prompt). None of these differences were statistically significant ($p = .155$ and $p = .064$, respectively), but their opposite sign gave rise to the interaction. The instruction that low-frequency words are better recalled may have increased immediate JOLs for words preceded by the prompt of low frequency, supporting the effect of beliefs on immediate JOLs. Since the interaction was marginal, we sought further empirical confirmation in Experiment 3 before discussing it.

8.5.2.3. Recall. A 2 (frequency) x 2 (prompt) ANOVA showed higher recall for high frequency words (*M* = .28, *SD* = 19) than low-frequency words (*M* = .15, *SD* = .12), $F(1, 32) = 36.07$, $p < .001$, $\eta_p^2 = .53$, and higher recall for words preceded by a congruent prompt (*M* = .25, *SD* = 19) than by an incongruent prompt (*M* = .17, *SD* = .13), $F(1, 32) = 12.62$, $p = .001$, $\eta_p^2 = .28$,. A significant interaction, $F(1, 32) = 8.63$, $p = .006$, $\eta_p^2 = .21$, showed that the recall difference between prompts was significant for high-frequency words, $t(32) = 3.94$, $p < .001$, Cohen's $d = 0.69$, but not for low-frequency words, $t(32) = 0.88$, $p = .386$, Cohen's $d = 0.15$.

8.5.3. Discussion

In Experiment 2, participants were led to believe that low-frequency words are better remembered than high-frequency words. This information would contradict the idea that high-frequency words are better remembered (Jia et al., 2016). This manipulation affected participants' judgments, eliminating word frequency effects on pre-study JOLs (i.e., equal for prompts of high and low frequency). Participants' immediate JOLs were still higher for high-frequency words, suggesting that they based these judgments on the actual frequency and direct experience with the items. This also suggests the existence of an a priori belief that high-frequency words are better remembered, which is somehow updated (but not completely altered) with the counter-belief. Regarding memory performance, Experiment 2 replicated the results from Experiments 1a and 1b.

8.6. Experiment 3

In the previous experiments, immediate JOLs were affected by actual frequency but not by the prompts. Further, in Experiment 2, we added a counter-belief that eliminated the effects of word frequency on pre-study JOLs and showed a small impact on immediate JOLs. These findings suggest that the word frequency effect on JOLs may be explained by both beliefs and the direct experience with the item, with a more prominent role of the latter factor. Experiment 3 aimed to extend findings from Experiment 2 that suggested that manipulating people's beliefs through an instruction can alter JOLs, but that, still, immediate JOLs are mostly based on direct experience with the items. Here, we instructed participants that either high-frequency words (better-for-high) or low-frequency words (better-for-low, replication of Experiment 2) are better remembered. In the better-for-high condition, we expected that the word frequency effect on JOLs would be larger, while in the better-for-low condition we expected a pattern of results similar to that in Experiment 2.

8.6.1. Method

8.6.1.1. Participants. In Experiment 3, the sample size was computed based on the effect size of word frequency effects on immediate JOLs in Experiment 2, in which a counter-belief was introduced (η_p^2 = .30). Ten participants per group were needed to obtain a statistical power of .95 and to detect a large effect size (*f* = .65). We aimed for a sample size of approximately 20 participants per group. Two participants were excluded for not being native speakers of European Portuguese. A final sample of 40 university students (7 male) with a mean age of 23.11 years-old (*SD* = 5.12) took part in this study in exchange for course credits.

8.6.1.2. Design and Materials. A 2 (prompt about word frequency: congruent, incongruent) x 2 (word frequency: high, low) x 2 (instruction: better-for-high, better-for-low) mixed design was used. Prompt and frequency were manipulated within-subjects, and instruction was manipulated between-subjects. Materials were the same as in Experiment 1a.

8.6.1.3. Procedure. The procedure was the same as in Experiment 2. Before instructions, participants read two paragraphs about which words are better remembered. For the low-frequency condition, the two paragraphs were identical to those in Experiment 2. For the high-frequency condition, the first paragraph was the same, but the second was replaced by this information "Curiously, our memory is very good for high-frequency words because, since they are so common, we find them more in our daily lives. This could be a reason why, when we learn a common word, we start seeing it everywhere because it has become familiar".

8.6.2. Results

Table 7 presents the main descriptive statistics for immediate JOLs and recall.

8.6.2.1. Pre-study JOLs. A 2 (prompt) x 2 (instruction) ANOVA showed higher pre-study JOLs for prompts of high frequency (M = 54.75, SD = 18.66) than for prompts of low frequency (M = 47.27, SD = 12.89), $F(1, 38) = 4.15$, $p = .049$, $\eta_p^2 = .10$, no effect of instructions, $F(1, 38) = 0.18$, $p = .677$, $\eta_p^2 = .005$, and a significant interaction, $F(1, 38) = 9.79$, $p = .003$, $\eta_p^2 = .21$. Participants in the better-for-low condition gave similar pre-study JOLs for prompts of high (M = 49.85, SD = 17.89) and low frequency (M = 53.85, SD = 12.89), as in Experiment 2, $t(19) = 0.72$, $p = .478$, Cohen's d = 0.16, while participants in the better-for-high condition gave higher pre-study JOLs to prompts of high frequency (M = 59.65, SD = 18.54) than for prompts of low frequency (M = 40.70, SD = 18.56), $t(19) = 3.93$, $p < .001$, Cohen's d = 0.88.

8.6.2.2. Immediate JOLs. A 2 (frequency) x 2 (prompt) x 2 (instruction) ANOVA showed higher JOLs to high-frequency (M = 57.30, SD = 18.39) than low-frequency words (M = 39.68, SD = 18.39), $F(1, 38) = 18.61$, $p < .001$, $\eta_p^2 = .33$, no effect of prompt, $F(1, 38) = 0.30$, $p = .587$, $\eta_p^2 = .008$, and no effect of instruction, $F(1, 38) = 0.97$, $p = .330$, $\eta_p^2 = 0.03$. There was a significant interaction between frequency, prompt, and instruction, $F(1, 38) = 4.68$, $p = .037$, $\eta_p^2 = .11$.

To analyze this interaction, we conducted two ANOVAs frequency x prompt for each instruction. With the better-for-high instruction, the only significant difference was that high-frequency words were rated with higher immediate JOLs than low-frequency words, $F(1, 19) = 19.40$, $p < .001$, $\eta_p^2 = .51$. However, with the better-for-low instruction the only significant result was the interaction between frequency and prompt, $F(1, 19) = 6.46$, $p = .020$, $\eta_p^2 = .25$. This interaction replicated that from Experiment 2: the low-frequency prompt increased immediate JOLs for both high- (incongruent prompt) and low-frequency words (congruent prompt). The differences were not significant ($p = .232$ and $p = .070$, respectively), but as in Experiment 2, their opposite sign caused the interaction. This result suggests a small but detectable effect of beliefs.[10] When the belief is consistent with previous knowledge, immediate JOLs are not affected. However, when it is inconsistent (i.e., a counter-belief), it increases JOLs for the low-frequency words, and the word frequency effect disappears.

8.6.2.3. Recall. A 2 (frequency) x 2 (prompt) x 2 (instruction) mixed ANOVA showed higher recall for high-frequency words (M = .28, SD = 17) than for low-frequency words (M = .19, SD = .15), $F(1, 38) = 13.49$, $p < .001$, $\eta_p^2 = .26$, no effect of prompt, $F(1, 38) = 0.12$, $p = .734$, $\eta_p^2 = .003$, nor

10 We explored the effect of beliefs on immediate JOLs by conducting linear mixed models using pre-study JOLs as a measure of beliefs. A complete report of that analysis is presented in the Supplemental Materials.

instruction, $F(1, 38) = 2.97$, $p = .093$, $\eta_p^2 = .07$. There was an interaction between frequency, prompt, and instruction, $F(1, 38) = 6.29$, $p = .017$, $\eta_p^2 = .14$.

To analyze this interaction, we conducted two ANOVAs frequency x prompt for each instruction. With the better-for-high instruction, high-frequency words were better recalled than low-frequency words, $F(1, 19) = 5.05$, $p = .037$, $\eta_p^2 = .21$. This result was qualified by the significant interaction between prompt and frequency, $F(1, 19) = 5.10$, $p = .036$, $\eta_p^2 = .21$, showing better memory for high- than low-frequency words for words preceded by congruent prompts, $t(19) = 3.18$, $p = .005$, Cohen's $d = 0.71$, but not for words preceded by incongruent prompts, $t(19) = 0.46$, $p = .649$, Cohen's $d = 0.10$. With the better-for-low instruction, high-frequency words were better recalled than low-frequency words, $F(1, 19) = 8.74$, $p = .008$, $\eta_p^2 = .32$.

8.3.3. Discussion

In Experiment 3, participants were led to believe that either high-frequency words (better-for-high) or low-frequency words (better-for-low) are better remembered. Regarding pre-study and immediate JOLs, results were consistent with previous experiments. Participants with the better-for-high instructions showed a similar pattern of results as in Experiment 1a, suggesting that when the instruction tried to strengthen the belief that high-frequency words are better remembered, JOLs do not seem to be affected. However, participants with the better-for-low instructions produced similar results as in Experiment 2. When the instructions presented a belief inconsistent with previous knowledge (i.e., a counter-belief), it increased JOLs for the low-frequency words, and the word frequency effect disappears. In Experiment 3, this latter finding was true not only for pre-study JOLs but also for immediate JOLs.

When considering memory performance, better recall for high-frequency words was observed for all participants regardless of instruction, replicating the word-frequency effect on recall. However, the congruity effect for high-frequency words disappeared. It is difficult to explain why, particularly because Experiment 2 and the better-for-low condition in Experiment 3 were direct replicas and showed a reversed pattern. Thus, the question of how word frequency, prompts, beliefs, and counter-beliefs interact with memory will remain open. Future research focused on their effects on memory should shed light on this issue.

8.7. General Discussion

The present research aimed to examine the mechanisms underlying the word frequency effect on JOLs. For that purpose, we manipulated two cues, word frequency and the congruity of the prompt paired with each word, and measured participants' pre-study JOLs, immediate JOLs, and recall. This section will mainly focus on the effects of word frequency on both types of JOLs, and results in memory performance will only be briefly discussed. The main conclusion is that direct experience with the items seems to play an essential role in the word frequency effect on JOLs with a minor role of beliefs.

Consistent with previous findings (Jia et al., 2016), in Experiment 1a, we found higher pre-study JOLs for prompts of high frequency, suggesting a strong influence of beliefs. However, immediate JOLs were higher for high-frequency words regardless of the prompt presented before, suggesting that participants relied on the actual frequency of words after direct experience with the items. Similarly, when we paired the prompt with the immediate JOLs in Experiment 1b to ensure that the cue was accessible to participants, they still rated high-frequency words with higher JOLs regardless of the prompt. While participants may believe that high-frequency words are better remembered, as suggested by pre-study JOLs, their immediate JOLs seem to be mostly based on experience-driven processes.

Our findings provide evidence for a joint, although unequal, contribution of theory-driven and experience-driven processes in the word frequency effect on JOLs. In Experiments 2 and 3, we introduced the counter-belief that low-frequency words are better remembered to further test if JOLs would shift when beliefs were manipulated. The counter-belief was effective because pre-study JOLs for prompts of both high and low frequency were similar, unlike in Experiment 1a. Interestingly, the counter-belief had a small effect and increased immediate JOLs for words preceded by a prompt of low frequency. This suggests that the information in the counter-belief was used as a cue for immediate JOL, but also that its effect was rather limited by comparison with the strong and consistent influence of actual word frequency. However, participants may integrate their own knowledge and beliefs in combination with the presented instruction in their JOLs.

Results from Experiments 2 and 3 are in line with findings showing that a counter-belief can decrease the typical effect of a cue on JOLs, but that a belief consistent with previous knowledge does not impact JOLs (Blake & Castel, 2018). Here, we found that countering participants' beliefs about word frequency altered their JOLs, but also that information in the same direction of the belief did not. It is unclear why strengthening a pre-existing belief does not affect JOLs, but Blake and Castel (2018) found similar results, even when presenting three intensity levels of the same-direction belief. It may be that whereas the challenge posed by the counter-belief leads to a re-evaluation of the belief, same-direction beliefs, no matter how intense, do not.

Finally, regarding memory performance, we consistently replicated the word frequency effect on correct recall, with a better recall for high-frequency words across all experiments. In addition, we observed a congruity effect favoring memory for words preceded/followed by congruent prompts. This latter effect was observed for high-frequency words, but not always for low-frequency words. It is possible that because recall is already low for low-frequency words, congruity of the prompt may not improve memory for these words.

In sum, our findings support the idea that word frequency effects on JOLs are not exclusively based on beliefs but hold an important contribution of experience-driven processes. This contrasts with other variables that affect JOLs mainly through beliefs such as the font size (Luna et al., 2019; Undorf & Zimdahl, 2019) and strongly suggests that different cues are differently influenced by beliefs and direct experience. Future research should try to identify why some cues are more affected by one factor or the other.

References

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease 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. https://doi.org/10.1016/0749-596X(89)90016-8

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31*(2), 297–305. https://doi.org/10.3758/BF03194388

Bein, O., Livneh, N., Reggev, N., Gilead, M., Goshen-Gottstein, Y., & Maril, A. (2015). Delineating the effect of semantic congruency on episodic memory: The role of integration and relatedness. *PloS One, 10(2)*, e0115624. https://doi.org/10.1371/journal.pone.0115624

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. https://doi.org/10.1037/xlm0000246

Blake, A. B., & Castel, A. D. (2018). On belief and fluency in the construction of judgments of learning: Assessing and altering the direct effects of belief. *Acta Psychologica, 186*, 27–38. https://doi.org/10.1016/j.actpsy.2018.04.004

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. https://doi.org/10.3758/BF03193146

Fiacconi, C. M., & Dollois, M. A. (2020). Does word frequency influence judgments of learning (JOLs)? A meta-analytic review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale.* Advance online publication. https://doi.org/10.1037/cep0000206

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(5), 680–693. https://doi.org/10.1037/xlm0000332

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2017). The influences of valence and arousal on judgments of learning and on recall. *Memory & Cognition, 45*(1), 121–136. https://doi.org/10.3758/S13421-016-0646-3

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. https://doi.org/10.1371/journal.pone.0142351

Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2016). The effect of word frequency on judgments of learning: Contributions of beliefs and processing fluency. *Frontiers in Psychology, 6*:1995. https://doi.org/10.3389/fpsyg.2015.01995

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. https://doi.org/10.1037/0096-3445.126.4.349

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 Psychology: General, 133*(4), 643–656. https://doi.org/10.1037/0096-3445.133.4.643

Luna, K., Nogueira, M., & Albuquerque, P. B. (2019). Words in larger font are perceived as more important: Explaining the belief that font size affects memory. *Memory, 27*(4), 555–560. https://doi.org/10.1080/09658211.2018.1529797

Mendes, P. S., Luna, K., & Albuquerque, P. B. (2019). Word frequency effects on judgments of learning: More than just beliefs. *The Journal of General Psychology*. Advance online publication. https://doi.org/10.1080/00221309.2019.1706073

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. https://doi.org/10.1016/j.jml.2013.09.007

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. https://doi.org/10.1016/j.jml.2016.10.008

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. https://doi.org/10.1016/j.jml.2016.12.003

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 65–80). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.4

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. https://doi.org/10.1037/a0013684

Schilling, H. E. H., 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. http://doi.org/10.3758/BF03201199

Schulman, A. I. (1974). Memory for words recently classified. *Memory & Cognition, 2*(1), 47–52. https://doi.org/10.3758/BF03197491

Soares, A. P., Iriarte, A., Almeida, J. J., Simões, A., Costa, A., França, P., Machado, J., & Comesaña, M. (2014). Procura-PALavras (P-PAL): Uma nova medida de frequência lexical do Português Europeu contemporâneo. *Psicologia: Reflexão e Crítica, 27*(1), 110–123. https://doi.org/10.1590/s0102-79722014000100013

Soares, A. P., Costa, A., Machado, J., Comesaña, M., & Oliveira, H. M. (2017). The Minho Word Pool: Norms for imageability, concreteness, and subjective frequency for 3,800 Portuguese words. *Behavior Research Methods, 49*(3), 1065–1081. https://doi.org/10.3758/s13428-016-0767-4

Soderstrom, N. C., Yue, C. L., & Bjork, E. L. (2016). Metamemory and education. In J. Dunlosky & S. Tauber (Eds.), *The Oxford handbook of metamemory*. https://doi.org/10.1093/oxfordhb/9780199336746.013.6

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review, 22*(4), 1014–1019. https://doi.org/10.3758/s13423-014-0768-1

Undorf, M., &, Bröder, A. (2020). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology, 73*(4), 629–642. https://doi.org/10.1177/1747021819882308

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition, 43*(4), 647–658. https://doi.org/10.3758/s13421-014-0479-x

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. https://doi.org/10.1037/xlm0000571

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. https://doi.org/10.1016/j.jml.2016.07.003

Yang, C., Huang, T. S.-T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language, 99*, 99–110. https://doi.org/10.1016/j.jml.2017.11.005

9. Study 3

On the Pervasive Effect of Word Frequency in Metamemory

This study was conducted at the University of Mannheim while a guest doctoral student under the supervision of Dr. Monika Undorf. It was recently submitted to an international journal in the field of experimental psychology.

Reference:

Mendes, P. S., & Undorf, M. (2020). *On the pervasive effect of word frequency in Metamemory*. [Manuscript submitted for publication].

Additionally, an abstract of this work was accepted to be presented as a conference presentation at the International APPE-SEPEX Meeting 2020. Unfortunately, due to the COVID-19 pandemic this event was canceled. The instructions and materials used in this study, can be found in Appendix D.

9.1. Abstract

Predictions of one's future memory performance – judgments of learning (JOLs) – are based on the cues learners regard as diagnostic of memory performance. One of these cues is word frequency or how often words are experienced in the language. It is not clear, however, whether word frequency would affect JOLs when other cues are also available. The current study aims to close this gap by testing whether objective and subjective word frequency affect JOLs in the presence of font size as an additional cue. Across three experiments, participants studied words that varied in word frequency (Experiment 1: high and low objective frequency; Experiment 2: a whole continuum from high to low objective frequency; Experiment 3: high and low subjective and objective frequency) and were presented in a large (48pt) or a small (18pt) font size, made JOLs, and completed a free recall test. Results showed that people based their JOLs on both word frequency and font size. We conclude that word frequency is an important cue that affects metamemory even in multiple-cue situations.

Keywords: judgments of learning, metamemory, word frequency, font size, cue integration

9.2. Introduction

Our behavior is often guided by our thoughts about our memory and its functioning. For example, if we think that we are forgetful of people's names, we may become more attentive when someone is introduced to us. This ability to think about our memory and the way it works and to adjust our behavior accordingly is known as metamemory (Rhodes, 2016). One popular way to assess metamemory is through judgments of learning (JOLs), which are predictions of later performance in a memory test (e.g., Nelson & Narens, 1990; Rhodes, 2016). In a typical JOL experiment, people indicate how likely they will later remember a specific stimulus. Because people cannot directly read out their memories, these judgments are based on cues and heuristics (Koriat, 1997; Rhodes, 2016).

According to Koriat's cue-utilization approach (1997), cues that inform JOLs are organized into three categories: intrinsic, extrinsic, and mnemonic. Cues such as word frequency or concreteness, which are characteristics of the study material, are referred to as intrinsic, whereas cues such as item repetition or study duration, which are specific to the study conditions, are referred to as extrinsic. Mnemonic cues are subjective internal factors that indicate to what extent an item has been learned, such as the ease with which information is processed (i.e., processing fluency). Thus, JOLs are assumed to rely on theory-driven processes such as beliefs or knowledge about one's memory that entail analytic inferences about how intrinsic or extrinsic cues affect memory (Koriat, 1997). In addition, JOLs are also based on experience-driven processes that draw on people's nonanalytic experiences at study such as ease of processing or fluency (Koriat, 1997).

Much research on JOLs has focused on identifying the cues that influence these judgments to better understand the basis of metamemory. Among the cues that have been shown to affect people's JOLs are word frequency (e.g., Begg et al., 1989), font size (e.g., Rhodes & Castel, 2008), concreteness (e.g., Witherby & Tauber, 2017), relatedness of word pairs (e.g., Mueller et al., 2013), and motoric fluency (e.g., Susser & Mulligan, 2015). One key finding is that some of these cues are more diagnostic of memory performance than others. For example, memory is better for concrete words than for abstract words, and people give higher JOLs to concrete than to abstract words (e.g., Witherby & Tauber, 2017), showing that they are metacognitively aware of this effect. In contrast, font size has at best a small effect on memory, with words studied in large font sizes being remembered slightly better or similar to words studied in small font sizes. Participants, however, attribute much higher JOLs to words studied in large font sizes. This metamemory illusion became known as the font size effect (e.g., Hu et al., 2015; Luna et al., 2018; Mueller et al., 2014; Rhodes & Castel, 2008; Undorf & Zimdahl, 2019).

Over the years, research has examined a number of cues that affect JOLs, but most studies have manipulated only one single cue at a time (Rhodes, 2016; Undorf et al., 2018). In contrast, in real world situations, people typically have multiple cues at their disposal to make metacognitive judgments. Hence, it would be essential to simultaneously manipulate more than one cue in order to understand memory monitoring outside of the lab (Rhodes, 2016). This is particularly important because there is evidence for cue interactions, demonstrating that cue effects may differ between one-cue situations and multiple-cue situations. For instance, Hourihan et al. (2017) found that both arousal (Experiment 1) and valence (Experiment 2) affected participants' JOLs when manipulated in isolation. Importantly, however, when arousal, valence, and word frequency were simultaneously manipulated, only word frequency but not valence and arousal affected JOLs (Experiment 3).

It is not yet clear under what circumstances cues that affect JOLs when manipulated in isolation also affect JOLs when manipulated in combination with additional cues. In order to shed light on this issue, the present study examined if word frequency affects JOLs when manipulated in combination with another cue. As will be detailed below, word frequency has a pervasive effect on memory performance and has recently received some attention in JOL studies (Fiacconi & Dollois, 2020; Jia et al., 2016; Mendes et al., 2019, 2020).

9.2.1. Word Frequency Effects on Memory and Metamemory

Word frequency corresponds to the number of times a word occurs in a specific language (e.g., Brysbaert et al., 2018). Objective word frequency refers to how often a word occurs in the corpus of a certain lexical database or to how frequently it occurs in a million words from the database (frequency per million or fpm, Brysbaert et al., 2018). Subjective word frequency refers to a word's subjective familiarity and is obtained through estimates of how many times people encounter a word either in its written or spoken form in their daily lives (Balota et al., 2001). Objective and subjective frequency are highly correlated. There is evidence that good measures of objective frequency are sufficient to explain word frequency effects in word recognition (Brysbaert & Cortese, 2011). In this domain, subjective frequency is mostly useful when exploring individual differences in word-person interactions. However, frequency measures that are closer to daily usage of words and thereby more akin to subjective frequency such as frequency measures based on subtitles from films and television shows and subtitles explain more variance than frequency measures based on written corpora (Brysbaert et al., 2011). Also, subjective frequency explains additional variance in response times from word identification tasks and naming tasks (Brysbaert et al., 2018; see also Adelman et al., 2006).

