

# Experimental Psychology

## Do beliefs about font size affect retrospective metamemory judgments in addition to prospective judgments? A registered report --Manuscript Draft--

<b>Manuscript Number:</b>	MS-1796R4	
<b>Full Title:</b>	Do beliefs about font size affect retrospective metamemory judgments in addition to prospective judgments? A registered report	
<b>Short Title:</b>	Metamemory beliefs and confidence	
<b>Article Type:</b>	Registered Report	
<b>Keywords:</b>	metamemory; beliefs; confidence; font size	
<b>Abstract:</b>	<p>Beliefs about how memory works explain several effects on prospective metamemory judgments (e.g., the effect of font size on JOLs). Less is known about the effect of beliefs on retrospective judgments (i.e., confidence). Here, we tested whether font size also affects confidence ratings and whether beliefs play a similar role in confidence than in JOLs. In two experiments, participants studied words in small and large size, rated JOLs, and completed a font-size test in which they indicated the font size at study and a standard old/new recognition test. The results confirmed that font size affected both JOLs and confidence ratings. The presentation of the counter-belief that memory is better for words in small font size in Experiment 2 and the analyses of confidence for participants who did not believe that large fonts improved memory suggested that the effect of font size on confidence was based on beliefs. This research shows that the debate on theory-based and experience-based factors should not be limited to prospective metamemory judgments, but also encompass retrospective judgments.</p>	
<b>Manuscript Classifications:</b>	16: Memory	
<b>Funding Information:</b>	Portuguese Foundation for Science and Technology and the Portuguese Ministry of Science, Technology and Higher Education through national funds and co-financed by FEDER through COMPETE2020 under the PT2020 Partnership Agreement (POCI-01-0145-FEDER-007653)	Not applicable
<b>Corresponding Author:</b>	Karlos Luna, Ph.D. Universidad Nacional de Colombia Bogotá, Bogotá COLOMBIA	
<b>Corresponding Author Secondary Information:</b>		
<b>First Author:</b>	Karlos Luna, Ph.D.	
<b>First Author Secondary Information:</b>		
<b>Other Authors:</b>	Pedro B Albuquerque, PhD	
<b>Author Comments:</b>	This is the Stage 2 report of a previously accepted Registered Report-Stage 1.	
<b>Additional Information:</b>		
<b>Question</b>	<b>Response</b>	
Has the manuscript or any component of it already been published or is currently under consideration by another journal?	No	
If the paper has been written by more than one person, can the corresponding author attest that each author endorsed the manuscript in the form submitted,	Yes	

agreed to be listed as a coauthor, and has accepted the order of authorship?	
Is there any conflict of interests concerning the content of this manuscript?	No
The manuscript comprehensively reports study design and data analysis, including all manipulations and all measures used; how sample size was determined; and which data, if any, were excluded from the analyses.	Yes
Does the length of the manuscript fall within the journal's word limits? [Preregistration of a single experiment: 5,000 words; Two and more experiments: 10,000 words – including title page, abstract, text, references, notes, appendices, tables, and figures]	Yes
Please state the word count. as follow-up to "Does the length of the manuscript fall within the journal's word limits? [Preregistration of a single experiment: 5,000 words; Two and more experiments: 10,000 words – including title page, abstract, text, references, notes, appendices, tables, and figures]"	A bit over 10000 words to accomodate Reviewers comments
Has the manuscript been prepared according to the Publication Manual of the American Psychological Association (7th ed.)?	Yes
Has this manuscript or another version of it been submitted to this journal previously?	No
Has the study that is being proposed in this manuscript already been conducted?	No
Before acceptance for publication, the manuscript must contain a permanent URL pointing to the raw data as well as any other required additional information (e.g., variable names, codebook, analysis script, etc.; please use open format, e.g., .csv, .txt format). Raw data must be deposited in a public or institutional repository that provides a permanent digital object identifier (DOI) for the electronic files (for a list of repositories see <a href="http://www.re3data.org">http://www.re3data.org</a> ). Please provide a paragraph specifying the DOI at	The manuscript includes a DOI generated by a public data repository pointing to the raw data underlying the findings reported in the article.

<p>the end of the manuscript after the reference section.</p>	
<p>Please provide the URL here as well. as follow-up to "Before acceptance for publication, the manuscript must contain a permanent URL pointing to the raw data as well as any other required additional information (e.g., variable names, codebook, analysis script, etc.; please use open format, e.g., .csv, .txt format). Raw data must be deposited in a public or institutional repository that provides a permanent digital object identifier (DOI) for the electronic files (for a list of repositories see <a href="http://www.re3data.org">http://www.re3data.org</a>). Please provide a paragraph specifying the DOI at the end of the manuscript after the reference section."</p>	<p><a href="https://osf.io/z62gk/?view_only=6497c613a6604d97a8aae05e2470cdf8">https://osf.io/z62gk/?view_only=6497c613a6604d97a8aae05e2470cdf8</a></p>
<p>If your original materials (including all stimuli, instructions where relevant, item wordings where relevant, and precise references to materials taken from other sources, such as tests or stimuli databases) are available from a stable online data repository, your manuscript may be eligible for the open materials badge. Please provide a URL pointing to the data repository at the end of the manuscript after the reference section.</p>	<p>To the best of my ability and knowledge, I have provided all original materials and clear references to all other materials via a stable online repository.</p>
<p>Please provide the URL here as well. as follow-up to "If your original materials (including all stimuli, instructions where relevant, item wordings where relevant, and precise references to materials taken from other sources, such as tests or stimuli databases) are available from a stable online data repository, your manuscript may be eligible for the open materials badge. Please provide a URL pointing to the data repository at the end of the manuscript after the reference section."</p>	<p><a href="https://osf.io/z62gk/?view_only=6497c613a6604d97a8aae05e2470cdf8">https://osf.io/z62gk/?view_only=6497c613a6604d97a8aae05e2470cdf8</a></p>
<p><b>Plain Language Summary</b> Please enter your Plain Language Summary below. It should be a short (between 50 and 100 words), simple, non-technical explanation or summary of your work (not an abstract). This will make it easier for colleagues from outside of your specific field or lay persons to find and</p>	<p>Future estimations of memory retrieval (JOLs) are affected by beliefs about the factors influencing memory. In this research, we showed that beliefs also affect participants' confidence that they selected the correct answer. In particular, beliefs about the