Interestingly, the effects of objective and subjective word frequency on memory performance depend on the type of memory test. Less familiar low-frequency words such as *strait* are usually more easily recognized than high-frequency words such as *time* (Benjamin, 2003; Tulving & Kroll, 1995). In contrast, recall performance is better for high-frequency than for low-frequency words or presents a U-shaped pattern favoring both extremes of word frequency over middle ranges of frequency (Deese, 1959; Lohnas & Kahana, 2013; Peters, 1936; for a review and new theory on the effects of word frequency on memory, see Popov & Reder, 2020). It has been argued that high-frequency words, which are more familiar, produce stronger memory traces and require less working memory resources, which could explain the effects of frequency on memory (Popov & Reder, 2020).

Regardless of the memory test, however, participants give higher JOLs to high-frequency words (Begg et al., 1989; Benjamin, 2003; Jia et al., 2016; Mendes et al., 2019; for a review see Fiacconi & Dollois, 2020). This was found both when word frequency was manipulated in two levels (e.g., Jia et al., 2016) and when it was manipulated on a continuum (e.g., Hourihan et al., 2017). Studies that examined the basis of word frequency effects on JOLs demonstrated that high-frequency words are processed more fluently than low-frequency words and that these differences in processing fluency contribute to the word frequency effect on JOLs (Mendes et al., 2019, 2020). At the same time, people's belief that common words are better remembered was also found to contribute to the effect of word frequency on JOLs (Jia et al., 2016).

Most studies on word frequency effects in memory and metamemory have used objective measures of frequency. In particular, no studies have investigated the role of subjective word frequency in metamemory (see Fiacconi & Dollois, 2020; for memory studies see Deese, 1959; Loewenthal & Gibbs, 1974; Peters, 1936; Reder et al., 2016). This is unfortunate, because the focus on objective word frequency ignores potential differences in subjective word frequency among individuals. In particular, frequent encounters with specific words of low objective frequency increase their subjective frequency and consequently produce mismatches between objective and subjective word frequency. For instance, *cognition* – a word with low objective word frequency (6.66 fpm) according to norms from the English Lexicon Project (Balota et al., 2007) – presumably has a low subjective word frequency for the average statistician but a high subjective word frequency for the average cognitive psychologist. Consistent with this idea, Monaghan et al. (2017) showed that the advantage in lexical and semantic access for words of high objective frequency over words of low objective frequency decreased as a function of word exposure. This is probably due to words of low objective frequency becoming more subjectively frequent. It is plausible to assume that subjective word frequency also affects metamemory. Specifically, due to high familiarity, words with high subjective frequency may be processed more fluently than words with low subjective frequency (Oppenheimer, 2006). Moreover, people may base their JOLs on the explicit belief that words they are familiar with are easier for them to remember (see, e.g., Dunlosky et al., 2015; Mueller et al., 2016; Mueller et al., 2013; for the impact of idiosyncratic strategies on JOLs see Bröder & Undorf, 2019). To our knowledge, only a single study addressed the effects of familiarity on JOLs. Shanks and Serra (2014) found that participants assigned higher JOLs to facts from topics they were more familiar with than to facts from topics they were less familiar with. It remains to be seen, however, whether subjective word frequency affects JOLs. The third experiment of the present study examined this question.

9.2.2. The Basis of JOLs in Multiple-cue Situations

Researchers have recently begun to address the effects of multiple cues on JOLs when simultaneously manipulated (Tatz & Peynircioğlu, 2020; Undorf et al., 2018). In one experiment (Undorf et al., 2018, Experiment 1), participants studied words in smaller and larger font sizes once or twice and made JOLs after each study presentation. Results revealed that both font size and frequency of presentation affected JOLs, with higher JOLs for words presented in a larger than in a smaller font size and for words studied twice than studied once. Also, individual-level analyses showed that the majority of participants based their JOLs on both font size and frequency of presentation. Together, these findings were interpreted as indicative of cue integration in JOLs. In three subsequent experiments, Undorf et al. (2018) found similar results when manipulating multiple cues in two discrete levels (concreteness and emotionality in Experiment 2; number of presentations, font size, concreteness and emotionality in Experiment 3) and when manipulating multiple cues on a continuum (font size, concreteness, emotionality in Experiment 4). Tatz and Peynircioglu (2020) also reported evidence for cue integration in JOLs. Across several experiments, they found that manipulations of font size and font blur simultaneously affected JOLs. Moreover, additionally manipulating the size or blur of backgrounds reduced the effects of font size and font blur on JOLs, suggesting that participants integrated item-specific cues and contextual cues in JOLs. Collectively, these findings show that people have the amazing ability to integrate multiple cues in JOLs.

However, it cannot be taken for granted that people integrate multiple cues in their JOLs. There is evidence to suggest that cue effects on JOLs are attenuated or eliminated when other cues are simultaneously manipulated. In particular, three experiments revealed mixed findings regarding the effects of word frequency on JOLs in multiple cue situations. Susser and Mulligan (2015, Exp. 2) and Susser et al. (2017) obtained JOLs for words of high and low frequency that participants wrote down with their dominant or non-dominant hand. Results showed that writing hand but not word frequency affected JOLs. In contrast, Hourihan et al. (2017, Exp. 3) observed that word frequency was the only cue to affect JOLs when simultaneously manipulating word frequency, emotional valence, and arousal.

The present study focused on the effects of objective and subjective word frequency on JOLs in multiple cue situations. In each of the three experiments, we simultaneously manipulated objective and/or subjective word frequency and font size. While it was an open question whether objective and subjective word frequency would affect JOLs, prior studies suggested that font size would affect JOLs. Rhodes and Castel (2008) found that font size affected JOLs even when font format (aLtErNaTiNg vs. standard) was manipulated. Later studies showed that font size affected JOLs when manipulated simultaneously with font style (Price et al., 2016), number of presentations (Kornell et al., 2011; Undorf et al., 2018, Exp. 1), relatedness of word pairs (Price & Harrison, 2017), and with number of presentations, concreteness, and emotionality combined (Undorf et al., 2018).

If people integrate word frequency and font size in their JOLs, then JOLs should be higher for high-frequency than for low-frequency words, and higher for words presented in a large than in a small font size. Importantly, this pattern should emerge not only at the aggregate level but also at the individual level to exclude the possibility that each participant's JOLs were based on only one cue but different individuals based their JOLs on different cues (see Undorf et al., 2018). If, in contrast, people base their JOLs only on font size, then JOLs should be similar for high-frequency and for low-frequency words but higher for words presented in a large than in a small font size.

9.2.3. JOL Accuracy in Multiple-cue Situations

In addition to assessing the effects of different cues on JOLs, metamemory research explored the accuracy of people's JOLs. Apart from absolute accuracy (or calibration), that is, the overall correspondence between JOLs and memory performance (Rhodes, 2016), research has focused on relative accuracy (or resolution). Relative accuracy refers to the extent to which JOLs discriminate between items that will be remembered and those which will not be remembered (Rhodes, 2016). Resolution is often measured by Goodman and Kruskal's gamma coefficient (Goodman & Kruskal, 1954). Gamma ranges from -1 to +1. A negative gamma indicates low accuracy since higher JOLs are attributed to non-remembered items, while a positive gamma indicates high accuracy since higher JOLs are attributed to remembered items. A gamma correlation of around zero means an absence of a relationship between JOLs and memory performance.

Gamma was argued to be better than other measures of resolution because it is widely applicable and makes no scaling assumptions beyond the ordinal level (Nelson, 1984). However, it has also been criticized. For example, Benjamin and Diaz (2008) advised against the use of gamma when interactions, between-group comparisons, or across-scale comparisons are used, and instead proposed a signal-detection measure for resolution. Similarly, Masson and Rotello (2009) found that gamma varies with response bias and also suggested alternative measures. Despite this criticism and the availability of alternative resolution measures, gamma is still widely used in JOL research (see Bröder & Undorf, 2019). One critical aspect of most resolution measures is that they conflate random noise and metacognitive ability (Bröder & Undorf, 2019). This implies that limited resolution of JOLs (see, e.g., Rhodes & Tauber, 2011) is not necessarily due to metacognitive misconceptions or missing knowledge but could result from unreliability in JOLs and measures of memory performance. Moreover, none of these measures can assess the contributions of individual cues to JOLs and memory performance in multiple-cue situations.

Recently, Bröder and Undorf (2019) have proposed an alternative measure of JOL resolution that could overcome these issues: the matching index (G) from Brunswik's (1952) lens model. Originally proposed as a framework to describe how people make judgments in the perceptual environment, Brunswik's (1952) lens model can be applied to any judgmental task in which people predict a distal variable in the environment on the basis of probabilistic cues (for a review see Karelaia & Hogarth, 2008). Typically, in perceptual studies, people make a judgment about a distal variable in the environment that is not directly available, such as the size of a distant object. To infer the distal variable and make a judgment, people therefore rely on probabilistic cues, such as linear perspective or retinal size (see Hammond & Stewart, 2001). When applied to metamemory, the environment to be predicted is memory performance and JOLs rely on the cues available at study (cf. Bröder & Undorf, 2019). In the lens model, two regression analyses determine the optimal regression weights for each cue in order to predict the environment and the weights given to the cues by the judge, respectively. The Pearson correlation between environment and judgments is referred to as achievement (r_a). In JOL studies, this correlation is a surface measure of JOL resolution. Additionally, the lens model offers a competence measure that is corrected for unreliability in memory performance and judgment processes. This is referred as matching index (*G*) and reflects how closely a person's cue weights correspond to the effects of cues on their memory performance.

Using Brunswik's (1952) lens model allows decomposing the Pearson correlation between JOLs and recall into the following components: the linear predictability of memory performance on the basis of cues (R_{REC} or validities), the consistency with which the person weighs the cues when making JOLs (R_{JOL} or cue utilization), the matching index (*G*) that is indicative of JOL resolution, and a nonlinear component (*C*) that indicates systematic covariation between judgments and recall that cannot be attributed to a linear combination of the measured cues. Bröder and Undorf (2019) applied the lens model to five experiments that manipulated multiple cues and concluded that the lens model's matching index *G* was a better measure of JOL resolution than the gamma coefficient or the Pearson correlation between JOLs and recall.

In the present research, we used Brunswik's (1952) lens model to examine the contribution of font size and word frequency to JOLs and to measure JOL accuracy through the matching *G* index. At the same time, we compared the matching *G* index with surface resolution measures that do not correct for unreliability in memory performance and judgment processes (i.e., r_a and Gamma).

9.2.4. The Current Study

In order to understand if word frequency affects JOLs when an additional cue is present, the current study aimed to test whether objective and subjective word frequency affect JOLs when font size is simultaneously manipulated. In Experiment 1, we manipulated objective word frequency and font size in two discrete categories. In Experiment 2, we combined a continuous manipulation of objective word frequency with a discrete manipulation of font size. Finally, in Experiment 3, we manipulated objective word frequency, subjective word frequency, and font size in two discrete categorie. In all experiments, participants completed a study phase with immediate JOLs and a recall test. We used Brunswik's (1952) lens model to examine the contributions of font size and word frequency to JOLs and JOL accuracy.

9.3. Experiment 1

Experiment 1 aimed to explore if objective word frequency would affect JOLs when an additional cue (i.e., font size) is present. If so, JOLs should be higher for words with a high than a low objective word frequency and for words printed in a large than in a small font size. In contrast, we expected better memory performance for high-frequency than for low-frequency words, but similar memory performance for words printed in a large and a small font size.

9.3.1. Method

9.3.1.1. Participants. Based on the large effects of word frequency and font size found in Mendes et al. (2020) and Undorf et al. (2018), respectively, we aimed for $N = 30$ to obtain a statistical power of $(1 - \beta) = .99$ to detect large effects ($f = .40$) of both manipulations in repeated measures ANOVAs with $\alpha = .05$ (all power analyses conducted via G*Power 3; Faul et al., 2007). Thirty University of Mannheim students (6 male) with an average age of 22.43 years ($SD = 3.30$) participated in this experiment in exchange for course credits or money.

9.3.1.2. Design and Materials. We used a 2 (word frequency: high, low) x 2 (font size: 48 pt, 18 pt) within-subjects design. JOLs and correct recall were measured.

For all experiments presented here, words were taken from the Berlin affective word list reloaded (BAWL-R, Võ et al., 2009), with absolute frequency per million taken from the dlex-DB online tool (type_normalized, Heister et al., 2011). A list of 40 German nouns (20 high-frequency, 20 low-frequency) was created (see Appendix D). Frequency of high-frequency words (e.g., *night*, *Nacht* in the original German) was higher ($M = 162.15$, $SD = 80.01$, range [100.85, 428.15]), than frequency of low-frequency words (e.g., *garland*, *Girlande* in German; $M = .48$, $SD = .27$, range [0.01, 0.92]), $t(19) = 9.06$, $p < .001$, $d = 2.03$. High- and low-frequency words did not differ in concreteness ($M_{HF} = 4.04$, $M_{LF} = 4.01$, $t(19) = 0.08$, $p = .94$, $d = 0.02$), valence ($M_{HF} = 0.36$, $M_{LF} = -0.25$, $t(19) = 1.93$, $p = .07$, $d = 0.43$), arousal ($M_{HF} = 2.62$, $M_{LF} = 2.63$, $t(19) = 0.06$, $p = .96$, $d = 0.01$), number of letters ($M_{HF} = 6.25$, $M_{LF} = 6.45$, $t(19) = 0.44$, $p = .66$, $d = 0.10$), or number of syllables ($M_{HF} = 2.05$, $M_{LF} = 2.25$, $t(19) = 1.17$, $p = .26$, $d = 0.26$). Additionally, four words (2 high- and 2 low-frequency) served as primacy and recency buffers and were not analyzed.

9.3.1.3. Procedure. Participants completed the experiment individually on a computer. After giving informed consent, participants were instructed to study a list of words for a later free recall test. They were also informed that they would have to predict the probability of remembering each word ("Chance to recall (0% -100%)?"). On each study trial, a fixation cross was presented in the middle of the screen for 500 ms. Immediately afterwards, a word was presented either in a large 48-pt font or in a small 18-pt font for 2 seconds. Participants then made a self-paced JOL on a new screen by clicking on an 11-point scale ranging from 0 to 100 (0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100). The same primacy and recency buffers were presented to all participants at the beginning and at the end of the study list. Items were presented in one of six fixed orders for each participant. Item orders were random with the restriction that no more than three items from each combination of word frequency and font size appeared consecutively. Half of the low-frequency words and high-frequency words were presented in 18-pt font, and the other half were presented in 48-pt font. Assignment of words to font size was counterbalanced across participants. After the study phase, participants completed a 3-min filler task (arithmetic calculus) followed by a 4-min free recall task. Participants were asked to type in as many words from the study phase as they could remember. Overall, the experiment took about 20 minutes.

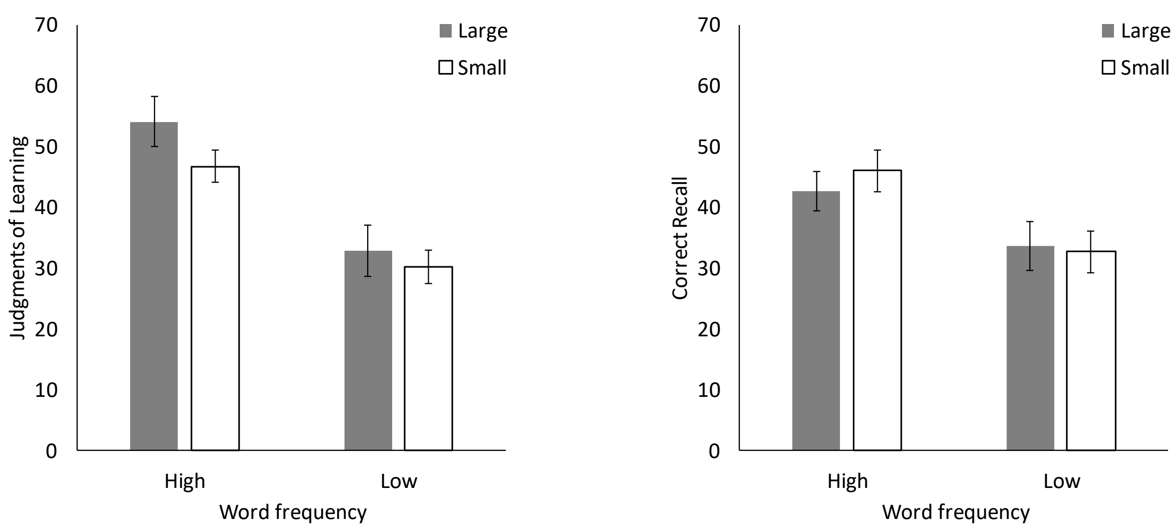
9.3.2. Results

For all analyses, the alpha level was set to .05. All effect sizes are reported in terms of Cohen's *d* for *t*-tests and η_p^2 for ANOVAs.

9.3.2.1. Judgments of Learning. JOLs are presented in Figure 4. A 2 (frequency) x 2 (font size) ANOVA revealed that JOLs were higher for high-frequency words than for low-frequency words, $F(1, 29) = 30.84$, $p < .001$, $\eta_p^2 = .52$, and higher for words presented in the large font size than in the small font size, $F(1, 29) = 9.73$, $p = .004$, $\eta_p^2 = .25$. The interaction was not significant, $F(1, 29) = 3.58$, $p = .07$, $\eta_p^2 = .11$. These results show that, at the aggregate level, word frequency and font size were integrated in JOLs.

Figure 4

Mean Judgments of Learning and Correct Recall in Experiment 1



Note: Error bars represent one standard error of the mean.

9.3.2.2. Recall. Correct recall is presented in Figure 4. A 2 (frequency) x 2 (font size) ANOVA revealed better recall for high-frequency words than for low-frequency words, $F(1, 29) = 17.54$, $p < .001$, $\eta_p^2 = .38$, but no effect of font size, $F(1, 29) = .16$, $p = .69$, $\eta_p^2 = .006$, and no interaction, $F(1, 29) = 1.00$, $p = .33$, $\eta_p^2 = .03$.

9.3.2.3. JOL Accuracy. We analyzed JOLs accuracy by calculating the Gamma correlation between JOLs and recall and by performing a lens model analysis. We applied the lens model to our data by following the steps provided by Bröder and Undorf (2019) for SPSS. The model included the effect coded independent variables (+1 = high frequency, large font size; -1 = low frequency, small font size) and the control variables number of letters, number of syllables, valence, arousal, and concreteness were centered. Descriptive statistics are presented in Table 8. Gamma was significantly positive, $t(29) = 7.71$, $p < .001$, $d = 1.41$, indicating that participants' JOLs were accurate in predicting which items they would and would not recall. The lens model analysis revealed that the consistency of cue use in JOLs and the linear predictability of recall were limited, as shown by relatively low lens model parameters R_{JOL} and R_{REC}. Parameter C significantly differed from zero, $t(29) = 4.37$, $p < .001$, $d = .80$, indicating that the linear combination of the cues included in the lens model could not account for all systematic covariance between JOLs and recall (for similar results, see Bröder & Undorf, 2019). The matching index G significantly exceeded r_a, $t(29) = 3.44$, $p = .002$, $d = 0.63$, and was numerically but not significantly higher than Gamma, $t(29) = 1.05$, $p = .303$, $d = 0.19$. This suggests that the lens model's matching index G is a valid measure of JOL resolution.

9.3.2.4. Cue Integration at the Individual Level. To evaluate individual cue use, we examined standardized regression weights from the linear regression of JOLs on cues obtained in the lens model analysis. As in previous studies (Undorf et al.,2018, Exp. 4; Undorf & Bröder, 2020, Exp. 2), we conjectured that participants based their JOLs on a cue if the cue's regression weight was positive. 18 participants (60%) integrated word frequency and font size in their JOLs. The remaining participants based their JOLs either on word frequency (9 participants) or font size (3 participants).[11] When we defined cue use as regression weights that exceeded the convention for small effect sizes proposed by Cohen (1977, $|r| \geq .10$), 11 participants (36.67%) integrated both cues in their JOLs. The remaining participants based their JOLs on word frequency (13 participants), on font size (4 participants), or neither cue (2 participants).

11 A linear regression analysis with only word frequency and font size as predictors (i.e., without control variables) revealed similar results: 18 participants used both cues, 8 only word frequency, and 4 only font size. These results are also consistent with an analysis based on simple mean differences.

Table 8

Mean Values of Gamma and the Parameters From Brunswik's (1952) Lens Model in Experiments 1 - 3

	Gamma	r_a	G	R_{JOL}	R_{REC}	C
Experiment 1	.38	.26	.43	.62	.46	.17
Experiment 2	.36	.24	.37	.51	.40	.20
Experiment 3	.40	.27	.42	.63	.46	.20

Notes: Gamma = Goodman-Kruskal gamma correlation between JOLs and recall, r_a = Pearson correlation between JOLs and recall, G = matching index from the lens model, R_{JOL} = multiple correlation between cues and JOLs, R_{REC} = multiple correlation between cues and recall, C = nonlinear component from the lens model.

9.3.3. Discussion

Results revealed higher JOLs and better recall for high-frequency words than for low-frequency words, reflecting the usual word frequency effect in metamemory and recall (e.g., Mendes et al., 2019). Replicating the font size illusion in metamemory, participants also made higher JOLs for words presented in the large font than in the small font although font size did not affect recall performance (e.g., Rhodes & Castel, 2008). Analysis of JOL accuracy using gamma correlations and Brunswik's (1952) lens model showed that people accurately predicted what they would and would not remember, and suggested that the lens model's matching index G is a valid measure of resolution. Examining cue utilization at the individual level revealed that individual participants based their JOLs on both word frequency and font size. Thus, this experiment demonstrates that word frequency affected JOLs when the additional cue font size was present. An interesting observation was that the consistency with which participants used the available cues when making JOLs (R_{JOL}) was higher than in previous lens model studies with JOLs (Bröder & Undorf, 2019). This finding is consistent with the idea that word frequency has a pervasive effect on JOLs

9.4. Experiment 2

Consistent with previous findings regarding the pervasiveness of word frequency effects on JOL in the presence of additional cues (Hourihan et al., 2017), Experiment 1 showed that word frequency affected JOLs when font size was also manipulated. In Experiment 2, we aimed to extend this finding to a situation where word frequency varied on a continuum. Considering that a non-representative selection of stimuli is claimed to change the processes that underlie people's judgments (e.g., Dhami et al., 2004; Gigerenzer et al., 1991; see also Undorf & Bröder, 2020; Undorf et al., 2018), using study lists that include only words with very low or high word frequency might have exaggerated or even produced word frequency effects in metamemory in Experiment 1 and most prior studies (e.g., Begg et al., 1989; Benjamin, 2003; Jia et al., 2016; Mendes et al., 2019; for a notable exception, see Hourihan et al., 2017). Thus, it was important to examine whether word frequency affected JOLs over and above font size when manipulated on a continuum. We expected JOLs to increase monotonically with increasing word frequency (Hourihan et al., 2017). In contrast, we expected higher recall performance for both low-frequency and high-frequency words as compared to mid-frequency words, that is, a U-shaped relationship between word frequency and recall (e.g., Lohnas & Kahana, 2013).

9.4.1. Method

9.4.1.1. Participants. Based on the effect sizes obtained in Experiment 1, we aimed for a sample size of $N = 30$ to obtain a statistical power of $(1 - \beta) = .99$ to detect large effects ($f = .40$) of both manipulations in repeated-measures ANOVAs. Thirty University of Mannheim students (8 male) with an average age of 23.33 years ($SD = 6.78$) participated in this experiment in exchange for course credit or money.

9.4.1.2. Design and Materials. An 8 (frequency bin, ranging from 8 = high to 1 = low) x 2 (font size: 48 pt, 18 pt) within-subjects design was used. JOLs and correct recall were measured. A list of 48 German nouns was created, with 6 words in each word-frequency bin. Word frequency differed significantly across bins ($p < .001$) but concreteness, valence, arousal, number of letters, and number of syllables were held constant across bins (all $p > .45$). Table 9 presents descriptive statistics for all word characteristics.

9.4.1.3. Procedure. The procedure was identical to Experiment 1. Within each frequency bin, the three words presented in 18 pt and the three words presented in 48 pt were similar in all word characteristics (see Table 9).

9.4.2. Results

9.4.2.1. Judgments of Learning. JOLs are presented in Figure 5. As one can see in the figure, JOLs increased with increasing word frequency. However, this increase was not strictly monotonic. An 8 (frequency bin) x 2 (font size) ANOVA revealed that word frequency affected JOLs, $F(7, 203) = 8.28$, $p < .001$, $\eta_p^2 = 0.22$, and that JOLs were higher for words presented in the large font size rather than in the small font size, $F(1, 29) = 28.27$, $p < .001$, $\eta_p^2 = 0.49$. There was no significant interaction, $F(7, 203) = 0.88$, $p = .59$, $\eta_p^2 = 0.03$. Post hoc t-tests revealed reliably higher JOLs in Bin 6 than in Bins 1 (*p* = .010), 2 (*p* < .001), 3 (*p* < .001), and 4 (*p* = .004), and reliably higher JOLs in Bin 8 than in Bins 2 (*p* = .015) and 3 (*p* = .022). Thus, JOLs were generally higher for bins with higher frequency than for bins with lower frequency.

Table 9

Means (and Standard Deviations) of Word Frequency, Concreteness, Valence, and Arousal for the Words in Each Frequency Bin Used in Experiment 2.

Frequency Bin (Range in fpm)	Frequency	Concreteness	Valence	Arousal
1 (0-1)	0.44 (0.37)	4.81 (0.84)	0.20 (0.86)	2.43 (0.19)
2 (1-5)	4.30 (0.58)	4.10 (1.19)	0.68 (0.57)	2.29 (0.84)
3 (7-15)	15.58 (2.77)	5.48 (1.11)	0.75 (0.35)	2.21 (0.50)
4 (20-40)	32.05 (6.98)	5.34 (1.06)	0.54 (0.64)	2.30 (0.23)
5 (40-70)	60.54 (11.46)	4.55 (1.72)	0.53 (0.89)	2.35 (0.38)
6 (75-100)	91.15 (5.93)	4.54 (1.39)	1.22 (0.82)	2.47 (0.61)
7 (100-120)	111.99 (5.27)	4.38 (1.73)	0.99 (1.04)	2.27 (0.51)
8 (125-400)	257.33 (105.81)	4.00 (1.21)	0.54 (0.78)	2.24 (0.31)

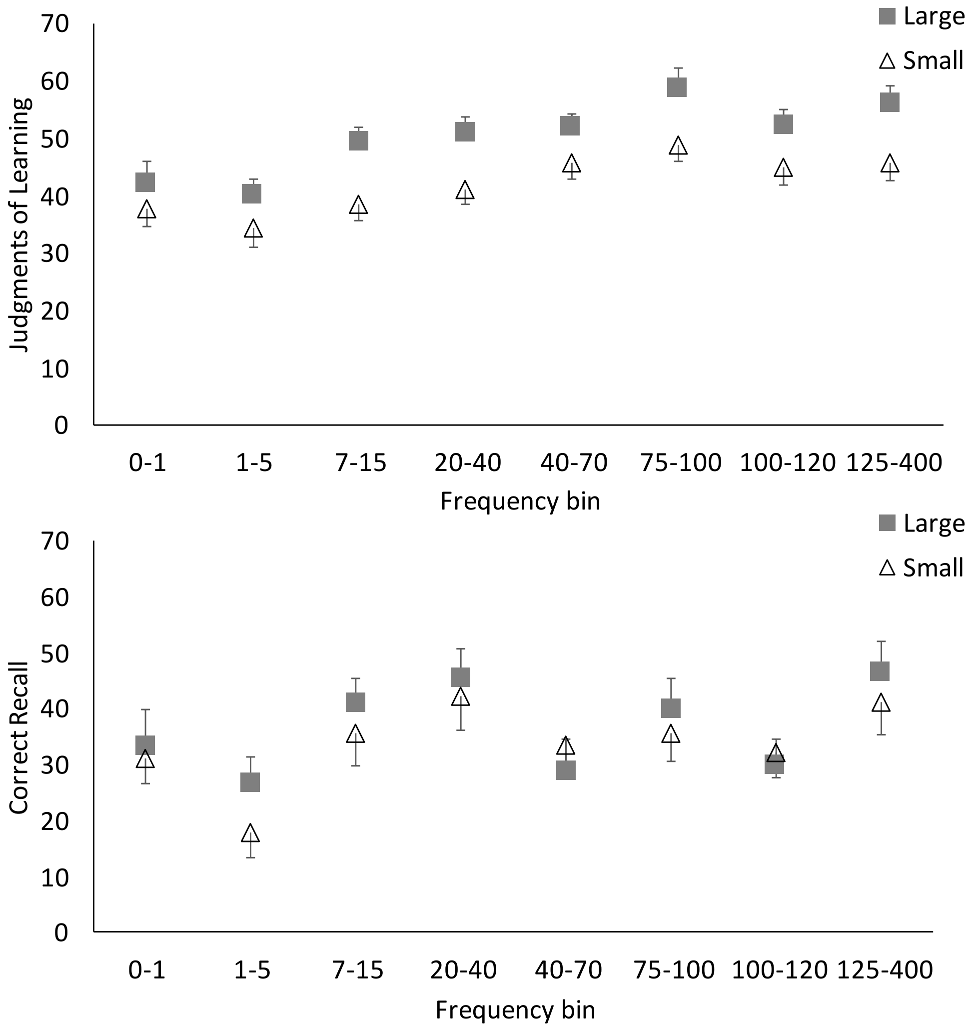
Notes: Frequency values (fpm = frequency per million) were taken from dlex-DB (Heister et al., 2011), and concreteness, valence, and arousal values were taken from BAWL-R (Võ et al., 2009).

9.4.2.2. Recall. Correct recall is presented in Figure 5. An 8 (frequency bin) x 2 (font size) ANOVA revealed that recall differed across frequency bins, $F(7, 203) = 5.35$, $p < .001$, $\eta_p^2 = .16$. There was no effect of font size on recall, $F(1, 29) = 2.03$ $p = .17$, $\eta_p^2 = 0.07$, and no interaction, $F(7, 203) = 0.47$, $p = .86$, $\eta_p^2 = 0.02$. Post hoc *t*-tests revealed reliably lower recall in Bin 2 than in Bins 3 (*p* = .034), 4 (*p* = .002), and 8 (*p* < .001). Thus, word frequency had no consistent effect on correct recall.

9.4.2.3. JOL Accuracy. As in Experiment 1, JOL accuracy was analyzed by the Gamma correlation between JOLs and recall and by Brunswik's (1952) lens model. Descriptive statistics are presented in Table 8. Gamma was significantly positive, $t(29) = 8.32$, $p < .001$, $d = 1.52$, indicating that participants' JOLs were accurate in discriminating which items they would and would not recall. The lens model included the centered independent variables frequency bin (+3.5 = Bin 8; +2.5 = Bin 7; +1.5 = Bin 6; +0.5 = Bin 5; -0.5 = Bin 4; -1.5 = Bin 3; -2.5 = Bin 2; -3.5 = Bin 1) and font size (+1 = large font size; -1 = small font size), and the same control variables as in Experiment 1 (number of letters, number of syllables, valence, arousal, and concreteness). The lens model parameters indicating consistency of cue use and predictability of memory performance from the cues, R_{JOL} and R_{REC}, were

Figure 5

Mean Judgments of Learning and Correct Recall in Experiment 2



Note: Error bars represent one standard error of the mean.

again quite low. Maybe, manipulating frequency on a continuum reduced the predictability of JOLs and recall by producing more fine-grained differences in word frequency. The nonlinear component *C* significantly differed from zero, $t(29) = 6.95$, $p < .001$, $d = 1.27$. The matching index *G* again significantly exceeded r_a, $t(29) = 3.03$, $p = .005$, $d = .55$, and was numerically but not significantly larger than Gamma, $t(29) = 0.22$, $p = .829$, $d = .04$. A numerically lower matching index *G* than in Experiment 1 suggests that JOLs were less accurate in this experiment. This is presumably due to the absence of an effect of word frequency on recall.

9.4.2.4. Cue Integration at the Individual Level. As in Experiment 1, we evaluated individual cue use by examining regression weights from the linear regression of JOLs on cues obtained in the lens model analysis. Twenty participants (66.67%) integrated word frequency and font size in their JOLs. The remaining participants based their JOLs either on word frequency (3 participants) or on font size (5 participants) or used neither cue (2 participants).[12] When defining cue use as regression weights that exceeded Cohen's (1977) convention for small effects ($|r| \geq .10$), 14 participants (46.67%) integrated both cues in their JOLs. The remaining participants based their JOLs on word frequency (4 participants) on font size (7 participants), or on neither cue (5 participants).

9.4.3. Discussion

In Experiment 2, JOLs increased with word frequency and thereby revealed a word frequency effect. At the same time, only some pairwise comparisons between frequency bins were reliable. This might suggest that manipulating word frequency on a continuum rather than discretely reduced the size of its effect on JOLs. It is also possible, however, that non-significant pairwise comparisons resulted from limitations in reliability because each frequency bin comprised only very few words. JOLs did not increase monotonically with word frequency: Participants assigned highest JOLs to words from Bin 6 (75 to 100 fpm) rather than Bin 8 (125 to 400 fpm). While we cannot exclude that the latter finding is related to numerically but not significantly elevated values of valence in Bin 6, a more plausible explanation is that mismatches between objective and subjective word frequency are at play. As mentioned in the introduction, the frequency of exposure to words (e.g., Monaghan et al., 2017) and the diversity of situations in which words are encountered (e.g., Brysbaert et al., 2018) may lead to differences in word frequency effects at the individual level. It is possible that JOLs are guided by subjective word frequency.

Unlike JOLs, recall performance was unrelated to word frequency. The finding that recall performance was better in Bins 3, 4, and 8 than in Bin 2 was inconsistent with the U-shaped relationship between word frequency and recall performance reported by Lohnas and Kahana (2013). As with JOLs, one potential explanation for the lack of a word frequency effect on recall performance are limitations in reliability due to the limited number of words per frequency bin.

Experiment 2 again demonstrated the font size illusion, with higher JOLs for words presented in a large font than in a small font, even though font size did not affect recall performance.

Gamma correlations between JOLs and recall performance and the matching index *G* from Brunswik's (1952) lens model converged on the conclusion that people accurately predicted which

12 A linear regression analysis with only word frequency and font size as predictors (i.e., without the additional control variables) revealed similar results: 22 participants (73%) integrated word frequency and font size in their JOLs, or used only word frequency (2 participants), only font size (3 participants), or neither cue (3 participants).

words they would and would not recall. At the same time, JOL resolution was lower than in Experiment 1, where word frequency was manipulated discretely. The lens model analysis suggests that reduced accuracy may be due to participants using word frequency less consistently for their JOLs, presumably because manipulating word frequency on a continuum required participants to take fine-grained differences in word frequency into account. At the same time, the fact that word frequency increased JOLs but left recall performance unaffected also contributed to reduced JOL resolution.

When investigating cue utilization at the individual level, results again revealed that participants based their JOLs on both word frequency and font size, suggesting integration of both cues. Thus, manipulating word frequency on a continuum did not interfere with cue integration in JOLs.

9.5. Experiment 3

The previous experiments demonstrated that word frequency affected JOLs when font size was simultaneously manipulated. This was found both when word frequency was manipulated in two discrete levels and when it was manipulated on a continuum. As all prior studies addressing the effects of word frequency effects on JOLs, Experiments 1 and 2 focused on objective word frequency but did not take subjective word frequency into account. As explained in the introduction, there is evidence that subjective word frequency affects memory (e.g., Deese, 1959; Tulving & Kroll, 1995; see also Popov & Reder, 2020) and it is plausible to assume that the same is true for metamemory. For instance, students who major in psychology presumably encounter the word *cognition* more often and in more diverse contexts than students who major in math. Because of that, it is likely that psychology students expect to remember this specific word better than math students.

Experiment 3 therefore aimed to explore the role of subjective frequency in the word frequency effect on JOLs. To achieve this goal, psychology and math students studied a list of words that included 10 words of each of the following types: words of high objective frequency (standard high-frequency words), words of low objective frequency unrelated to math or psychology (standard low-frequency words), words of low objective frequency that are related to maths and, hence, very familiar to math students (math words), and words of low objective frequency that are related to psychology and, hence, very familiar to psychology students (psychology words).

We expected that JOLs would increase with both objective and subjective frequency, as well as with font size. Recall performance was expected to be better for words of high objective and subjective frequency than for low-frequency words.

9.5.1. Method

9.5.1.1. Participants. The sample size for this experiment was based on the effect sizes obtained in Experiment 1. As in the previous experiment, we aimed for a total $N = 30$ to obtain a statistical power of $(1 - \beta) = .99$ to detect large effects ($f = .40$). Thirty University of Mannheim students (11 male) with an average age of 21.77 years ($SD = 3.38$) participated in this experiment in exchange for course credit or money. Half the participants majored in psychology and the other half majored in courses with math-related curricula (Mathematics in Business and Economics, Business Informatics).

9.5.1.2. Design and Materials. We used a 4 (word type: standard high-frequency, standard low-frequency, math, psychology) x 2 (font size: 48 pt, 18 pt) x 2 (major: math, psychology) mixed design, with word type and font size manipulated within subjects. JOLs and correct recall were measured. A new list of 40 German nouns was created (see Appendix D), comprising 10 words from each word type. We selected math and psychology words in a pilot study. First, math and psychology students provided us with a list of words they frequently encountered in their classes (23 math words and 31 psychology words). Then, $N = 5$ math and $N = 5$ psychology students rated all 54 words in terms of valence, arousal, and concreteness. Finally, we selected 10 math and 10 psychology words that were similar to each other and to the standard high-frequency and low-frequency words in valence, arousal, concreteness, number of syllables, and number of letters (all $p > .09$, see Table 10). Also, math, psychology, and standard low-frequency words were similar in objective frequency (all $p > .33$) and were of reliably lower objective word frequency than standard high-frequency words (all $p < .001$).

9.5.1.3. Procedure. The procedure was identical to the previous experiments. Again, half of the words from each type (standard high-frequency, standard low-frequency, math, psychology) were presented in 48 pt and the other half was presented in 18 pt.

9.5.2. Results

9.5.2.1. Judgments of Learning. JOLs are presented in Figure 6. A 4 (word type) x 2 (font size) x 2 (major) mixed ANOVA revealed a main effect of word type, $F(3, 84) = 13.96$, $p < .001$, $\eta_p^2 = .33$, that was qualified by a significant interaction between word type and major, $F(3, 84) = 16.32$, $p < .001$, $\eta_p^2 = .37$. A main effect of font size, $F(1, 28) = 24.38$, $p < .001$, $\eta_p^2 = .47$, indicated that JOLs were higher for words presented in the large font size rather than in the small font size. There was no

Table 10

Means (and Standard Deviations) of Word Frequency, Concreteness, Valence, and Arousal for the Words Used in Experiment 3.

Word type	Objective frequency (fpm)	Concreteness	Valence	Arousal
Standard high-frequency	164.28 (49.55)	3.76 (1.33)	0.22 (0.87)	2.60 (0.41)
Standard low-frequency	0.37 (0.24)	4.01 (1.36)	-0.31 (1.26)	2.70 (0.57)
Math	0.41 (0.66)	3.06 (0.95)	-0.24 (0.33)	2.28 (0.34)
Psychology	0.94 (1.68)	3.10 (0.88)	-0.50 (0.63)	2.57 (0.57)

Notes: Objective frequency values (fpm = frequency per million) were taken from dlex-DB (Heister et al., 2011), and concreteness, valence, and arousal values were taken from BAWL-R (Võ et al., 2009) for standard high- and low-frequency words and from a pilot study for math and psychology words (see main text). Math = words with low objective frequency that were related to math; Psychology = words with low objective frequency that were related to psychology.

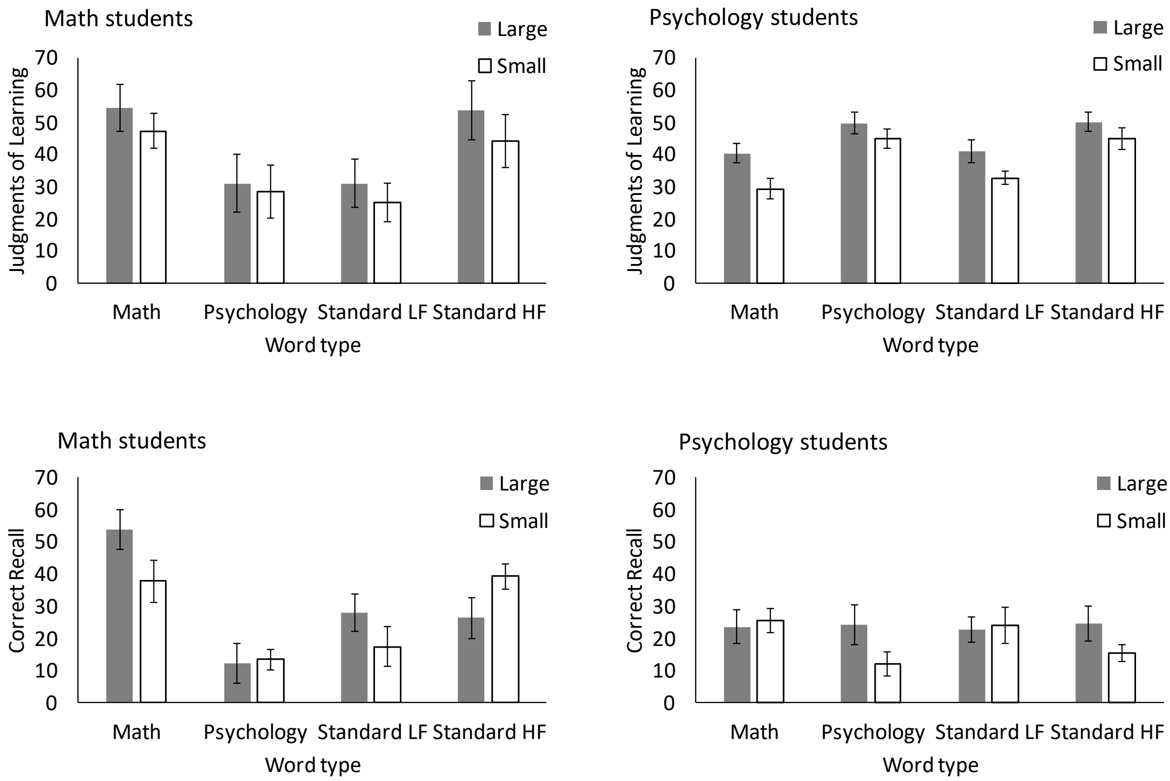
main effect of major, $F(1,28) = 0.22$, $p = .645$, $\eta_p^2 = .008$, and neither the Font size x Major interaction, $F(1, 28) = 0.12$, $p = .731$, $\eta_p^2 = .004$, nor the Word type x Font size x Major interaction, $F(3, 84) = 0.67$, $p = .572$, $\eta_p^2 = .02$, were significant. We followed up on the Word type x Major interaction with separate repeated-measures ANOVAs for each major with word type as the only factor.

For math students, we found a significant main effect of word type, $F(3,42) = 29.02$, $p < .001$, $\eta_p^2 = .68$. Post-hoc *t*-tests showed that JOLs were reliably higher for standard high-frequency words and math words than for psychology words or standard low-frequency words (all $p < .001$). Also, JOLs were similar for standard high-frequency words and math words and for psychology words and standard low-frequency words (both $p = 1.00$).

For psychology students, there was also a significant main effect of word type, $F(3,42) = 5.88$, $p = .002$, $\eta_p^2 = .30$, with overall higher JOLs for standard high-frequency and psychology words than for math and standard low-frequency words. Post-hoc *t*-tests confirmed similar JOLs for standard high-frequency and psychology words and similar JOLs for math and standard low-frequency words (both $p = 1.00$) as well as higher JOLs for psychology and standard high-frequency words than math words ($p = .016$ and $p = .015$, respectively). However, JOLs for psychology words were only marginally higher than JOLs for standard low-frequency words ($p = .065$), and there were no reliable differences between JOLs for standard high-frequency words and standard low-frequency words ($p = .058$).

Figure 6

Mean Judgments of Learning and Correct Recall in Experiment 3.



Notes: Math = words with low objective frequency related to math; Psychology = words with low objective frequency related to psychology; Standard LF = words with low objective frequency unrelated to math or psychology; Standard HF = words with high objective frequency. Error bars represent one standard error of the mean.

9.5.2.2. Recall. Correct recall is presented on Figure 6. A 4 (frequency) x 2 (font size) x 2 (major) mixed ANOVA revealed a main effect of word type, $F(3, 84) = 11.03$, $p < .001$, $\eta_p^2 = .28$, that was qualified by significant interactions between word type and major, $F(3, 84) = 6.14$, $p < .001$, $\eta_p^2 = .18$, and among word type, font size, and major, $F(3, 84) = 5.54$, $p = .002$, $\eta_p^2 = .17$. There was no effect of font size on recall performance, $F(1, 28) = 2.88$, $p = .100$, $\eta_p^2 = .09$. A main effect of major indicated that math students showed better memory performance than psychology students, $F(1, 28) = 5.60$, $p = .025$, $\eta_p^2 = .17$. This finding was unexpected and will be addressed in the discussion. The interaction between font size and major was not significant, $F(3, 84) = 0.09$, $p = .764$, $\eta_p^2 = .003$. To explore the significant interactions, we performed separate 4 (word type) x 2 (font size) repeated-measures ANOVAs for each major.

For math students, correct recall differed significantly between word types, $F(3, 42) = 15.41$, $p < .001$, $\eta_p^2 = .52$. Post hoc *t*-tests showed that recall was significantly better for math than for psychology words ($p < .001$) and for standard low-frequency words ($p = .007$), and marginally better for math than for standard high-frequency words (p = .087). Also, recall was significantly better for standard high-frequency than for psychology words (p = .003). There was no significant effect of font size on memory performance, $F(1, 14) = 1.10$, $p = .311$, $\eta_p^2 = .07$. A significant Word type x Font size interaction, $F(3, 42) = 5.38$, $p = .003$, $\eta_p^2 = .28$, occurred because of numerically higher recall of psychology words presented in the small font size than in the large font size and inverse patterns for all other word types (see Figure 6).

For psychology students, neither of the main effects nor the interaction were significant, word type: $F(3, 42) = 0.79$, $p = .505$, $\eta_p^2 = .05$, font size: $F(1, 14) = 1.80$, $p = .202$, $\eta_p^2 = .11$, interaction: $F(3, 42) = 1.39$, $p = .260$, $\eta_p^2 = .09$.

9.5.2.3. JOL Accuracy. As in Experiments 1 and 2, JOL accuracy was analyzed by the Gamma correlation between JOLs and recall and by Brunswik's (1952) lens model. Descriptive statistics are presented in Table 8. Gamma was significantly positive, $t(29) = 8.89$, $p < .001$, $d = 1.62$, indicating that participants' JOLs were accurate in discriminating which items they would and would not recall. The lens model included the effect-coded variables objective frequency (+1 = standard high-frequency; -0.33 = standard low-frequency, math, psychology), font size (+1 = large font size; -1 = small font size), and subjective frequency (0 = standard high-frequency; +1 = math words for math students, psychology words for psychology students; -0.5 = standard low-frequency, psychology words for math students, math words for psychology students), as well as the same control variables as in Experiments 1 and 2 (number of letters, number of syllables, valence, arousal, and concreteness). The lens model analysis revealed relatively low consistency of cue use in JOLs (R_{JOL}) and linear predictability of recall (R_{REC}). Parameter C significantly differed from zero, $t(29) = 5.86$, $p < .001$, $d = 1.07$. As in Experiments 1 and 2, the matching index G significantly exceeded r_a, $t(29) = 4.56$, $p < .001$, $d = .83$, and exceeded Gamma numerically but not significantly, $t(29) = 0.70$, $p = .490$, $d = .13$.

9.5.2.4. Cue Integration at the Individual Level. As in Experiments 1 and 2, we evaluated individual cue use by examining regression weights from the linear regression of JOLs on cues obtained in the lens model analysis. Eighteen participants (60%) integrated objective frequency, subjective frequency, and font size in their JOLs. The remaining participants based their JOLs on objective and subjective frequency (5 participants), on objective frequency and font size (1 participant), on subjective frequency and font size (2 participants), only on objective frequency (1 participant), only on subjective frequency (2 participants), or only on font size (1 participant).[13] Defining cue use as regression weights that exceeded Cohen's (1977) convention for small effect sizes ($|r| \geq .10$) revealed that 11 participants (36.67%) integrated all three cues in their JOLs. The remaining participants based their JOLs on objective and subjective frequency (5 participants), on objective frequency and font size (1 participant), on subjective frequency and font size (4 participants), or only on objective frequency (3 participants), only on subjective frequency (4 participants), or only on font size (2 participants).

9.5.3. Discussion

As in the previous experiments, both objective word frequency and font size affected JOLs, with higher JOLs for words of high objective word frequency and words presented in a large font size. A new finding was that high subjective word frequency increased JOLs: Participants predicted better recall for words with low objective frequency that were related to their major than for other low-frequency words. This was likely the first demonstration that subjective word frequency affects JOLs over and above objective word frequency.

Regarding memory performance, math and psychology students differed in how word type and font size affected their memory as well as in their overall level of performance. Math students recalled math words and standard high-frequency words better than psychology words and standard low-frequency words. These results demonstrate the expected memory benefits for frequent and familiar words over rare and unfamiliar words (e.g., Popov & Reder, 2020). In contrast, psychology students showed similar memory performance for all word types. In combination with the finding that math students outperformed psychology students, this could suggest that psychology students were less motivated to succeed in the memory test. Alternatively, or additionally, recurrent exposition to words in memory experiments, in which psychology students participate much more often than math students, might have placed psychology students at a disadvantage.

As in the previous experiments, gamma correlations between JOLs and recall performance and the matching index *G* from Brunswik's (1952) lens model revealed that people accurately predicted which words they would and would not recall.

Examining cue integration at the individual level revealed that the majority of participants based their JOLs on both objective and subjective word frequency and on font size. This demonstrates that

---

13 A linear regression analysis without the control variables revealed similar results: 20 participants used all three cues, 3 used objective and subjective frequency but not font size, 3 used subjective frequency and font size but not objective frequency, 2 used only subjective frequency, 1 used only objective frequency, and 1 used only font size. These results are also consistent with an analysis based on simple mean differences.

previous findings on cue integration of word frequency and font size extend to subjective word frequency.

9.6. General Discussion

The present research examined whether objective and subjective word frequency affect JOLs when an additional cue is present. Although researchers have recently begun to address the issue of whether people integrate multiple cues in their JOLs (e.g., Tatz & Peynircioğlu, 2020; Undorf et al., 2018), it was an open question whether word frequency effects on JOLs persisted in multiple cue situations. Across three experiments, we simultaneously manipulated word frequency and font size, requested immediate JOLs, and assessed free recall performance. In all experiments, the majority of participants based their JOLs on both word frequency and font size. Overall, JOLs were higher for high-frequency words than for low-frequency words, and higher for words presented in a large font size rather than a small font size.

This study provided several novel insights into the role of word frequency in metamemory. First, this study highlighted that word frequency has a pervasive effect on JOLs. In particular, word frequency affected JOLs when font size was simultaneously manipulated. This finding is in line with a previous study demonstrating that word frequency affected JOLs when valence and arousal were also manipulated (Hourihan et al., 2017). In the current Experiments 1 and 3, in which objective word frequency was manipulated in two discrete levels, JOLs increased with word frequency. This finding was extended in Experiment 2, in which objective word frequency was manipulated on a continuum, which made the selection of words more representative and arguably reduced the saliency of the word frequency manipulation (e.g., Dhami et al., 2004).

Moreover, the present study is the first to demonstrate that subjective word frequency affects JOLs over and above objective word frequency. Experiment 3 revealed that students majoring in math or psychology assigned high JOLs not only to high-frequency words but also to low-frequency words related to their respective major. This demonstrated that subjective word frequency plays an important role in metamemory. We manipulated subjective word frequency by selecting words participants frequently encountered in their lectures and that were therefore familiar to them. While our results clearly supported the validity of this manipulation, future research should use alternative manipulations of subjective word frequency. In particular, it might be interesting to examine how differences in subjective word frequency between individual participants rather than groups of participants affect metamemory judgments. Also, future research could benefit from including an independent measure of subjective word frequency.

Another novel finding was that word frequency dissociated JOLs and recall performance when using a continuous manipulation of word frequency in Experiment 2. With discrete manipulations of word frequency, prior studies typically found higher JOLs and higher recall performance for high-frequency than for low-frequency words. In contrast, in Experiment 2, JOLs continued to increase with increasing word frequency, whereas word frequency was unrelated to actual recall performance. Notably, mismatches between effects of word frequency on JOLs and memory would have occurred even if we replicated the U-shaped relationship between word frequency and memory performance found by Lohnas & Kahana (2013). The very different effects of word frequency on JOLs and memory performance underline the inferential nature of JOLs and demonstrate that JOL accuracy depends on the validity of the cues that inform JOLs (Koriat, 1997).

In the three experiments reported here, font size consistently affected JOLs in addition to word frequency. Namely, JOLs were higher for words presented in a large rather than a small font size. This finding replicates previous research showing that font size affected JOLs when manipulated simultaneously with other variables (e.g., Undorf et al., 2018). Even though we manipulated font size in two discrete levels in all experiments, there is little reason to expect that results would change when varying font size on a continuum (see also Undorf & Zimdahl, 2019). Because font size did not affect memory performance, the present study replicated prior work on the font size illusion (Hu et al., 2015; Luna et al., 2018; Mueller et al., 2014; Rhodes & Castel, 2008; Undorf & Zimdahl, 2019).

The current findings also contribute to the literature on cue integration in JOLs. In particular, the present finding of people basing their JOLs on both word frequency (objective or objective and subjective) and font size provides further evidence for cue integration in JOLs. The strongest evidence for this conclusion came from individual-level analyses showing that across all three experiments, the majority (60%-66.67%) of participants integrated both word frequency and font size in their JOLs.

Also, the present study suggested that Brunswik's (1952) lens model is useful for analyzing the accuracy of metacognitive judgments in multiple cue situations (cf. Bröder & Undorf, 2019). Particularly, in all three experiments, we found that the lens model's matching index G was numerically higher than Gamma correlations and significantly higher than Pearson correlations between JOLs and recall. This was expected, since the matching index G is corrected for unreliability in memory performance and judgment processes, while the surface correlation between JOL and recall conflates random noise and the metacognitive ability of the judge (Bröder & Undorf, 2019). Even though the matching index *G* did not significantly exceed Gamma (see Bröder & Undorf, 2019), the current findings suggest that the matching index *G* is a valid measure of resolution. Across the three experiments reported here, the lens model's parameter *C* was slightly lower in Experiment 1 where font size and objective word frequency were manipulated in two discrete values than in Experiment 2 where word frequency varied on a continuum and in Experiment 3 where subjective word frequency was also manipulated. Considering that the *C* coefficient was hypothesized to reflect idiosyncratic cues (Bröder & Undorf, 2019), this might be taken to suggest that the impact of idiosyncratic cues on JOLs increases when there are more fine-grained differences in word frequency and with a higher number of cues.

Unexpectedly, Experiment 3 revealed several differences between participants who majored in math and psychology. First, subjective word frequency had a stronger effect on JOLs made by math students than on JOLs made by psychology students. For instance, differences between the JOLs psychology students attributed to psychology and standard low-frequency words were only marginal, whereas differences between the JOLs math students attributed to math and standard low-frequency words were reliable. Second, recall performance in math students was affected by objective and subjective frequency, whereas neither type of frequency affected recall performance in psychology students. Finally, math students achieved better recall performance than psychology students. Further research will be needed to determine whether differences between the two student populations replicate and if so, to understand the reasons for these differences. Until this research has been done, we can only speculate. One possibility is that proactive interference from prior memory experiments plays a role (Underwood, 1957). Because psychology students take part in such experiments more often than math students, this might contribute to performance differences. It is also possible that math students were more motivated to perform well than psychology students, which may or may not be related to participating less often in memory experiments.

In summary, the current study demonstrated the pervasiveness of word frequency effects on JOLs when an additional cue was simultaneously manipulated. Furthermore, it revealed that subjective word frequency affected JOLs over and above objective word frequency.

References

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814–823. https://doi.org/10.1111/j.1467-9280.2006.01787.x

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29*(4), 639–647. https://doi.org/10.3758/BF03200465

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445–459. https://doi.org/10.3758/BF03193014

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. https://doi.org/10.1016/0749-596X(89)90016-8

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31*(2), 297–305. https://doi.org/10.3758/BF03194388

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky, & R. A. Bjork (Eds.). *Handbook of metamemory and memory* (pp. 73–94). Psychology Press.

Bröder, A., & Undorf, M. (2019). Metamemory viewed through the judgment lens. *Acta Psychologica, 197*, 153–165. https://doi.org/10.1016/j.actpsy.2019.04.011

Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago University Press.

Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *The Quarterly Journal of Experimental Psychology, 64*(3), 545–559. https://doi.org/10.1080/17470218.2010.503374

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology, 58*(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Academic Press. https://doi.org/10.1016/c2013-0-10517-x

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17–22. https://doi.org/10.1037/h0046671

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin, 130*(6), 959–988. https://doi.org/10.1037/0033-2909.130.6.959

Dunlosky, J., Mueller, M. L., & Tauber, S. K. (2015). The contribution of processing fluency (and beliefs) to people's judgments of learning. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas & H. L. Roediger III (Eds.), *Remembering: Attributions, processes, and control in human memory* (Part I, pp. 46–64). Psychology Press. https://doi.org/10.4324/9781315752808

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fiacconi, C. M., & Dollois, M. A. (2020). Does word frequency influence judgments of learning (JOLs)? A meta-analytic review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale.* Advance online publication. https://doi.org/10.1037/cep0000206

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*(4), 506–528. https://doi.org/10.1037/0033-295X.98.4.506

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association, 49*(268), 732–764. https://doi.org/10.2307/2281536

Hammond, K. R., & Stewart, T. R. (Eds.). (2001). *The essential Brunswik: Beginnings, explications, applications*. Oxford University Press.

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB - eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau, 62*(1), 10–20. https://doi.org/10.1026/0033-3042/a000029

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2017). The influences of valence and arousal on judgments of learning and on recall. *Memory & Cognition, 45*(1), 121–136. https://doi.org/10.3758/S13421-016-0646-3

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. https://doi.org/10.1371/journal.pone.0142351

Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2016). The effect of word frequency on judgments of learning: Contributions of beliefs and processing fluency. *Frontiers in Psychology, 6*:1995. https://doi.org/10.3389/fpsyg.2015.01995

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134*(3), 404–426. https://doi.org/10.1037/0033-2909.134.3.404.

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. https://doi.org/10.1037/0096-3445.126.4.349

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. https://doi.org/10.1177/0956797611407929

Loewenthal, K., & Gibbs, G. (1974). Word familiarity and retention. *Quarterly Journal of Experimental Psychology, 26*(1), 15–25. https://doi.org/10.1080/14640747408400383

Lohnas, L. J., & Kahana, M. J. (2013). Parametric effects of word frequency in memory for mixed frequency lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(6), 1943–1946. https://doi.org/10.1037/a0033669

Luna, K., Martín-Luengo, B., & Albuquerque, P. B. (2018). Do delayed judgements of learning reduce metamemory illusions? A meta-analysis. *Quarterly Journal of Experimental Psychology, 71*(7), 1626–1636. https://doi.org/10.1080/17470218.2017.1343362

Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 509–527. https://doi.org/10.1037/a0014876

Mendes, P. S., Luna, K., & Albuquerque, P. B. (2019). Word frequency effects on judgments of learning: More than just beliefs. *The Journal of General Psychology*. Advance online publication. https://doi.org/10.1080/00221309.2019.1706073

Mendes, P. S., Luna, K., & Albuquerque, P. B. (2020). Experience matters: Effects of (in)congruent prompts about word frequency on judgments of learning. *Zeitschrift für Psychologie, 228*(4), 254–263. https://doi.org/10.1027/2151-2604/a000423

Monaghan, P., Chang, Y. N., Welbourne, S., & Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language*, *93*, 1–21. https://doi.org/10.1016/j.jml.2016.08.003

Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people's metamemory judgments: What are contributions of processing fluency and beliefs about memory? *Quarterly Journal of Experimental Psychology, 69*(4), 781–799. https://doi.org/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. https://doi.org/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 and Review, 20*(2), 378–384. https://doi.org/10.3758/s13423-012-0343-6

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.). *The psychology of learning and motivation* (Vol. 26., pp.125–141). Academic Press.

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology, 20*(2), 139–156. https://doi.org/10.1002/acp.1178

Peters, H. N. (1936). The relationship between familiarity of words and their memory value. *The American Journal of Psychology, 48*(4), 572–584. https://doi.org/10.2307/1416508

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review, 127*(1), 1–46. https://doi.org/10.1037/rev0000161

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. https://doi.org/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. https://doi.org/10.1080/13825585.2015.1102194

Reder, L. M., Liu, X. L., Keinath, A., & Popov, V. (2016). Building knowledge requires bricks, not sand: The critical role of familiar constituents in learning. *Psychonomic Bulletin & Review, 23*(1), 271–277. https://doi.org/10.3758/s13423-015-0889-1

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 65–80). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.4

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. https://doi.org/10.1037/a0013684

Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin, 137*(1), 131–148. https://doi.org/10.1037/a0021705

Shanks, L. L., & Serra, M. J. (2014). Domain familiarity as a cue for judgments of learning. *Psychonomic Bulletin & Review, 21*(2), 445–453. https://doi.org/10.3758/s13423-013-0513-1

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review, 22*(4), 1014–1019. https://doi.org/10.3758/s13423-014-0768-1

Tatz, J. R., & Peynircioğlu, Z. F. (2020). Judgments of learning in context: Backgrounds can both reduce and produce metamemory illusions. *Memory & Cognition, 48*(4), 581–595. https://doi.org/10.3758/s13421-019-00991-9

Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin & Review, 2*(3), 387–390. https://doi.org/10.3758/BF03210977

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. https://doi.org/10.1037/xlm0000571

Undorf, M., & Bröder, A. (2020). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology, 73*(4), 629–642. https://doi.org/10.1177/1747021819882308

Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory and Cognition, 46*(4), 507–519. https://doi.org/10.3758/s13421-017-0780-6

Võ, M.L.-H., Conrad, M., Kuchinke, L., Urton, K, Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL–R). *Behavior Research Methods, 41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534

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. https://doi.org/10.3758/s13421-016-0681-0

10. Using Brunswik's (1952) Lens Model With Previous Data

After considering the importance of the lens model to examine JOL accuracy, we decided to employ it to data from our previous studies. Note, however, that these were meant as complementary analyses, so they will not be discussed exhaustively.

We employed the lens model for immediate JOLs in all experiments. As in Study 3, we examined JOLs accuracy by calculating the Gamma correlation between JOLs and recall and by performing a lens model analysis. We applied the lens model to our data by following the steps provided by Bröder and Undorf (2019) for SPSS. The model included the effect coded independent variable (+1 = high frequency; -1 = low frequency) and the centered control variables number of letters, number of syllables, concreteness, and subjective frequency (and prompt for experiments from Study 2).

Additionally, we examined standardized regression weights from the linear regression of JOLs on cues obtained in the lens model analysis. This allowed us to evaluate if participants used word frequency in their JOLs and further explore whether participants integrated word frequency and information given by the prompts in Study 2.

10.1. Results

10.1.1. JOL Accuracy

Descriptive statistics are presented in Table 11. Across all experiments, Gamma was significantly positive (all $p < .001$), indicating that participants' JOLs were accurate in predicting which items they would and would not recall. The lens model analysis revealed that the consistency of cue use in JOLs and the linear predictability of recall were limited, as shown by relatively low lens model parameters R_{JOL} and R_{REC}, especially concerning the Pilot Study. Parameter C also significantly differed from zero in all experiments (all $p < .001$), indicating that the linear combination of the cues included in the lens model could not account for all systematic covariance between JOLs and recall. These findings are similar to what we found in Study 3, and in previous literature (see Bröder & Undorf, 2019). The matching index *G* significantly exceeded r_a in all experiments (all $p < .002$) except in Experiment 2 of Study 1, $t(29) = 1.82$, $p = .080$, $d = 0.33$, and in Experiment 3 of Study 2, $t(39) = 1.74$, $p = .090$, $d = 0.28$. *G* only significantly exceeded Gamma in the Pilot Study, $t(27) = 2.58$, $p = .016$, $d = 0.78$, and in the Experiment 1 from Study 1, $t(35) = 3.76$, $p < .001$, $d = 0.63$. In the remaining experiments, *G* did not reliably differ from Gamma (all $p > .132$). This is in line with what we found in Study 3 from this dissertation, suggesting that the lens model's matching index G is a valid measure of JOL resolution.

Table 11

Mean Values of Gamma and the Parameters From Brunswik's (1952) Lens Model in the Pilot Study, Studies 1, and 2.

Study	Experiment	Gamma	r_a	G	R_{JOL}	R_{REC}	C
Pilot	-	.41	.27	.59	.33	.25	.24
Study 1	1	.47	.31	.62	.67	.39	.22
	2	.44	.25	.36	.64	.33	.22
	4	.38	.24	.42	.63	.39	.24
Study 2	1a	.42	.22	.38	.66	.40	.17
	1b	.46	.30	.53	.62	.41	.21
	2	.42	.21	.36	.62	.44	.14
	3 (global)	.38	.22	.30	.63	.44	.18
	Instruction HF	.42	.23	.34	.68	.44	.19
	Instruction LF	.35	.20	.27	.59	.43	.18

Notes: Gamma = Goodman-Kruskal gamma correlation between JOLs and recall, r_a= Pearson correlation between JOLs and recall, G = matching index of the lens model, R_{JOL} = multiple correlation between cues and JOLs, R_{REC} = multiple correlation between cues and recall, C = nonlinear component of lens model equation, HF – high-frequency words "better remembered", LF – low-frequency words "better remembered".

10.1.2. Individual Cue Use

As in our Study 3, we conjectured that participants based their JOLs on a cue if the cue's regression weight was positive, and when regression weights exceeded the convention for small effect sizes proposed by Cohen (1977, $|r| \geq .10$). The number of participants that used objective word frequency to make their immediate JOLs is presented in Table 12.

Additionally, we explored if in Study 2 participants integrated word frequency and prompts in their immediate JOLs. This would be a complementary analysis of the individual cue use based on the mean difference provided in the Supplemental Materials of Study 2 (cf. Appendix C.3). When considering positive regression weights, in Experiment 1a, 9 participants (27%) used both word frequency and prompt, 11 (33%) used only frequency, 6 (18%) only used prompt, and the remaining 7 participants (21%) did not use either cue. In Experiment 1b, 9 participants (30%) used both word frequency and prompt, 8 (27%) used only frequency, 8 (27%) only used prompt, and the remaining 5 participants (17%) did not use either cue. In Experiment 2, 4 participants (12%) used both word frequency and prompt, 15 (45%) used only frequency, 4 (12%) only used prompt, and the remaining 10 participants (30%) did not use either cue. Finally, in Experiment 3, 6 participants (15%) used both word frequency and prompt, 8 (20%) used only frequency, 14 (35%) only used prompt, and the remaining 12 participants (30%) did not use either cue. When considering Cohen's (1977) convention for small effect sizes ($|r| \geq .10$), the pattern of results was similar, but fewer participants used the mentioned cues. These results are in line with our findings, showing that when making immediate JOLs, the most participants relied on actual word frequency instead of the prompt's information.

Table 12

Number of Participants who Used Word Frequency as a Cue Based on Regression Weights Obtained Through Brunswik's (1952) Lens Model in the Pilot Study, Studies 1, and 2.

Study	Experiment	Positive regression weights	Regression weights > .10
Pilot	- (*N* = 28)	15 (54%)	4 (14%)
Study 1	1 (*N* = 36)	26 (72%)	20 (56%)
	2 (*N* = 30)	22 (73%)	19 (63%)
	4 (*N* = 40)	20 (50%)	18 (45%)
Study 2	1a (*N* = 33)	20 (61%)	17 (52%)
	1b (*N* = 30)	17 (57%)	10 (33%)
	2 (*N* = 33)	19 (57%)	15 (45%)
	3 (global) (*N* = 40)	14 (35%)	13 (33%)
	Instruction HF	6 (30%)	5 (25%)
	Instruction LF	8 (40%)	8 (40%)

10.2. Discussion

After employing Brunswik's (1952) lens model to explore JOL accuracy in Study 3, we recognized its contribution as a measure of resolution that corrects for unreliability in memory performance and judgment processes. Hence, we decided to test whether it would be a valid resolution measure when applied to our previous data.

Our results showed that, overall, the matching index *G* is a superior measure of resolution than the surface correlation between JOLs and recall (r_a). However, it did not exceed Gamma correlations in most of our experiments. Nonetheless, we think *G* is at least as valid as Gamma when assessing JOL accuracy, with the advantage of controlling for unreliability in memory performance and judgment processes, which could be mostly useful in multiple-cue scenarios.

We noted that the matching index *G* was relatively lower in Experiments 2 and 3 (better-for-low) from Study 2. It is not clear why that is the case, but one possibility is that this indication of lower accuracy may be related to the instruction that low-frequency words are better remembered, which had a small effect on participant's immediate JOLs. Attributing higher JOLs to low-frequency words and later fail to recall them would result in lower accuracy. However, this was not verified by gamma correlations, so it is mere speculation. Another interesting result is related to the lowest consistency of cue use in JOLs (R_{JOL}) in the Pilot Study. Again, we can only speculate, but this could be because words were mainly of mid-frequency, showing that the consistency of frequency use when making JOLs was very limited. This is also in line with the lowest linear predictability of recall (R_{REC}), showing that word frequency did not predict recall.

When assessing individual cue use based on standardized beta weights, results revealed that, overall, the majority of participants used word frequency as a cue to inform their JOLs. Interestingly, the percentage of participants that relied on word frequency was higher in experiments from Study 1 than in experiments from Study 2. This could be to the presence of an additional cue in Study 2 (i.e., prompt), which was incongruent for half of the items. These findings provided further support that immediate JOLs were mostly based on actual frequency (see also Appendix C).

In summary, Brunswik's (1952) offers a valid measure of JOL resolution that, although not reliably exceeding gamma correlations, controls for unreliability in memory performance and judgment processes and allows to explore individual cue use in multiple-cue scenarios. This is certainly useful for future metamemory research.

References

Bröder, A., & Undorf, M. (2019). Metamemory viewed through the judgment lens. *Acta Psychologica, 197*, 153–165. https://doi.org/10.1016/j.actpsy.2019.04.011
Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago University Press.
Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Academic Press.

PART III

FINAL DISCUSSION

11. The Contribution of the Present Work to Metamemory Research

The aim of this dissertation was to better understand the mechanisms behind the word frequency effect on JOLs. More specifically, this research aimed to test if in addition to theory-driven processes, experience-driven processes also contributed to the word frequency effect on JOLs. We accomplished this by manipulating word frequency alone (Pilot Study, Studies 1, and 2) and in combination with font size (Study 3). Taking into consideration all experiments from this work, results suggest that: (1) theory-driven factors do not fully account for the word frequency effect on JOLs; (2) processing fluency, measured by response times in a lexical decision task, also contributes to the effect of word frequency on JOLs; (3) beliefs about word frequency can be easily changed using a counter-belief, with an impact on JOLs; (4) word frequency affected JOLs even when another salient cue (i.e., font size) was present, showing that people can integrate word frequency and other cues in their JOLs; (5) subjective measures of frequency seem to play an important role in word frequency effects on JOLs; and (6) Brunswik's (1952) lens model offers a valid measure of JOL resolution. This chapter will summarize the main findings from this dissertation, interpreting in detail their contributions to the field of metamemory research.

11.1. Theory-driven Factors Do not Fully Explain the Word Frequency Effect on JOLs

Attending to Koriat's (1997) cue-utilization approach, intrinsic cues like word frequency would directly impact JOLs through theory-driven processes such as previous knowledge or beliefs about its impact on memory. In addition to this explanation, it is also possible that word frequency affects JOLs indirectly through mnemonic cues and, thus, through experience-driven processes, such as processing fluency (Koriat, 1997). The idea is that high-frequency words would be more easily processed than low-frequency words while studying, which would result in the subjective feeling that they are more easily learned (Begg et al., 1989). Based on this experience, that subjective feeling would guide people's JOLs. Previous research has explained the word frequency effect on JOLs based exclusively on beliefs (Benjamin, 2003; Jia et al., 2016; Tullis & Benjamin, 2012). However, in the introduction of this dissertation, we presented several arguments against this view, and the conducted studies provide evidence supporting our claims. We argued that relying on beliefs of how word frequency affects memory to make JOLs would require that people can identify word frequency as a relevant cue. Results from Study 1 (Experiments 1, 2, and 4) showed that some participants indeed identified word frequency as a cue, further stating they relied on it to make their JOLs. However, some participants did not base their JOLs on word frequency, but their JOLs were still significantly higher for high-frequency words than for low-frequency words. This finding suggested that theory-driven processes could not fully account for the effect of word frequency on JOLs for these participants. Hence, our work challenged the previous findings supporting a belief-only explanation of the effect. Accordingly, if the effect of word frequency on JOLs is not based on the belief of how word frequency impacts memory, then there must be other processes at play.

11.2. The Contribution of Processing Fluency to the Word Frequency Effect on JOLs

As described before, processing fluency may also guide the effect of word frequency on JOLs. Previous research from word processing shows that high-frequency words are processed faster than low-frequency words (e.g., Brysbaert et al., 2018), suggesting high-frequency words to be more fluently processed. This was also true for the materials we used in Study 1 (Experiments 3 and 4). In addition, we directly explored the role of processing fluency, measured through response times in a lexical decision task (Experiment 4), in the word frequency effect on JOLs. We found evidence that processing fluency mediated the effect of word frequency on JOLs. Of course, there is the possibility that beliefs about fluency also contribute to the effect, which we did not explore. Beliefs about fluency were previously found to impact JOLs (Mueller & Dunlosky, 2017), so it is likely that they could also be involved regarding word frequency. For example, participants could find high-frequency words easier to read or to pronounce (i.e., more fluent) than low-frequency words and believe that words that are easier to read or to pronounce are better remembered. Nonetheless, our results showed that fluency directly impacted JOLs, which is strong evidence against a belief-only explanation of the word frequency effect on JOLs.

Results from Study 2 provided further evidence favoring our claims against a belief-only explanation of the word frequency effect on JOLs. The strong role of beliefs in the word frequency effect was found using pre-study JOLs (Jia et al., 2016) when there was no direct experience with the words. In Study 2, we asked participants to make both pre-study and immediate JOLs (Experiments 1a, 2, and 3). The to-be-studied words were preceded by a prompt referring to their frequency, but, critically, these prompts (e.g., "You will see a high-frequency word") could be congruent or incongruent with the actual frequency of the words. This design allowed us to assess the role of beliefs, reflected in pre-study JOLs, but also the combined effect of beliefs plus experience, reflected in immediate JOLs. Results from Experiment 1a revealed higher pre-study JOLs for prompts of high frequency, but immediate JOLs were higher for high-frequency words regardless of the prompt presented. We replicated these results when asking only immediate JOLs paired with the (in)congruent prompts in Experiment 1b. Participants' immediate JOLs were based on actual word frequency, which indicates reliance on direct experience with the items, challenging the belief-only explanation of the word frequency effect on JOLs.

11.3. The Impact of a Counter-Belief on the Word Frequency Effect on JOLs

While we found compelling evidence that experience-driven processes contribute to the word frequency effect on JOLs, the influence of beliefs is also evident from past research. To explore how strongly beliefs contribute to the effect of word frequency on JOLs, we directed participants' beliefs by introducing a counter-belief. Previous research found that counter-beliefs may not reverse the font-size bias in immediate JOLs but that they may be enough to eliminate it (Blake & Castel, 2018). It is possible that the same might be true for the effects of other variables on JOLs. Thus, we sought to test the effect of counter-beliefs about word frequency on JOLs.

In Study 2, the instructions of Experiment 2 informed participants that low-frequency words are in general better remembered for they are more salient than high-frequency words. The impact of this manipulation was clear on pre-study JOLs, which were similar for prompts of high and low frequency. In other words, the previously observed effect favoring prompts of high frequency was eliminated, showing that beliefs are malleable enough to be manipulated utilizing an instruction.

Interestingly, the instruction that low-frequency words are better recalled numerically increased immediate JOLs for words preceded by the prompt of low frequency, supporting a small effect of beliefs on immediate JOLs. Still, immediate JOLs were higher for high-frequency words, suggesting reliance on direct experience with the items. Similar results were observed in Experiment 3, in which participants were either instructed that high-frequency words are better remembered (congruent with the belief that was proposed as the basis of the word frequency effect on JOLs) or that low-frequency words are better remembered (i.e., a counter-belief). Participants who received the instruction favoring high-frequency words showed a pattern identical to that of Experiment 1a: higher pre-study JOLs for prompts of high frequency, but higher immediate JOLs for high-frequency words regardless of previous prompt. Participants who received the instruction with a counter-belief showed a pattern identical to that of Experiment 2: similar pre-study JOLs for prompts of high and low frequency, but higher immediate JOLs for high-frequency words, with numerically higher JOLs for words preceded by prompts of low frequency. Together, these findings provide further evidence supporting the role of direct experience with the items in the word frequency effect on immediate JOLs. Additionally, they suggest a small but detectable effect of beliefs about how word frequency impacts memory on immediate JOLs. This finding was partially supported by linear mixed models using pre-study JOLs as a measure of beliefs.

We reviewed the evidence supporting our claim against a belief-only explanation of the word frequency effect on JOLs that derived from our first two studies. These studies challenged the previously established body of research that favored beliefs in terms of the basis of JOLs (e.g., Jia et al., 2016). As previously mentioned in the introduction of this dissertation, the explanation for the impact of different cues on JOLs seemed to move from experience-driven processes, such as fluency, to theory-driven processes, such as beliefs. For example, this occurred for font size (Rhodes & Castel, 2008; Mueller et al., 2014) and, before the empirical evidence presented in this thesis, for word frequency (Begg et al., 1989; Benjamin, 2003; Jia et al., 2016). More recently, however, researchers have found evidence supporting a combination of both types of processes on JOLs. Regarding font size, for example, not only beliefs (e.g., Blake & Castel, 2018) but also fluency (e.g., Undorf et al., 2017; Yang et al., 2018) have been shown to mediate effects on JOLs, suggesting a contribution of both types of processes. Regarding word frequency, which is a less explored variable in metamemory (Fiacconi & Dollois, 2020), the present dissertation provides empirical evidence that supports a combination of both types of processes on JOLs. The combined contribution of both theory-driven and experience-driven factors is in line with Koriat's (1997) cue-utilization approach, which expresses the complexity of metamemory judgments and the basis of JOLs in particular. To disentangle each factor's contribution to the word frequency effect on JOLs is challenging, but this dissertation presents the first step to address this problem.

As mentioned, a critical piece of our findings supported the contribution of processing fluency to the word frequency effect on JOLs. This aspect underlines the importance of matching the methodologies adopted to capture differences in processing fluency with the type of fluency to be assessed. We argued that high- and low-frequency words may not differ in perceptual fluency, which was measured by Jia et al. (2016) by presenting easy vs. difficult font styles, but differ instead in lexical fluency, which could be measured through response times in lexical decision tasks. Although previous research has also found differences in perceptual fluency through lexical decision tasks (e.g., Luna et al., 2019), other measures, such as a continuous identification task (Yang et al., 2018), seem more adequate. In fact, a recent work of Vogel et al. (2020) suggests that fluency effects on judgments are more evident when the type of processing fluency fits the type of judgment to be made. Accordingly, this underlines that processing fluency measures must be adequate and sensitive to better understand its role as a basis of different judgments. Indeed, we emphasize that future research exploring the influence of fluency on metamemory should carefully attend to the variety of types of fluency and methodologies used to assess them. Additionally, it also may be relevant to develop new measures or manipulations that provide better insight on disentangling the role of different types of fluency, such as the recent continuous identification task (Yang et al., 2018).

11.4. The Word Frequency Effect on JOLs Endures Multiple-cue Scenarios

To better understand the word frequency effect on JOLs, we also extended our findings to multiple-cue scenarios. Situations in which there is more than one cue may be closer to real-world decision-making, such as when making JOLs (Rhodes, 2016). When multiple cues are available, it might be harder to identify each cue and to rely on beliefs of how each impacts memory. There has been a recent interest in this topic, with findings suggesting the ability to integrate multiple cues in JOLs (e.g., Tatz & Peynircioğlu, 2020; Undorf et al., 2018). For example, it was learned that people could integrate font size and up to three additional cues in their JOLs (Undorf et al., 2018). Regarding word frequency, previous studies found mixed results when word frequency was manipulated in the presence of other cues (Hourihan et al., 2017; Susser & Mulligan, 2015; Susser et al., 2017). Study 3 of this dissertation tested if objective and subjective word frequency affected JOLs when an additional cue was present and if participants would be able to integrate word frequency and font size in their JOLs. Evidence supported cue integration both at the aggregate and individual levels, with both word frequency and font size affecting JOLs. Overall, we observed that JOLs were higher for high-frequency words than for low-frequency words and higher for words presented in large font size than for words presented in small font size. This was true when word frequency was manipulated discretely (Experiment 1) and in a continuum (Experiment 2), and even when the subjective frequency was also manipulated (Experiment 3).

The results from Study 3 provide new insights into the word frequency effect on JOLs, showing its pervasiveness when another cue is present. This was the first demonstration of word frequency affecting JOLs in combination with another cue and is in line with results from Hourihan et al. (2017), who found effects of word frequency JOLs in the presence of additional cues. Importantly, here both word frequency and font size affected JOLs at an aggregated level, but also at the individual level, which further supports cue integration. Across all three experiments, font size consistently affected JOLs but not recall, replicating the metamemory illusion known as the font size effect (e.g., Rhodes & Castel, 2008).

Another novel finding was the dissociation between JOLs and recall performance regarding the word frequency effect on JOLs in Experiment 2. Previous studies with discrete manipulations of word frequency typically found both higher JOLs and recall performance for high-frequency words than for low-frequency words, as we found in Experiment 1. When manipulating word frequency continuously in Experiment 2, JOLs increased with word frequency while recall was nearly unaffected by word frequency. Even if we had found the U-shaped curve relationship between word frequency and recall (e.g., Lohnas & Kahana, 2013), we would still observe mismatches between JOLs and recall performance. This is further evidence for the inferential nature of JOLs, demonstrating that JOL accuracy depends on the validity of cues informing JOLs.

11.5. Subjective Frequency Also Affected JOLs

As previously mentioned, Study 3 addressed an issue that had not been explored before in the literature, which is the role of subjective word frequency on JOLs. High-frequency words are usually more familiar to people since they also have high subjective frequency (e.g., Brysbaert et al., 2018). Familiarity could promote fluency (e.g., Oppenheimer, 2006), which could explain why people give higher JOLs to high-frequency words. In Experiment 3, we presented high- and low-frequency words, as in our previous experiments, but added words related to math and psychology with low objective frequency. This manipulation allowed us to explore if words related to students' major, albeit with low objective frequency, would be given higher JOLs. In other words, this enabled us to examine the role of subjective frequency on JOLs. Results collected with psychology and math students showed that JOLs were higher for standard high-frequency words and for words related to students' own major than standard low-frequency words and words related to the other major. These findings provide the first evidence that subjective word frequency also affects JOLs. However, it is noteworthy that the manipulation of subjective word frequency was conducted by selecting words that were related or unrelated to the participant's major. While our results supported the effectiveness of this manipulation, future research should use alternative manipulations of subjective word frequency, such as those taken from independent databases, or by asking participants to rate each word's frequency in addition to making JOLs (see suggestions in the next chapter).

11.6. Brunswik's (1952) Lens Model to Examine JOL Accuracy

Finally, another relevant contribution of this dissertation was the use of Brunswik's (1952) lens model to explore JOL accuracy in Study 3. The lens model was recently employed to examine JOL accuracy by Bröder and Undorf (2019), who directly explored the contribution of cues validities (i.e., the cue weights on memory) and utilizations (i.e., cue weights on JOLs) in multiple-cue scenarios. Here, we tested if the lens model would be a valid resolution measure when word frequency and font size were manipulated together. Our results suggested that while the matching index *G* of lens model seems to be an adequate measure of resolution, it did not exceed gamma correlations. However, we argue in favor of the application of this model since it considers random noise in the covariation between JOLs and recall which cannot be attributed to the manipulated cues. Moreover, by decomposing JOL accuracy, the lens model allows to better analyze cue utilization, with the potential to describe memory monitoring at the individual level. This proved useful in this dissertation and certainly will be in future research.

12. Limitations and Future Directions

This chapter focuses on the limitations of this dissertation and future directions for metamemory research. First, we reflect on the issues about reporting word frequency. Then, we consider the difficulties in isolating the effects of subjective word frequency on JOLs and suggest possible solutions. Subsequently, we wonder whether low-frequency words might be so rare that the word frequency effect on JOLs could be considered a nonword effect. Additionally, we discuss how manipulations used here with recall tests, namely the use of a counter-belief, would fare in the face of recognition tests. Finally, we reflect upon what might be the next step regarding metamemory research.

12.1. About Reporting Word Frequency

In the introduction of this dissertation, we reviewed different measures of objective frequency, each of them with advantages and limitations. According to the most recent literature review about the word frequency effects on word processing (i.e., Brysbaert et al., 2018), the most reliable measure would be Zipf. Despite this, we reported values of fpm in all the studies conducted for this dissertation. Although this could have been a problem, especially regarding the Pilot Study, in the subsequent studies we confirmed (a posteriori) that high- and low-frequency words also differed accordingly in Zipf. Nevertheless, we could have reported Zipf values. This could have started a shift in the memory and metamemory literature to use more reliable word frequency measures. By providing both measures here (see Appendixes), the conclusions from this thesis remain, since we made sure high-frequency words and low-frequency words differed significantly across different measures of frequency.

Additionally, it might be relevant to consider and report subjective measures of word frequency. While the role of familiarity in word processing and memory processes is well explored, the same is not true in metamemory, in which the role of subjective frequency has been mostly neglected.

12.2. Clarifying the Role of Subjective Word Frequency

A novelty of this work was the exploration of the effect of subjective word frequency on JOLs in Study 3. However, we operationalized subjective word frequency based on how words related to two groups of people (math and psychology students), and not considering individual differences. For example, a psychology student who is training in clinical psychology may be more familiar with the word "therapy" than a psychology student who is training in organizational psychology. Similarly, a math student that is going through therapy may be similarly familiarized with the word "therapy" as the clinical psychology student. This relates to the differences between item familiarity (e.g., "therapy") and category familiarity (e.g., "psychology"). In fact, it seems that item familiarity modulates the effects of category familiarity on memory performance (Ning et al., 2018), which could also occur for metamemory. Thus, it might be interesting to inspect how individual differences in subjective word frequency, rather than group differences, affect metamemory judgments.

Future research could also benefit from including an independent measure of subjective word frequency, such as those obtained in validated lexical databases. This could allow to disentangle the effects of objective and subjective frequencies. For example, one could orthogonally manipulate objective and subjective frequency. However, while it may be relatively easy to find words with low objective frequency that differ in subjective frequency, it is very difficult to find words with high objective frequency and low subjective frequency. A solution to this problem could be working with a new or a made-up language and manipulating familiarity by means of different rates of exposure. Certainly, such questions could be the focus of future research about word frequency effects on JOLs.

Regardless of how these issues about objective and subjective frequency are explored in the future, it is important to consider that manipulating extreme values of frequency could lead people to think that low-frequency words are nonwords, as we discuss below.

12.3. Might the Word Frequency Effect on JOLs Actually Be a Nonword Effect?

The absence of the effect across different ranges of word frequency in the Pilot Study presented in this dissertation, along with the effect found when word frequency was manipulated with extremes of frequency, left an open question. Since most studies rely on extremes of frequency, low-frequency words may be so rare that most participants may have no lexical representations of them. Hence, low-frequency words could have no meaning, being, in truth, nonwords. Because nonwords are given lower JOLs than words (e.g., Mueller et al., 2014), low-frequency words being considered nonwords could explain the lower JOLs attributed to low-frequency words. It could be interesting to test if JOLs given to nonwords would be similar to JOLs given to extremely low-frequency words such as those used in this dissertation.

Preliminary results from a study not included in this dissertation suggest this may not be the case, and that JOLs are higher for low-frequency words than for nonwords. We hope those results and additional research may shed light on this idea.

That study and the ones included here explored several questions about the contribution of theory-driven and experience-driven processes to the word frequency effect on JOLs in face of a recall test. Below, we reflect on how some of our manipulations would fare when using recognition tests.

12.4. Counter-Beliefs in Recognition Tests

The results reported here, in addition to previous research, suggest that there is a general belief that high-frequency words are better remembered. This belief matches the usual recall performance benefit for high-frequency words over low-frequency words when word frequency is manipulated discretely (e.g., Hall, 1954; Jia et al., 2016; Sumby, 1963). However, in recognition tests, the memory performance advantage is for low-frequency words (e.g., Benjamin, 2003; Clark, 1992; Gorman, 1961; Malmberg & Murnane, 2002). This means that the concept of a counter-belief regarding word frequency would depend on the final memory test. In other words, indicating that low-frequency words are better remembered in the face of a recognition test would be congruent with actual memory performance (hence, not a counter-belief). How would different manipulations of participants' beliefs about word frequency fare when presenting different final memory tests?

The research about the word frequency on JOLs using recognition tests (e.g., Benjamin, 2003; Tullis & Benjamin, 2012) found higher JOLs for high-frequency words in a first study-test cycle but, after a recognition test, JOLs were higher for low-frequency words in a second study-test cycle. Those results suggest that experience may also be informing people's beliefs. Since experience seems to play a bigger role regarding word frequency effects on JOLs, how easy would it be to manipulate people's beliefs in the face of different memory tests? It would be interesting to know the effects of directing people's beliefs congruently with their experience in a memory test. Would such manipulations help to clarify the exact contribution of theory and experience to JOLs? These questions are left open for future research. In the next section we consider what the next steps in metamemory research might be.

12.5. Beyond Theory and Experience: What Is Next for Metamemory Research?

As stated before, there is much evidence that both theory- and experience-driven processes are the basis for the effects of different cues on JOLs. The next step around the basis of JOLs in terms of theory vs. experience may be to understand how much each factor contributes to the observed effects. It seems that for effects of font size on JOLs beliefs are more important (e.g., Frank & Kuhlmann, 2017; Hu et al., 2015; Undorf & Zimdahl, 2019), but with word frequency, we found here that experience is more relevant. This could be different for other cues, and even for distinct metamemory judgments. As mentioned above, one crucial approach to tackle this may be the careful manipulation of fluency, which should be in accordance with the type of fluency measured.

Koriat's (1997) cue-utilization approach and the recently proposed analytic-processing theory (Dunlosky et al., 2015; Mueller et al., 2016; Mueller et al., 2013) already present a theoretical framework for the basis of JOLs that comprise both theory- and experience-driven processes. However, considering the importance of making metacognitive judgments in multiple-cue scenarios, where it might be more difficult to rely on beliefs about the individual contribution of each cue to JOLs, how would these theoretical frameworks fare? Despite the ability of cue integration on JOLs, are people aware of combined effects of different cues? Are they basing their JOLs more on beliefs or on fluency? This has not yet been directly explored, so future research could address it.

The core reason to understand how people make JOLs and other metacognitive judgments is probably related to how accurately they monitor memory processes in the real world. While different measures of resolution exist, research has mostly focused on gamma correlations. To better explore metamemory accuracy, it may be relevant to explore different resolution measures, such as those provided in Brunswik's (1952) lens model, which is more adequate when multiple cues are available. Additionally, how does the accuracy of metacognitive judgments relate to theory-driven and experience-driven processes? Are judgments based on experience more or less accurate than those based on beliefs? Future metamemory research could try to improve accuracy measures and face these issues.

Another critical aspect for real-world situations is the impact of monitoring processes on people's behavior. Past research has also explored how monitoring processes expand to control processes, such as study time allocation after making metacognitive judgments. Even though this field of research has been extensively explored, much work could still be done. For example, when given the chance, would people allocate more time to study low-frequency words? This could be extended, of course, to more ecological materials regarding daily-life scenarios.

Finally, we still do not know why some people’s metacognitive judgments are more accurate than others, or why some people seem to be more reflective of their own memories than others. Are these differences more based on individual differences, such as personality traits? Or is this reflective behavior more socially dependent, and must be learned? These and more naturalistic scenarios could also be interesting to develop in future research.

In summary, by understanding the basic processes behind metacognitive judgments, future research should focus on developing strategies and methods to improve their accuracy. This would allow us to provide tools for people to use in their daily lives, whether they are students monitoring their learning before an exam, older people with failing metacognitive abilities, or just a forgetful person going to the grocery store without a shopping list.

13. Final comments

The comprehension of metamemory judgments in general, and JOLs in particular, is of great importance to the field of Psychology and even other fields of knowledge, such as Education. To clarify how people monitor cognitive processes is relevant in different contexts, including their daily lives. Examples such as knowing if one is ready to take an exam or should prepare better, or if an alarm is needed to remind them of an appointment, or even if they will remember every item and do not need to bring the shopping list to the grocery store underline the importance of metamemory. This dissertation is but a small contribution to the field of metamemory research that shed light into the comprehension of the basic processes underlying the complex mechanism of making JOLs. We hope that the importance of research focused on basic processes, such as this dissertation, is recognized as crucial for applied research to thrive.

References

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. https://doi.org/10.1016/0749-596X(89)90016-8

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31*(2), 297–305. https://doi.org/10.3758/BF03194388

Blake, A. B., & Castel, A. D. (2018). On belief and fluency in the construction of judgments of learning: Assessing and altering the direct effects of belief. *Acta Psychologica, 186*, 27–38. https://doi.org/10.1016/j.actpsy.2018.04.004

Bröder, A., & Undorf, M. (2019). Metamemory viewed through the judgment lens. *Acta Psychologica, 197*, 153–165. https://doi.org/10.1016/j.actpsy.2019.04.011

Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago University Press.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition, 20*(3), 231–243. https://doi.org/10.3758/BF03199660

Dunlosky, J., Mueller, M. L., & Tauber, S. K. (2015). The contribution of processing fluency (and beliefs) to people's judgments of learning. In D. S. Lindsay (Ed.) & C. M. Kelley (Trans.) & A. P. Yonelinas, H. L. Roediger III (Eds.), *Psychology Press festschrift series. Remembering: Attributions, processes, and control in human memory: Essays in honor of Larry Jacoby* (pp. 46–64). Psychology Press. https://psycnet.apa.org/record/2014-56350-004

Fiacconi, C. M., & Dollois, M. A. (2020). Does word frequency influence judgments of learning (JOLs)? A meta-analytic review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale.* Advance online publication. https://doi.org/10.1037/cep0000206

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(5), 680–693. https://doi.org/10.1037/xlm0000332

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology, 61*(1), 23–29. https://doi.org/10.1037/h0040561

Hall, J. F. (1954). Learning as a function of word-frequency. *The American Journal of Psychology, 67*(1), 138–140. https://doi.org/10.2307/1418080

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2017). The influences of valence and arousal on judgments of learning and on recall. *Memory & Cognition, 45*(1), 121–136. https://doi.org/10.3758/S13421-016-0646-3

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). https://doi.org/10.1371/journal.pone.0142351

Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2016). The effect of word frequency on judgments of learning: Contributions of beliefs and processing fluency. *Frontiers in Psychology, 6*:1995. https://doi.org/10.3389/fpsyg.2015.01995

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. https://doi.org/10.1037/0096-3445.126.4.349

Lohnas, L. J., & Kahana, M. J. (2013). Parametric effects of word frequency in memory for mixed frequency lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1943–1946. https://doi.org/10.1037/a0033669

Luna, K., Nogueira, M., & Albuquerque, P. B. (2019). Words in larger font are perceived as more important: Explaining the belief that font size affects memory. *Memory, 27*(4)*,* 555–560. https://doi.org/10.1080/09658211.2018.1529797

Malmberg, K. J., & Murnane, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 616–630. https://doi.org/10.1037/0278-7393.28.4.616

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? *Quarterly Journal of Experimental Psychology, 69*(4), 781–799. https://doi.org/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. https://doi.org/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. https://doi.org/10.3758/s13423-012-0343-6

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. https://doi.org/10.1016/j.jml.2016.10.008

Ning, X., Li, C., & Yang, J. (2018). Word familiarity modulated the effects of category familiarity on memory performance. *Frontiers in Psychology, 9:*1429. https://doi.org/10.3389/fpsyg.2018.01429

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology, 20*(2), 139–156. https://doi.org/10.1002/acp.1178

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review, 127*(1), 1–46. http://doi.org/10.1037/rev0000161

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky, & S. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 90–117). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.4

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. https://doi.org/10.1037/a0013684

Sumby, W. H. (1963). Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior, 1*(6), 443–450. https://doi.org/10.1016/S0022-5371(63)80030-4

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review, 22*(4), 1014–1019. https://doi.org/10.3758/s13423-014-0768-1

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. https://doi.org/10.1016/j.jml.2017.03.002

Tatz, J. R., & Peynircioğlu, Z. F. (2020). Judgments of learning in context: Backgrounds can both reduce and produce metamemory illusions. *Memory & Cognition, 48*(4), 581–595. https://doi.org/10.3758/s13421-019-00991-9

Tullis, J. G., & Benjamin, A. S. (2012). The effectiveness of updating metacognitive knowledge in the elderly: Evidence from metamnemonic judgments of word frequency. *Psychology and Aging, 27*(3), 683–690. https://doi.org/10.1037/a0025838

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. https://doi.org/10.1037/xlm0000571

Undorf, M., Söllner, A. & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory & Cognition, 46*(4), 507–519. https://doi.org/10.3758/s13421-017-0780-6

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. https://doi.org/10.1016/j.jml.2016.07.003

Vogel, T., Silva, R. R., Thomas, A., & Wänke, M. (2020). Truth is in the mind, but beauty is in the eye: Fluency effects are moderated by a match between fluency source and judgment dimension. *Journal of Experimental Psychology: General, 149*(8), 1587–1596. https://doi.org/10.1037/xge0000731

Yang, C., Huang, T. S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language, 99*, 99–110. https://doi.org/10.1016/j.jml.2017.11.005

APPENDIXES

Appendix A. Instructions and Stimuli Used in the Pilot Study

A.1. Instructions Used in the Pilot Study

Table A.1.1

Instructions for the Pilot Study.

Experiment phase	**Instruction**
General instructions	In this experiment you will see several words and for each you must judge the probability of remembering it later (on a scale of 0 to 100), using the keyboard and pressing the key <Send answer>. Later, you will have to recall the words you have seen. When you are ready to start the experiment, press the <spacebar>.
JOL (after each word)	What is the probability of later remembering the word you just saw? 0: I will not remember it; 100: I am sure I will remember the word
Filler task	In this phase of the experiment, you will have to solve some math calculations. You should insert the correct answer and press <Send answer>. To start press <spacebar>.
Recall task	Now, please write all the words from the list you have just studied in the paper sheet you have received. (When time ended): Recall time has ended! Before continuing, please call the researcher. (When ready): Press <spacebar> to continue.
Final question (after 4 study-test trials)	There was a characteristic that distinguished the words in the word lists you have studied. Can you identify that characteristic for each list? If so, please specify for each list (answer in the respective field). You can press <ENTER> to change the line in the response field. (When finished): The experience is over. Thank you very much for participating.

Note: If you would like the original instructions (in European Portuguese), please contact the author.

A.2. Stimuli Used in the Pilot Study

For all tables presented here, objective frequency is presented in frequency per million (fpm); Sub. Freq = Subjective frequency; Freq (Subtlex) = Objective frequency in fpm, taken from SUBTLEX-PT (Soares et al., 2015); values of objective frequency, letters, syllables, and Zipf were taken from P-PAL (Soares et al., 2018); values of concreteness and subjective frequency (ranging 1–7) were taken from Minho Word Pool (Soares et al., 2017).

Table A.2.1

Stimuli Used in the Very Wide Condition

Frequency	Word	Obj. Frequency	Concreteness	Letters	Syllables	Sub. Freq	Freq (Subtlex)	Zipf (P-Pal)
Low	falcão	5.24	6.40	6	2	3.09	7.59	3.72
	peste	5.55	3.51	5	2	3.53	8.95	3.74
	devoção	5.69	2.63	7	3	3.36	4.04	3.76
	granito	5.85	6.21	7	3	2.88	1.72	3.77
	anel	6.22	6.63	4	2	4.30	49.99	3.79
	muralha	6.46	6.34	7	3	3.36	2.88	3.81
	célula	6.64	5.21	6	3	3.61	7.46	3.82
	gravata	6.82	6.74	7	3	4.05	15.39	3.83
	bênção	7.09	2.67	6	2	3.02	8.25	3.85
	colheita	8.79	4.77	8	3	3.24	8.97	3.94
	nuvem	7.45	5.71	5	2	5.19	10.45	3.87
	réplica	7.96	4.07	7	3	3.72	2.91	3.90
	tela	8.37	6.16	4	2	3.32	12.36	3.92
	vedeta	8.60	4.19	6	3	3.32	0.77	3.93
	pimenta	8.63	6.20	7	3	4.65	6.19	3.94
	lagoa	8.80	6.30	5	3	3.27	2.99	3.94
	orador	7.86	4.19	6	3	3.11	2.53	3.90
	ombro	9.70	6.69	5	2	4.84	22.35	3.99
	carbono	9.79	4.39	7	3	2.89	11.09	3.99
	engenho	9.92	4.00	7	3	2.95	3.00	4.00
Mean (*SD*)		**7.57 (1.48)**	**5.15 (1.36)**	**6.10 (1.12)**	**2.65 (0.49)**	**3.58 (0.68)**	**9.49 (10.90)**	**3.87 (0.09)**
High	negócio	97.07	4.00	7	3	5.22	132.71	4.99
	distrito	97.49	4.18	8	3	4.40	12.82	4.99
	carro	99.33	6.55	5	2	6.59	516.27	5.00
	estação	101.64	5.49	7	3	4.74	62.19	5.01
	senhor	102.47	5.22	6	2	6.36	937.03	5.01
	redução	106.42	3.38	7	3	5.18	3.76	5.03
	canal	107.09	4.61	5	2	5.54	34.62	5.03
	versão	107.09	3.36	6	2	4.53	27.29	5.03
	carreira	109.16	3.85	8	3	4.60	51.14	5.04
	indústria	110.24	5.25	9	3	4.21	15.60	5.04
	idade	111.50	4.08	5	3	5.76	107.40	5.05
	agência	115.23	4.29	7	3	4.02	29.75	5.06
	passagem	116.79	4.25	8	3	4.98	34.40	5.07
	conselho	121.49	3.54	8	3	5.07	73.49	5.08
	pessoa	121.97	6.02	6	3	6.69	355.95	5.09
	índice	123.31	4.39	6	3	4.60	3.06	5.09
	espécie	123.80	3.95	7	3	4.21	73.19	5.09
	sequência	125.87	4.03	9	3	4.98	11.46	5.10
	página	127.13	6.05	6	3	6.27	31.43	5.10
	efeito	127.68	3.33	6	3	4.70	41.34	5.11
Mean (*SD*)		**112.64 (10.34)**	**4.49 (0.96)**	**6.80 (1.24)**	**2.80 (0.41)**	**5.13 (0.82)**	**127.74 (17.83)**	**5.05 (0.04)**

Table A.2.2

Stimuli Used in the Wide Condition

Frequency	Word	Obj. Frequency	Concreteness	Letters	Syllables	Sub. Freq	Freq (Subtlex)	Zipf (P-Pal)
Low	táxi	10.11	6.75	4	2	3.84	42.80	4.00
	hino	10.27	4.66	4	2	3.22	3.55	4.01
	xadrez	10.34	6.36	6	2	3.13	10.01	4.01
	garrafa	10.43	6.65	7	3	6.22	42.07	4.02
	rato	10.59	6.64	4	2	4.82	45.22	4.02
	camada	10.85	4.24	6	3	3.93	7.56	4.04
	fôlego	10.97	3.68	6	3	3.64	7.31	4.04
	gestor	16.06	5.32	6	2	4.53	1.20	4.21
	estátua	11.60	6.07	7	3	3.75	12.09	4.06
	parcela	11.64	4.02	7	3	2.77	1.24	4.07
	clareza	11.84	2.55	7	3	4.55	4.24	4.07
	óculos	12.09	6.63	6	3	6.33	29.15	4.08
	carrinha	12.63	6.58	8	3	4.72	44.35	4.10
	ofício	13.73	3.78	6	3	3.38	3.93	4.14
	brisa	14.70	4.26	5	2	3.64	7.11	4.17
	audição	15.09	4.31	7	3	4.60	12.79	4.18
	correio	17.86	5.96	7	3	4.80	28.43	4.25
	bilhete	18.50	6.35	7	3	4.93	37.99	4.27
	coesão	18.97	3.11	6	3	3.67	0.62	4.28
	paragem	19.47	5.05	7	3	5.40	14.98	4.29
Mean (*SD*)		**13.39 (3.21)**	**5.15 (1.35)**	**6.15 (1.14)**	**2.70 (0.47)**	**4.29 (0.97)**	**17.83 (16.54)**	**4.12 (0.10)**
High	notícia	80.45	4.78	7	3	6.63	44.98	4.91
	tendência	84.90	3.17	9	3	4.74	7.19	4.93
	década	81.01	4.19	6	3	3.60	8.98	4.91
	acto	81.74	3.88	4	2	5.00	25.57	4.91
	volume	82.23	4.16	6	3	4.91	10.46	4.92
	carácter	82.33	2.98	8	3	4.54	8.57	4.92
	dimensão	82.60	2.81	8	3	4.45	8.45	4.92
	moeda	83.32	6.65	5	3	6.48	20.84	4.92
	parque	85.20	5.60	6	2	5.05	60.95	4.93
	domínio	85.56	2.82	7	3	4.30	8.36	4.93
	língua	86.24	5.66	6	2	5.22	45.46	4.94
	sinal	86.53	4.48	5	2	5.36	130.65	4.94
	missão	87.18	3.63	6	2	4.56	107.58	4.94
	autarca	84.77	4.90	7	3	3.67	0.04	4.93
	palco	87.91	6.26	5	2	3.97	28.31	4.94
	ponte	89.42	6.46	5	2	4.70	64.59	4.95
	treinador	91.86	5.84	9	3	4.57	40.19	4.96
	mesa	92.21	6.66	4	2	6.41	127.83	4.96
	intenção	93.08	2.73	8	3	5.38	31.90	4.97
	comércio	94.81	4.58	8	3	4.62	10.51	4.98
Mean (*SD*)		**86.17 (4.24)**	**4.61 (1.35)**	**6.45 (1.54)**	**2.60 (0.50)**	**4.91 (0.84)**	**39.57 (40.22)**	**4.93 (0.02)**

Table A.2.3

Stimuli Used in the Narrow Condition

Frequency	Word	Obj. Frequency	Concreteness	Letters	Syllables	Sub. Freq	Freq (Subtlex)	Zipf (P-Pal)
Low	cartaz	20.10	5.96	6	2	4.46	6.74	4.30
	buraco	20.72	5.74	6	3	4.93	64.04	4.32
	cavalo	20.86	6.72	6	3	3.76	73.65	4.32
	hábito	20.95	3.76	6	3	5.39	12.95	4.32
	carteira	21.30	6.73	8	3	6.18	33.80	4.33
	montanha	21.45	6.40	8	3	3.96	39.16	4.33
	vocação	21.87	2.92	7	3	4.12	2.88	4.34
	cadeira	21.87	6.71	7	3	6.20	52.00	4.34
	cedência	22.09	2.80	8	3	3.83	0.31	4.34
	estúdio	22.41	5.76	7	3	3.68	21.70	4.35
	noção	23.45	3.03	5	2	5.59	14.68	4.37
	floresta	24.09	6.21	8	3	4.25	46.23	4.38
	parede	25.06	6.56	6	3	5.34	57.52	4.40
	janela	25.38	6.61	6	3	6.35	71.96	4.40
	virtude	26.31	2.50	7	3	4.04	5.22	4.42
	círculo	28.16	5.85	7	3	4.42	19.35	4.45
	fusão	28.28	3.34	5	2	3.32	10.37	4.45
	cliente	28.69	5.05	7	3	5.35	72.74	4.46
	antena	29.62	6.22	6	3	4.18	6.61	4.47
	postura	29.73	3.82	7	3	4.84	5.63	4.47
Mean (*SD*)		**24.12 (3.28)**	**5.14 (1.56)**	**6.65 (0.93)**	**2.85 (0.37)**	**4.71 (0.91)**	**30.88 (26.40)**	**4.38 (0.06)**
High	fábrica	65.39	6.10	7	3	4.69	32.53	4.82
	aspecto	66.28	3.47	7	3	5.12	43.98	4.82
	estilo	66.62	3.44	6	3	5.14	59.02	4.82
	distância	67.41	3.38	9	3	5.39	52.28	4.83
	feira	67.71	5.75	5	2	4.57	19.64	4.83
	certeza	68.87	2.91	7	3	5.80	559.67	4.84
	excepção	69.22	3.49	8	3	5.19	9.88	4.84
	criança	69.92	6.20	7	3	6.04	140.90	4.84
	silêncio	70.64	3.03	8	3	5.71	61.47	4.85
	juiz	71.18	5.85	4	2	3.90	50.68	4.85
	ouro	72.37	5.86	4	2	4.07	83.38	4.86
	coração	72.97	5.96	7	3	5.78	216.60	4.86
	percurso	73.94	4.46	8	3	4.74	6.54	4.87
	banda	75.35	5.35	5	2	4.84	58.69	4.88
	produto	75.85	4.77	7	3	5.58	18.61	4.88
	divisão	77.12	4.26	7	3	4.89	23.84	4.89
	cenário	78.03	4.02	7	3	4.41	14.18	4.89
	ciclo	79.04	3.47	5	2	4.42	10.47	4.90
	margem	79.04	4.47	6	2	4.04	11.27	4.90
	protecção	79.91	3.45	9	3	4.81	28.95	4.90
Mean (*SD*)		**72.34 (4.71)**	**4.49 (1.15)**	**6.65 (1.46)**	**2.70 (0.47)**	**4.96 (0.63)**	**75.13 (124.77)**	**4.86 (0.03)**

Table A.2.4

Stimuli Used in the Very Narrow Condition

Frequency	Word	Objective Frequency	Concreteness	Letters	Syllables	Sub. Freq	Freq (Subtlex)	Zipf (P-Pal)
Low	dúzia	30.05	4.98	5	2	4.54	16.84	4.48
	planeta	30.29	6.40	7	3	4.09	106.29	4.48
	prédio	30.33	6.49	6	2	5.40	35.72	4.48
	ficção	31.57	3.19	6	2	4.09	10.42	4.50
	onda	32.05	5.59	4	2	4.21	27.44	4.51
	sensação	32.08	2.61	8	3	5.07	37.05	4.51
	paisagem	32.25	4.81	8	3	4.82	6.83	4.51
	pacote	32.32	5.61	6	3	5.51	19.84	4.51
	rapaz	33.76	6.47	5	2	5.88	264.63	4.53
	império	37.94	3.74	7	3	2.91	17.73	4.58
	método	34.44	3.21	6	3	4.82	13.96	4.54
	profissão	36.03	4.40	9	3	5.18	13.84	4.56
	peixe	39.19	6.37	5	2	5.59	54.23	4.59
	gesto	36.95	4.31	5	2	4.65	11.66	4.57
	duração	37.36	3.41	7	3	5.33	3.10	4.57
	prestação	37.75	3.38	9	3	4.10	1.17	4.58
	limpeza	38.79	4.08	7	3	5.76	21.43	4.59
	rosto	39.00	6.25	5	2	5.12	54.42	4.59
	expansão	36.38	3.02	8	3	3.70	4.68	4.56
	colecção	38.87	4.91	8	3	4.53	12.37	4.59
Mean (*SD*)		**34.87 (3.29)**	**4.66 (1.31)**	**6.55 (1.47)**	**2.60 (0.50)**	**4.77 (0.76)**	**36.68 (58.95)**	**4.54 (0.04)**
High	convite	50.26	4.97	7	3	4.35	20.05	4.70
	contexto	50.72	3.17	8	3	5.09	6.33	4.71
	caixa	52.28	6.58	5	2	5.45	100.96	4.72
	véspera	52.32	3.48	7	3	4.26	9.77	4.72
	actor	52.67	5.91	5	2	4.98	18.96	4.72
	jardim	53.00	6.34	6	2	5.00	38.90	4.72
	reacção	53.02	3.18	7	3	5.00	17.51	4.72
	avião	53.65	6.84	5	3	4.29	114.47	4.73
	geração	54.92	3.43	7	3	4.36	16.98	4.74
	museu	55.30	6.30	5	2	3.45	26.66	4.74
	anúncio	56.82	5.00	7	3	5.38	20.57	4.75
	dólar	57.17	6.21	5	2	4.14	18.46	4.76
	escritor	57.68	5.55	8	3	4.64	22.62	4.76
	etapa	59.82	3.12	5	3	4.58	5.74	4.78
	visão	60.49	4.68	5	2	5.52	50.92	4.78
	álbum	60.75	5.49	5	2	3.91	10.82	4.78
	adesão	60.88	3.60	6	3	3.96	0.19	4.78
	bairro	62.98	5.52	6	2	4.34	32.41	4.80
	turismo	64.10	4.40	7	3	4.33	3.51	4.81
	género	64.31	3.07	6	3	5.21	34.04	4.81
Mean (*SD*)		**56.66 (4.50)**	**4.84 (1.32)**	**6.10 (1.07)**	**2.60 (0.50)**	**4.61 (0.57)**	**28.49 (29.93)**	**4.75 (0.03)**

Appendix B. Instructions and Stimuli Used in Study 1

B.1. Instructions Used in Study 1

Table B.1.1

Instructions for Experiments in Study 1.

<table>
<tr><th colspan="2">Experiment phase</th><th>Instruction</th></tr>
<tr><td colspan="2">General instructions</td><td>Please, read these instructions carefully.
In this experiment you will see several words and for each you must judge the probability of remembering it later (on a scale of 0 to 100), using the keyboard and pressing the button CONTINUE.
Later, you will have to recall the words you have seen.
If you have any question, please ask now.
When you are ready to start the experiment, press the button START.</td></tr>
<tr><td colspan="2">JOL (after each word)</td><td>What is the probability of remembering the presented word? (Insert a number between 0 and 100).</td></tr>
<tr><td colspan="2">Filler task</td><td>In this phase of the experiment, you will have to solve math calculations. You have 3 minutes to finish this task.
In the right column, write the result in front of the calculation from the left column (one result per line).
When you are ready, press the button START.</td></tr>
<tr><td colspan="2">Recall task</td><td>Please, read these instructions carefully.
You will have 4 minutes to write down all the presented words that you remember.
Write one word per line.
If you have any question, please ask now.
When you are ready to start the experiment, press the button START.</td></tr>
<tr><td rowspan="3">Final question</td><td>Experiment 1</td><td>There was a characteristic that distinguished the words from the list you have studied. Can you identify that characteristic? (Yes/No)
If “Yes” > Please say what is that characteristic.

(At end): The experience is over. Thank you for participating!</td></tr>
<tr><td>Experiment 2 (Asked before recall)</td><td>Question 1: Your judgments of learning for the words (these evaluations you made) varied throughout the experiment. What did you base your judgements on?
Question 2: There is a characteristic of the words from which you can divide the list in two groups. Can you tell which characteristic that is?
Question 3: In the list you studied, there were high- and low-frequency words (common and rare words). Did you base your judgments on the frequency of the words?</td></tr>
<tr><td>Experiment 4 (Either A or B was asked)</td><td>Question A: During the first phase of the experiment, what did you base your judgments on for each word?
Question B: During the first phase of the experiment (during judgments), did you notice that words differed in frequency? Did you use frequency as a cue to make your judgments?</td></tr>
<tr><td colspan="2">Lexical decision task (Experiments 3 and 4)</td><td>Please, read these instructions carefully.
In this (phase of the) experiment you must decide as quickly as possible whether the presented item is a word or not. Se the item is a word you must press the key (M/Z) or if it is not a word, you must press the key (Z/M).
Before the experiment starts, you will have a training session with 4 items.
If you have any question, please ask now.
When you are ready to start the experiment, press the button START.

(After training, a reminder of which key (M/Z) corresponded to word/nonword appeared on the screen until the participant pressed the START button.)</td></tr>
</table>

Note: If you would like the original instructions (in European Portuguese), please contact the author.

B.2. Stimuli Used in Study 1

Table B.2.1

Words Used in Study 1

Frequency	Word	Objective Frequency	Concreteness	Letters	Syllables	Sub. Freq	Freq (Subtlex)	Zipf (P-Pal)
Low	térmite	0.01	3.49	7	3	1.51	0.13	0.99
	begónia	0.05	4.41	7	3	1.84	0.03	1.73
	espora	0.09	3.94	6	3	1.96	0.19	1.97
	fuzil	0.12	4.43	5	2	1.69	0.91	2.07
	esturjão	0.17	3.95	8	3	1.53	0.22	2.23
	chicória	0.19	3.96	8	3	2.06	0.31	2.28
	neutrão	0.22	3.64	7	2	2.35	0.22	2.33
	goraz	0.23	3.41	5	2	1.42	0.03	2.36
	galé	0.24	3.83	4	2	1.48	0.41	2.39
	condor	0.27	3.93	6	2	1.82	1.06	2.44
	arenque	0.28	4.27	7	3	1.57	1.15	2.45
	peçonha	0.30	2.67	7	3	1.69	0.04	2.48
	groselha	0.15	6.16	8	3	3.32	0.55	2.17
	tatu	0.07	5.00	4	2	2.38	0.44	1.87
	quilate	0.56	4.35	7	3	2.17	0.36	2.75
	coador	0.05	6.09	6	3	3.22	0.27	1.73
	sebo	0.69	4.08	4	2	2.15	1.76	2.84
	moreia	0.15	4.82	6	3	1.85	0.06	2.17
	provação	0.81	2.64	8	3	2.84	1.18	2.91
	desertor	0.81	2.96	8	3	2.20	2.22	2.91
Mean (SD)		**0.27 (0.25)**	**4.10 (0.93)**	**6.40 (1.39)**	**2.65 (0.49)**	**2.05 (0.55)**	**0.58 (0.61)**	**2.25 (0.46)**
High	solução	160.64	3.61	7	3	5.56	35.56	5.21
	jornal	178.13	6.64	6	2	5.93	48.28	5.25
	sessão	174.84	3.91	6	2	4.74	22.43	5.24
	século	186.56	3.97	6	3	4.00	31.95	5.27
	produção	212.35	3.67	8	3	4.75	20.73	5.33
	presença	213.75	3.65	8	3	5.22	36.97	5.33
	vitória	213.82	3.80	7	3	4.91	34.20	5.33
	serviço	222.53	3.74	7	3	5.32	107.08	5.35
	comissão	226.13	4.12	8	3	4.13	23.14	5.35
	direcção	264.86	3.98	8	3	5.16	54.10	5.42
	posição	288.09	3.89	7	3	5.38	93.82	5.46
	decisão	299.04	3.66	7	3	5.60	83.03	5.48
	zona	300.96	3.81	4	2	5.04	72.20	5.48
	nome	313.50	4.27	4	2	6.22	579.20	5.50
	história	323.94	3.64	8	3	4.91	309.72	5.51
	frente	343.08	3.91	6	2	5.98	373.43	5.54
	questão	370.67	3.69	7	2	5.68	122.83	5.57
	lugar	450.07	4.15	5	2	5.95	613.73	5.65
	lado	567.71	4.25	4	2	6.07	503.12	5.75
	tempo	757.54	3.63	5	2	6.61	1317.77	5.88
Mean (SD)		**303.41 (146.74)**	**4.00 (0.66)**	**6.40 (1.39)**	**2.55 (0.51)**	**5.36 (0.68)**	**224.16 (323.73)**	**5.44 (0.18)**

Notes: Objective frequency is presented in frequency per million (fpm); Sub. Freq = Subjective frequency; Freq (Subtlex) = Objective frequency in fpm, taken from SUBTLEX-PT (Soares et al., 2015); values of objective frequency, letters, syllables, and Zipf were taken from P-PAL (Soares et al., 2018); values of concreteness and subjective frequency (ranging 1–7) were taken from Minho Word Pool (Soares et al., 2017).

Table B.2.2

Nonwords Used in Study 1

retitém	arquene	lusoção	pioçosa	urtejãos	tilaque	dopurção	hastírio
gonébia	cepanho	jalnor	cidesão	chacírio	doraco	serpença	trenfe
serapo	gresalho	sasões	onaz	tenurão	bose	rotívia	toasque
lufiz	tutá	céluso	onem	ragoz	armeio	versiço	gurla
gelá	vorpação	miscosão	dola	dronco	torseder	cerdição	mopte

B.3. Supplemental Materials for Study 1

These are the Supplemental Materials for Study 1 (available at https://osf.io/uyh9z), which focus on gamma correlations. There is some redundancy since we already presented resolution measures in Chapter 10, but we present these here for completeness.

B.3.1. Relative Accuracy (Gamma Correlations)

Experiment 1. A one-sample t-test was conducted, showing that the relationship between JOLs and recall (γ = .48, *SD* = .19) was significantly different from zero, *t*(35) = 14.99, *p* < .001, Cohen's *d* = 2.50. A paired samples t-test showed that participants' JOLs were similarly accurate in predicting correct recall of low-frequency words (γ = .32, *SD* = .52) and high-frequency words (γ = .35, *SD* = .36), *t*(35) = .09, *p* = .74, Cohen's *d* = .06.

Experiment 2. A one-sample t-test was conducted, showing that the relationship between JOLs and recall (γ = .44, *SD* = .39) was significantly different from zero, *t*(29) = 6.29, *p* < .001, Cohen's *d* = 1.15. A paired samples t-test showed that participants' JOLs were similarly accurate in predicting correct recall of low-frequency words (γ = .34, *SD* = .56) and high-frequency words (γ = .40, *SD* = .47), *t*(27) = .50, *p* = .62, Cohen's *d* = .10.

Experiment 4. A one-sample t-test was conducted, showing that the relationship between JOLs and recall (γ = .36, *SD* = .39) was significantly different from zero, *t*(39) = 5.83, *p* < .001, Cohen's *d* = .92. A paired samples t-test showed that participants' JOLs were similarly accurate in predicting correct recall of low-frequency words (γ = .30, *SD* = .55) and high-frequency words (γ = .33, *SD* = .41), *t*(37) = .33, *p* = .74, Cohen's *d* = .05.

Appendix C. Instructions, Stimuli and Supplemental Materials of Study 2

C.1. Instructions Used in Study 2

Table C .1.1

Instructions for Experiments in Study 2.

Experiment phase		Instruction
Initial instruction	Experiment 2	There are more than 200,000 words in the Portuguese language. By analyzing written texts and oral speech entries, it is possible to verify that some words appear very frequently, while others only appear sometimes. High-frequency words are those that appear many times (more than 100 times per million words), for example, 'house'. Low-frequency words are those which rarely appear (less than 1 time per million words), for example, 'reef'. Curiously, our memory is very good for low-frequency words because, since they are so rare, they capture our attention and become very salient among others. This could be a reason why, when we learn an uncommon word, we start seeing it everywhere, because it has become more salient.
	Experiment 3 (better-for-high)	There are more than 200,000 words in the Portuguese language. By analyzing written texts and oral speech entries, it is possible to verify that some words appear very frequently, while others only appear sometimes. High-frequency words are those that appear many times (more than 100 times per million words), for example, 'house'. Low-frequency words are those which rarely appear (less than 1 time per million words), for example, 'reef'. Curiously, our memory is very good for high-frequency words because, since they are so common, we find them more in our daily lives. This could be a reason why, when we learn a common word, we start seeing it everywhere because it has become familiar.
	Experiment 3 (better-for-low)	There are more than 200,000 words in the Portuguese language. By analyzing written texts and oral speech entries, it is possible to verify that some words appear very frequently, while others only appear sometimes. High-frequency words are those that appear many times (more than 100 times per million words), for example, 'house'. Low-frequency words are those which rarely appear (less than 1 time per million words), for example, 'reef'. Curiously, our memory is very good for low-frequency words because, since they are so rare, they capture our attention and become very salient among others. This could be a reason why, when we learn an uncommon word, we start seeing it everywhere, because it has become more salient.
General instructions		Please, read these instructions carefully. In this experiment you will study a list of words, one at a time. You will see high- and low-frequency words. Please be attentive because you will have to remember the words later. Before (*after* in Experiment 1b) you see each word, a prompt will appear on the screen informing you if the word is of high or low frequency. At that point you must indicate on a scale from 0 to 100 how confident you are of later remembering that word. If you are certain you will not remember the word, select 0. Se you are certain that you will remember it, then select 100. After each word, you will have to do that same assessment. If you have any question, please ask now. When you are ready to start the experiment, press the button START.
Pre-study JOL (before each word)		What is the probability of remembering the word? (Mouse-clicking scale: 0; 20; 40; 60; 80; 100)
Immediate JOL (after each word)		What is the probability of remembering the presented word? (Mouse-clicking scale: 0; 20; 40; 60; 80; 100)
Filler task		In this phase of the experiment, you will have to solve math calculations. You have 3 minutes to finish this task. In the right column, write the result in front of the calculation from the left column (one result per line). When you are ready, press the button START.
Recall task		Please, read these instructions carefully. You will have 5 minutes to write down all the presented words that you remember. Write one word per line. If you have any question, please ask now. When you are ready to start the experiment, press the button START.

Note: If you would like the original instructions (in European Portuguese), please contact the author.

C.2. Stimuli Used in Study 2

Table C.2.1

Words Used in Study 2

Frequency	Word	Objective Frequency	Concreteness	Letters	Syllables	Sub. Freq	Freq (Subtlex)	Zipf (P-Pal)
Low	térmite	0.01	3.49	7	3	1.51	0.13	0.99
	begónia	0.05	4.41	7	3	1.84	0.03	1.73
	espora	0.09	3.94	6	3	1.96	0.19	1.97
	fuzil	0.12	4.43	5	2	1.69	0.91	2.07
	esturjão	0.17	3.95	8	3	1.53	0.22	2.23
	chicória	0.19	3.96	8	3	2.06	0.31	2.28
	neutrão	0.22	3.64	7	2	2.35	0.22	2.33
	goraz	0.23	3.41	5	2	1.42	0.03	2.36
	galé	0.24	3.83	4	2	1.48	0.41	2.39
	condor	0.27	3.93	6	2	1.82	1.06	2.44
	arenque	0.28	4.27	7	3	1.57	1.15	2.45
	peçonha	0.30	2.67	7	3	1.69	0.04	2.48
	impostor	0.35	3.27	8	3	3.33	3.72	2.54
	páscoa	0.38	4.22	6	2	2.73	0.04	2.58
	quilate	0.56	4.35	7	3	2.17	0.36	2.75
	faia	0.63	4.21	4	2	1.67	5.18	2.80
	sebo	0.69	4.08	4	2	2.15	1.76	2.84
	toxina	0.69	4.16	6	3	3.09	3.33	2.84
	provação	0.81	2.64	8	3	2.84	1.18	2.91
	desertor	0.81	2.96	8	3	2.20	2.22	2.91
Mean (SD)		**0.36 (0.25)**	**3.79 (0.55)**	**6.40 (1.39)**	**2.60 (0.50)**	**2.05 (0.56)**	**1.12 (1.45)**	**2.39 (0.46)**
High	solução	160.64	3.61	7	3	5.56	35.56	5.21
	gestão	168.00	4.12	6	2	5.11	4.38	5.23
	sessão	174.84	3.91	6	2	4.74	22.43	5.24
	século	186.56	3.97	6	3	4.00	31.95	5.27
	produção	212.35	3.67	8	3	4.75	20.73	5.33
	presença	213.75	3.65	8	3	5.22	36.97	5.33
	vitória	213.82	3.80	7	3	4.91	34.20	5.33
	serviço	222.53	3.74	7	3	5.32	107.08	5.35
	comissão	226.13	4.12	8	3	4.13	23.14	5.35
	direcção	264.86	3.98	8	3	5.16	54.10	5.42
	posição	288.09	3.89	7	3	5.38	93.82	5.46
	decisão	299.04	3.66	7	3	5.60	83.03	5.48
	zona	300.96	3.81	4	2	5.04	72.20	5.48
	nome	313.50	4.27	4	2	6.22	579.20	5.50
	história	323.94	3.64	8	3	4.91	309.72	5.51
	frente	343.08	3.91	6	2	5.98	373.43	5.54
	questão	370.67	3.69	7	2	5.68	122.83	5.57
	lugar	450.07	4.15	5	2	5.95	613.73	5.65
	lado	567.71	4.25	4	2	6.07	503.12	5.75
	tempo	757.54	3.63	5	2	6.61	1317.77	5.88
Mean (SD)		**302.90 (147.21)**	**3.87 (0.22)**	**6.40 (1.39)**	**2.55 (0.51)**	**5.32 (0.67)**	**221.97 (325.13)**	**5.44 (0.18)**

Notes: Objective frequency is presented in frequency per million (fpm); Sub. Freq = Subjective frequency; Freq (Subtlex) = Objective frequency in fpm, taken from SUBTLEX-PT (Soares et al., 2015); values of objective frequency, letters, syllables, and Zipf were taken from P-PAL (Soares et al., 2018); values of concreteness and subjective frequency (ranging 1–7) were taken from Minho Word Pool (Soares et al., 2017).

C.3. Supplemental Materials of Study 2

These are the analyses provided in the Supplemental Materials of the published article, which were peer reviewed and are available at: https://osf.io/8qvay

C.3.1. Pre-study JOLs Across the Study Phase

In our experiments, participants could learn that some of the prompts are unreliable since they were incongruent with actual word frequency in half of the trials. This could explain why they disregarded the prompt when making immediate JOLs. If participants notice that the prompts are unreliable, the effect of prompts on pre-study JOLs might decline across the study phase. We graphed the mean pre-study JOLs given to prompts of high frequency (vs. low frequency) attending to the relative serial position (i.e., “position 1” is the first prompt presented of that type). See Figure C.3.1 below.

To explore if the effects on pre-study JOLs decrease from the beginning to the end of the study phase, we calculated the difference between pre-study JOLs given to the first and last three items for prompts of high and for prompts of low frequency. Overall, there is a small but not significant decrease over time (see Table C.3.1 below with the main descriptive and inferential statistics for each kind of prompt). More importantly, the difference between high- and low-frequency prompts for pre-study JOLs is the same at the beginning and at the end of the study phase: the difference is relatively constant in Experiments 1a and 3 (better-for-high condition), and it disappears when a counter-belief is introduced (Experiments 2 and 3 (better-for-low condition). Table C.3.2. shows the effect size for pre-study JOLs at the beginning and at the end of the study-phase. With the exception of Experiment 2, the effect on pre-study JOLs is observable across the study phase, suggesting that participants do not disregard the information given by the prompt. Hence, the effect on immediate JOLs is not caused by learning that prompts are unreliable.

Figure C.3.1

Pre-study JOLs for Prompts of High Frequency (HF) and Low Frequency (LF) Across the Study Phase for Experiments 1a, 2 and 3.

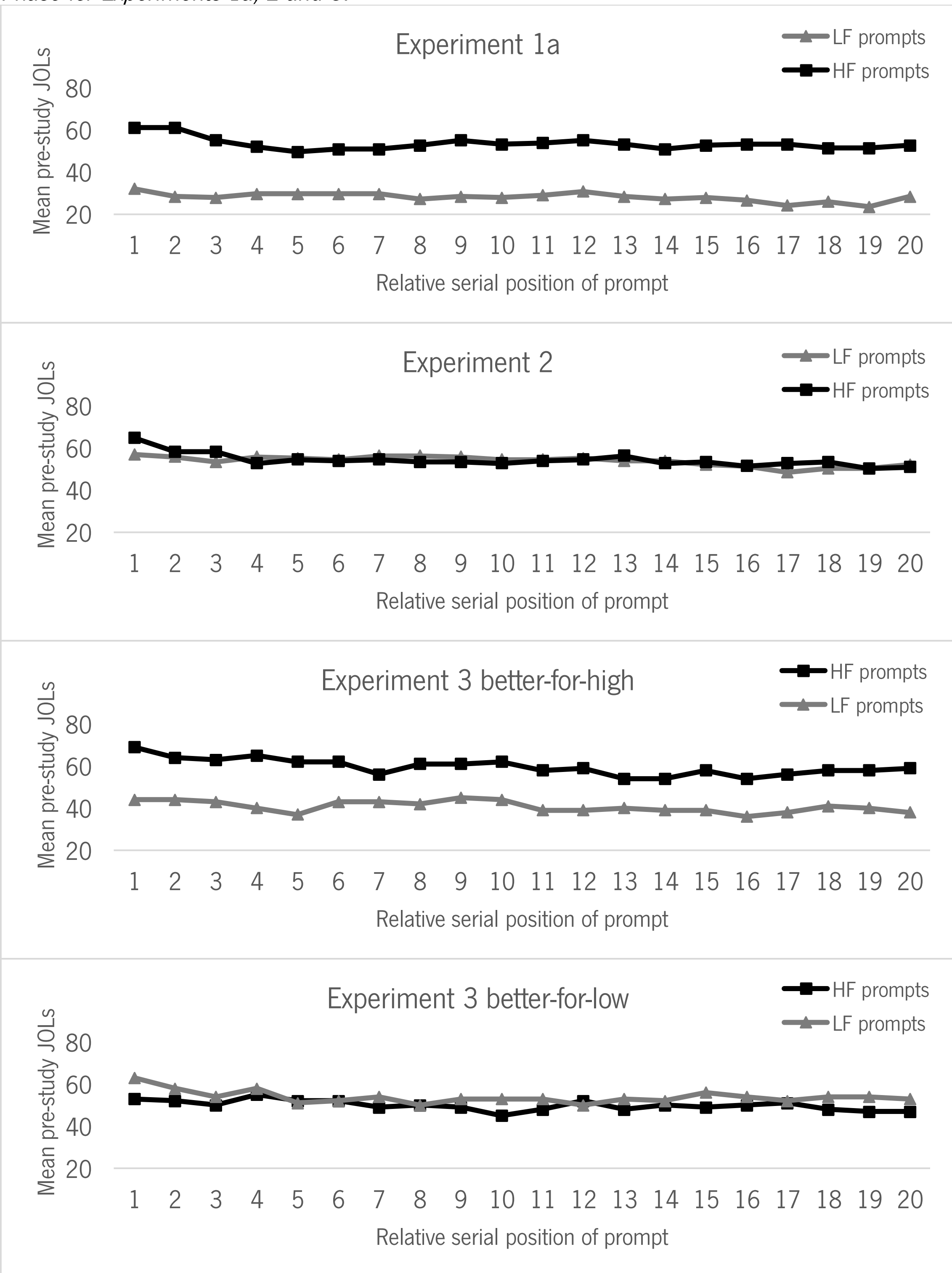


Table C.3.1

Mean Pre-study JOLs at the Beginning vs. End of the Study Phase

	First 3	Last 3	Difference	Wilcoxon signed-rank test
Experiment 1a				
Prompt of HF	59.00 (3.46)	52.33 (0.58)	-6.67	$Z = 6.00$, $p = .174$
Prompt of LF	29.33 (2.30)	26.00 (2.00)	-3.33	$Z = 3.00$, $p = .371$
Experiment 2				
Prompt of HF	60.33 (4.04)	51.33 (1.53)	-9.00	$Z = 6.00$, $p = .250$
Prompt of LF	55.33 (2.08)	50.67 (1.16)	-4.66	$Z = 6.00$, $p = .250$
Experiment 3 better-for-high				
Prompt of HF	65.33 (3.22)	58.33 (0.58)	-7.00	$Z = 6.00$, $p = .250$
Prompt of LF	43.67 (0.58)	39.67 (1.53)	-4.00	$Z = 6.00$, $p = .250$
Experiment 3 better-for-low				
Prompt of HF	51.67 (1.53)	47.33 (0.58)	-4.34	$Z = 6.00$, $p = .174$
Prompt of LF	58.33 (4.51)	53.67 (0.58)	-4.66	$Z = 6.00$, $p = .250$

Table C.3.2.

Effects of Frequency (Prompt) on Pre-study JOLs at the Beginning vs. End of Study Phase

	Cohen's *d*		Wilcoxon signed-rank test *r*	
Experiment	First 3	Last 3	First 3	Last 3
Experiment 1a	9.71**	17.24**	1.00	1.00
Experiment 2	1.67	0.32	1.00	0.67
Experiment 3 better-for-high	7.51**	8.97**	1.00	1.00
Experiment 3 better-for-low	-2.18^{+}	-10.97**	-1.00	-1.00

Notes: Non-parametric analysis showed no significant effects. $^{+}$ $p < .10$ ** $p < .01$

C.3.2. Individual Use of Word Frequency as a Cue

We explored the number of participants that used word frequency as a cue by examining pre-study and immediate JOLs at the individual level. For each participant, we calculated the individual means of pre-study JOLs given to prompts of high (vs. low) frequency and immediate JOLs given to actual high- and low-frequency words. Participants were then classified as to have given higher JOLs either for prompts/words of high frequency or for prompts/words of low frequency. Participants who had the exact same mean for prompts/words of high and low frequency were considered to have given "similar JOLs" for both kinds of prompts/words.

Experiment 1a, Pre-study JOLs. Most participants (91%) provided higher pre-study JOLs for prompts of high frequency and three participants (9%) provided higher pre-study JOLs for prompts of low frequency. These results suggest the existence of the belief that high-frequency words are better remembered.

Experiment 1a, Immediate JOLs. Most participants (97%) provided higher immediate JOLs for actual high-frequency words and only one participant (3%) provided higher immediate JOLs for actual low-frequency words. Regardless of prompt presented, most participants gave higher immediate JOLs to actual high-frequency words, suggesting that experience with the items is also important to the word frequency effect on JOLs

Experiment 1b, Immediate JOLs. Most participants (97%) provided higher immediate JOLs for actual high-frequency words and one participant (3%) provided higher immediate JOLs for low-frequency words. As in Experiment 1a, these results show that experience with the items seems to be important in the word frequency effect on JOLs.

Experiment 2, Pre-study JOLs. Twelve participants (36%) gave higher pre-study JOLs for prompts of high frequency, but the majority gave higher pre-study JOLs for prompts of low frequency (18 participants; 55%). The remaining three participants (9%) gave similar pre-study JOLs for both types of prompt. These results suggest that the manipulation of the belief was effective, at least to some extent.

Experiment 2, Immediate JOLs. Most participants (76%) gave higher immediate JOLs for actual high-frequency words, seven participants (21%) gave higher JOLs for low-frequency words, and one participant (3%) provided similar ratings for high- and low-frequency words. These results further support that direct experience with the items is more important for immediate JOLs.

Experiment 3, Pre-study JOLs. With the better-for-high instruction, 17 (85%) participants provided higher pre-study JOLs for prompts of high frequency, a figure similar to those in Experiment 1a. The remaining three participants (15%) gave higher pre-study JOLs for prompts of low frequency. With the better-for-low instruction, however, only eight participants (40%) provided higher pre-study JOLs for prompts of high frequency, a figure closer to that in Experiment 2. The remaining 12 participants (60%) gave higher pre-study JOLs for prompts of low frequency. These results suggest that belief manipulation was effective and integrated in pre-study JOLs.

Experiment 3, Immediate JOLs. With the better-for-high instruction, 16 participants (80%) gave higher immediate JOLs for high-frequency words and the other four participants (20%) gave higher JOLs for low-frequency words. With the better-for-low instruction, 13 participants (65%) gave higher immediate JOLs for actual high-frequency words. The remaining participants either gave higher JOLs for low-frequency words (six participants; 30%) or similar JOLs to high- and low-frequency words (one participant; 5%). These results provide further evidence that for immediate JOLs not only theory-driven processes, but also experience-based processes are at play in the word frequency effect.

C.3.3. Effects of Beliefs on Immediate JOLs (Linear Mixed Effects)

In the review process of our manuscript, it was suggested to further explore the role of beliefs about word frequency on immediate JOLs by analyzing linear mixed effects with pre-study JOLs as a measure of beliefs, and participants as a random factor, such as in Frank and Kuhlmann (2017).

We performed linear mixed regressions analysis in Experiments 1a, 2 and 3, using the R packages lme4 and lmerTest (Bates et al., 2015; Kuznetsova et al., 2016; R Core Team, 2016). All the factors, pre-study JOLs used as measure of beliefs, prompts (prompt of low frequency = 0, prompt of high frequency = 1) and frequency (low-frequency = 0, high-frequency = 1) were treated as fixed factors in the analysis. Participants were considered as random intercept.

Overall, results supported the effect of beliefs on immediate JOLs, except in Experiment 1a, where pre-study JOLs had no significant effect on immediate JOLs, and Experiment 3 (better-for-low), in which the effect of pre-study JOLs on immediate JOLs was marginal. These results provide further evidence for the joint contribution of theory-driven and experience-based processes in the word frequency effect on JOLs. The Table C.3.3 below presents all the statistics of the analysis, which are provided for completeness.

Table C.3.3

Mixed Model Results to Examine Effects of Beliefs and Experience on Immediate JOLs

Experiment	Estimate	SE	*df*	*t* value	*p* value
Experiment 1a					
Intercept	24.91	3.27	178.64	7.63	< .001
Frequency	25.61	3.50	1279.97	7.31	< .001
Prompt	6.67	5.14	1310.16	1.30	.194
Pre-study JOL	0.11	0.08	1311.57	1.28	.201
Frequency*Prompt	-17.23	6.81	1280.99	-2.53	.012
Frequency*Pre-stugy JOL	0.026	0.11	1280.12	0.24	.809
Prompt*Pre-study JOL	-0.20	0.11	1311.51	-1.71	.087
Frequency*Prompt*Pre-study JOL	0.33	0.15	1280.84	2.21	.027
Experiment 2					
Intercept	32.54	4.99	306.91	6.52	< .001
Frequency	13.82	5.27	1278.30	2.62	.009
Prompt	10.57	5.35	1294.59	1.974	.049
Pre-study JOL	0.18	0.077	1309.99	2.28	.023
Frequency*Prompt	-23.70	7.12	1278.43	-3.33	< .001
Frequency*Pre-stugy JOL	0.03	0.09	1278.41	0.32	.750
Prompt*Pre-study JOL	-0.24	0.09	1296.38	-2.60	.009
Frequency*Prompt*Pre-study JOL	0.44	0.12	1278.56	3.56	< .001
Experiment 3 (better-for-high)					
Intercept	19.08	5.01	200.95	3.81	< .001
Frequency	36.13	5.63	771.61	6.42	< .001
Prompt	43.76	7.03	791.99	6.23	< .001
Pre-study JOL	0.31	0.10	753.87	3.09	.002
Frequency*Prompt	-68.16	9.39	771.79	-7.26	< .001
Frequency*Pre-stugy JOL	-0.20	0.12	771.88	-1.64	.101
Prompt*Pre-study JOL	-0.76	0.13	791.45	-5.80	< .001
Frequency*Prompt*Pre-study JOL	1.12	0.17	772.11	6.54	< .001
Experiment 3 (better-for-low)					
Intercept	36.87	6.82	444.91	5.41	< .001
Frequency	22.85	8.03	774.10	2.85	.005
Prompt	22.24	8.25	781.83	2.70	.007
Pre-study JOL	0.20	0.11	791.78	1.74	.083
Frequency*Prompt	-48.998	10.23	773.49	-4.79	< .001
Frequency*Pre-study JOL	-0.26	0.14	774.17	-1.83	.067
Prompt*Pre-study JOL	-0.52	0.15	776.07	-3.46	< .001
Frequency*Prompt*Pre-study JOL	1.02	0.19	773.56	5.48	< .001

Notes: Regression weights were computed using R packages lme4 and lmerTest. For frequency, 0 = low-frequency, 1 = high-frequency; for prompt, 0 = low-frequency, 1 = high-frequency. Pre-study JOL were used as a measure for belief effects.

C.3.4. References

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

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(5), 680–693. http://doi.org/10.1037/xlm0000332

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: *Tests in linear mixed effects models* (R Package Version 2.0-33) [Computer software]. Retrieved from https://CRAN.Rproject.org/package=lmerTest

R Core Team. (2016). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing.

Appendix D. Instructions and Stimuli Used in Study 3

D.1. Instructions Used in Study 3

Table D.1.
Instructions for Experiments in Study 3.

Experiment phase	Instruction
General instructions	Please read the following information carefully. In this experiment you will learn 44 words that will be shown on the screen in sequence. Please try to remember all the words. Later you will be asked to write down all the words you have learned. After learning each word, rate the likelihood of remembering it on a scale from 0 to 100. If you are sure that you will not remember the word later, please select 0. If you are sure that you will remember the word later, please select 100. Use the values 10, 20, 30, 40, 50, 60, 70, 80 and 90 to grade your assessments. If you have any question, please ask now. When you are ready to start the experiment, press the button START.
JOL (after each word)	Chance to recall (0% -100%)? (Mouse-clicking scale: 0; 10; 20; 30; 40; 50; 60; 70; 80; 90; 100)
Filler task	In this phase of the experiment you will solve arithmetic problems. You have 3 minutes for this task. Please enter the correct result (one result per line) in the right column. When you are ready, press the button START.
Recall task	Please read the following instructions carefully. You now have 4 minutes to write down all the words you have learned and remembered. Please only write one word in each line. When you are ready to start the experiment, press the button START. (At the end): The experience is over. Thank you for participating!

Note: If you would like the original instructions (in German), please contact the author.

D.2. Stimuli Used in Study 3

For all tables presented here, frequency values (fpm = frequency per million) and number of letters and syllables were taken from dlex-DB (Heister et al., 2011); concreteness, valence, and arousal values were taken from BAWL-R (Võ et al., 2009). Zipf values were calculated considering the fpm values by taking $\log_{10}$(fpm) + 3, as suggested by van Heuven et al. (2014).

Table D.2.1

Words Used in Experiment 1

Frequency	Word	Objective Frequency	Concreteness	Letters	Syllables	Valence	Arousal	Zipf
Low	Amboss	0.01	4.64	6	2	-0.32	2.32	0.91
	Rasthaus	0.05	5.56	8	2	0.80	2.24	1.69
	Klette	0.14	3.82	6	2	-0.90	2.26	2.14
	Morphem	0.25	1.89	7	2	-1.00	2.00	2.39
	Pegel	0.31	3.50	5	2	-0.45	2.50	2.49
	Ampulle	0.32	3.82	7	3	-0.73	3.14	2.50
	Girlande	0.34	5.41	8	3	1.36	2.91	2.53
	Mundraub	0.19	3.22	8	2	-1.20	3.39	2.27
	Liaison	0.45	3.44	7	3	1.00	3.17	2.65
	Halunke	0.47	4.67	7	3	-1.40	3.50	2.67
	Ritze	0.53	3.77	5	2	-0.15	2.53	2.73
	Fibel	0.54	5.00	5	2	0.40	1.95	2.73
	Karaffe	0.54	4.55	7	3	0.41	1.95	2.73
	Kerbe	0.57	3.59	5	2	-0.20	2.55	2.76
	Alpdruck	0.72	2.89	8	2	-1.90	3.17	2.86
	Portwein	0.76	5.78	8	2	1.10	2.17	2.88
	Manko	0.80	1.91	5	2	-1.05	2.78	2.90
	Rubin	0.82	6.22	5	2	1.40	2.61	2.91
	Undank	0.82	1.95	6	2	-1.80	3.11	2.91
	Mörser	0.92	4.59	6	2	-0.45	2.32	2.96
	Mean (SD)	**0.48 (0.27)**	**4.01 (1.27)**	**6.45 (1.19)**	**2.25 (0.44)**	**-0.25 (1.03)**	**2.63 (0.49)**	**2.53 (0.49)**
High	Fenster	100.85	6.14	7	2	1.14	1.91	5.00
	Wahrheit	104.19	2.22	8	2	2.50	3.00	5.02
	Kultur	105.05	2.91	6	2	1.50	2.42	5.02
	Tod	105.47	4.44	3	1	-2.80	4.06	5.02
	Seele	105.69	2.68	5	2	1.80	2.35	5.02
	Abend	113.27	4.45	5	2	1.65	1.83	5.05
	Anfang	114.55	2.44	6	2	0.80	3.00	5.06
	Stunde	115.71	1.82	6	2	0.25	2.06	5.06
	Minister	116.58	4.56	8	3	-0.10	2.28	5.07
	Kampf	128.47	5.44	5	1	-1.60	4.05	5.11
	Professor	131.84	6.00	9	3	0.06	2.58	5.12
	Nacht	150.14	6.11	5	1	1.00	2.89	5.18
	Beispiel	163.21	2.69	8	2	0.30	1.94	5.21
	Stelle	169.25	3.38	6	2	0.00	2.22	5.23
	Wasser	182.08	5.68	6	2	1.55	1.83	5.26
	Partei	185.70	3.44	6	2	-0.60	2.62	5.27
	Politik	196.62	3.69	7	3	-0.88	3.00	5.29
	Mutter	247.78	5.68	6	2	2.05	2.67	5.39
	Ende	278.32	2.44	4	2	-0.60	2.81	5.44
	Regierung	428.15	4.62	9	3	-0.85	2.84	5.63
	Mean (SD)	**162.15 (80.01)**	**4.04 (1.45)**	**6.25 (1.59)**	**2.05 (0.60)**	**0.36 (1.34)**	**2.62 (0.63)**	**5.17 (0.17)**

Table D.2.2

Words Used in Experiment 2

Bin (range of frequency)	Word	Objective Frequency	Concreteness	Letters	Syllables	Valence	Arousal	Zipf
1 (0-1)	Amboss	0.01	4.64	6	2	-0.32	2.32	0.91
	Ritze	0.53	3.77	5	2	-0.15	2.53	2.73
	Rubin	0.82	6.22	5	2	1.40	2.61	2.91
	Delle	0.44	4.27	5	2	-0.90	2.33	2.64
	Stelze	0.01	4.73	6	2	0.18	2.64	0.91
	Imker	0.82	5.23	5	2	1.00	2.17	2.91
	Mean (*SD*)	**0.44 (0.37)**	**4.81 (0.84)**	**5.33 (0.52)**	**2.00 (0.00)**	**0.20 (0.86)**	**2.43 (0.19)**	**2.17 (0.98)**
2 (1-5)	Ruder	4.45	5.27	5	2	0.40	2.44	3.65
	Anreiz	3.52	2.27	6	2	1.30	3.74	3.55
	Roggen	4.93	4.59	6	2	0.65	1.79	3.69
	Laube	3.95	4.73	5	2	1.30	1.75	3.60
	Beutel	4.95	4.77	6	2	0.60	1.42	3.69
	Revier	3.99	2.95	6	2	-0.20	2.59	3.60
	Mean (*SD*)	**4.30 (0.58)**	**4.10 (1.19)**	**5.67 (0.52)**	**2.00 (0.00)**	**0.68 (0.57)**	**2.29 (0.84)**	**3.63 (0.06)**
3 (7-15)	Mond	19.46	6.89	4	1	1.30	1.83	4.29
	Teller	19.48	6.14	6	2	0.59	1.59	4.29
	Miene	13.60	4.18	5	2	0.55	2.90	4.13
	Hemd	14.43	6.18	4	1	0.68	1.95	4.16
	Reiter	19.33	5.23	6	2	1.00	2.56	4.29
	Orden	19.15	4.27	5	2	0.35	2.44	4.28
	Mean (*SD*)	**17.58 (2.77)**	**5.48 (1.11)**	**5.00 (0.89)**	**1.67 (0.52)**	**0.75 (0.35)**	**2.21 (0.50)**	**4.24 (0.07)**
4 (20-40)	Absatz	39.89	4.04	6	2	-0.09	2.24	4.60
	Regen	26.76	6.77	5	2	0.32	2.06	4.43
	Gast	32.87	4.81	4	1	1.41	2.28	4.52
	Stufe	31.18	4.59	5	2	0.30	2.15	4.49
	Keller	22.16	6.33	6	2	0.00	2.72	4.35
	Wind	39.44	5.50	4	1	1.26	2.33	4.60
	Mean (*SD*)	**32.05 (6.98)**	**5.34 (1.06)**	**5.00 (0.89)**	**1.67 (0.52)**	**0.54 (0.64)**	**2.30 (0.23)**	**4.50 (0.10)**
5 (40-70)	Inhalt	71.98	2.67	6	2	0.00	2.25	4.86
	Heimat	53.18	3.95	6	2	1.70	1.81	4.73
	Bank	54.59	6.35	4	1	-0.24	2.28	4.74
	Nase	44.59	6.00	4	2	0.32	2.32	4.65
	Grenze	66.09	5.78	6	2	-0.20	3.00	4.82
	Wunsch	72.81	2.56	6	1	1.60	2.45	4.86
	Mean (*SD*)	**60.54 (11.46)**	**4.55 (1.72)**	**5.33 (1.03)**	**1.67 (0.52)**	**0.53 (0.89)**	**2.35 (0.38)**	**4.78 (0.08)**
6 (75-100)	Mitte	84.96	3.36	5	2	0.75	2.11	4.93
	Geist	99.68	4.00	5	1	1.10	2.83	5.00
	Himmel	86.40	5.82	6	2	2.25	1.65	4.94
	Arzt	87.42	6.58	4	1	0.09	2.78	4.94
	Wagen	92.03	4.42	5	2	1.06	2.11	4.96
	Erfolg	96.40	3.05	6	2	2.10	3.32	4.98
	Mean (*SD*)	**91.15 (5.93)**	**4.54 (1.39)**	**5.17 (0.75)**	**1.67 (0.52)**	**1.22 (0.82)**	**2.47 (0.61)**	**4.96 (0.03)**

Table D.2.2 (continued)

Bin (range of frequency)	Word	Objective Frequency	Concreteness	Letters	Syllables	Valence	Arousal	Zipf
7 (100-200)	Buch	118.34	6.84	4	1	1.91	1.89	5.07
	Gesetz	116.01	2.86	6	2	-0.90	2.68	5.06
	Hilfe	106.34	3.14	5	2	1.30	3.00	5.03
	Kultur	105.05	2.91	6	2	1.50	2.42	5.02
	Abend	113.27	4.45	5	2	1.65	1.83	5.05
	Tisch	112.92	6.09	5	1	0.50	1.77	5.05
Mean (*SD*)		**111.99 (5.27)**	**4.38 (1.73)**	**5.17 (0.75)**	**1.67 (0.52)**	**0.99 (1.04)**	**2.27 (0.51)**	**5.05 (0.02)**
8 (125-400)	Gott	201.03	3.44	4	1	1.50	2.50	5.30
	Seite	272.54	4.09	5	2	0.25	1.89	5.44
	Herr	396.59	5.15	4	1	-0.29	2.56	5.60
	Gebiet	129.11	2.95	6	2	0.10	2.32	5.11
	Frage	362.61	2.68	5	2	0.15	2.33	5.56
	Wasser	182.08	5.68	6	2	1.55	1.83	5.26
Mean (*SD*)		**257.33 (105.81)**	**4.00 (1.21)**	**5.00 (0.89)**	**1.67 (0.52)**	**0.54 (0.78)**	**2.24 (0.31)**	**5.38 (0.19)**

Table D.2.3

Words Used in Experiment 3

Word Type	Word	Objective Frequency	Concreteness	Letters	Syllables	Valence	Arousal	Zipf
Math (related to math)	Algorithmus	0.01	3.20	11	4	-0.30	1.90	0.91
	Urbild	0.01	3.00	6	2	-0.23	2.13	0.91
	Portfolio	0.11	5.10	9	4	-0.20	2.50	2.03
	Theorem	0.02	2.88	7	3	0.13	2.40	1.21
	Arithmetik	0.02	3.10	10	4	-0.45	2.25	1.21
	Anfangswert	0.03	3.60	11	3	0.20	2.10	1.51
	Involution	0.06	1.25	10	5	-1.00	3.13	1.76
	Vektor	0.72	3.10	6	2	-0.10	2.10	2.86
	Metrik	1.24	2.50	6	2	-0.10	2.10	3.09
	Konvergenz	1.86	2.90	10	3	-0.30	2.18	3.27
	Mean (*SD*)	**0.41 (0.66)**	**3.06 (0.95)**	**8.60 (2.12)**	**3.20 (1.03)**	**-0.24 (0.33)**	**2.28 (0.34)**	**1.88 (0.90)**
Psychology (related to psychology)	Adaption	0.02	3.05	8	4	-0.10	2.03	1.21
	Screening	0.02	4.00	9	2	-0.70	2.90	1.21
	Selektion	0.02	3.40	9	4	-1.00	2.50	1.39
	Attribution	0.07	2.30	11	5	0.00	1.83	1.87
	Persistenz	0.07	2.13	10	3	-0.73	2.88	1.87
	Hemmung	5.49	2.68	7	2	-1.88	2.90	3.74
	Kognition	0.49	2.00	9	4	0.40	2.30	2.69
	Emotion	0.73	3.20	7	4	-0.40	3.80	2.86
	Rezeptor	0.80	4.85	8	3	-0.20	2.35	2.90
	Diagnostik	1.67	3.40	10	4	-0.40	2.20	3.22
	Mean (*SD*)	**0.94 (1.68)**	**3.10 (0.88)**	**8.80 (1.32)**	**3.50 (0.97)**	**-0.50 (0.63)**	**2.57 (0.57)**	**2.30 (0.90)**
Standard Low-frequency	Partisan	0.28	4.67	8	3	-0.80	2.74	2.44
	Reiberei	0.07	3.22	8	3	-1.40	3.50	1.87
	Missetat	0.43	3.33	8	3	-2.20	3.22	2.64
	Rücklage	0.60	1.96	8	3	0.03	2.11	2.78
	Rasthaus	0.05	5.56	8	2	0.80	2.24	1.69
	Wehklage	0.16	3.22	8	3	-1.90	3.41	2.19
	Karaffe	0.54	4.55	7	3	0.41	1.95	2.73
	Girlande	0.34	5.41	8	3	1.36	2.91	2.53
	Portwein	0.76	5.78	8	2	1.10	2.17	2.88
	Mischehe	0.51	2.44	8	3	-0.50	2.78	2.70
	Mean (*SD*)	**0.37 (0.24)**	**4.01 (1.36)**	**7.90 (0.32)**	**2.80 (0.42)**	**-0.31 (1.26)**	**2.70 (0.57)**	**2.44 (0.40)**
Standard High-frequency	Tätigkeit	100.17	2.92	9	3	0.74	2.76	5.00
	Beispiel	163.21	2.69	8	2	0.30	1.94	5.21
	Minister	116.58	4.56	8	3	-0.10	2.28	5.07
	Interesse	131.59	3.19	9	4	1.62	3.19	5.12
	Stelle	169.25	3.38	6	2	0.00	2.22	5.23
	General	175.37	5.12	7	3	-1.35	2.83	5.24
	Politik	196.62	3.69	7	3	-0.88	3.00	5.29
	Bedeutung	213.54	1.81	9	3	0.94	2.86	5.33
	Geschichte	258.50	3.77	10	3	0.76	2.71	5.41
	Straße	117.96	6.46	6	2	0.18	2.21	5.07
	Mean (*SD*)	**164.28 (49.55)**	**3.76 (1.33)**	**7.90 (1.37)**	**2.80 (0.63)**	**0.22 (0.87)**	**2.60 (0.41)**	**5.20 (0.13)**

Appendix E. Thesis' Ethical Approval

The project for this dissertation received the ethical approval for experiments with humans by the Ethics' Subcommittee for Human and Social Sciences of the University of Minho (SECSH 017/2018), as shown below.

Universidade do Minho

SECSH

Subcomissão de Ética para as Ciências Sociais e Humanas

Identificação do documento: SECSH 017/2018

Título do projeto: *: (Meta)memory: Effects of word frequency on judgments of learning and recall*

Investigador(a) Responsável: Pedro B. Albuquerque: Human Cognition Laboratory, CIPsi, University of Minho, Braga, Portugal (orientador)

Outros Investigadores: Pedro Simão Oliveira Mendes: Human Cognition Laboratory, CIPsi, University of Minho, Braga, Portugal e Karlos Luna: Human Cognition Laboratory, CIPsi, University of Minho, Braga, Portugal

PARECER

A Subcomissão de Ética para as Ciências Sociais e Humanas (SECSH) analisou o processo relativo ao projeto intitulado *": (Meta)memory: Effects of word frequency on judgments of learning and recall"*.

Os documentos apresentados revelam que o projeto obedece aos requisitos exigidos para as boas práticas na investigação com humanos, em conformidade com as normas nacionais e internacionais que regulam a investigação em Ciências Sociais e Humanas.

Face ao exposto, a SECSH nada tem a opor à realização do projeto.

Braga, 02 de maio de 2018.

O Presidente

Digitally signed by PAULO MANUEL PINTO PEREIRA ALMEIDA MACHADO
Date: 2018.05.04 15:52:13 +01'00'

Paulo Manuel Pinto Pereira Almeida Machado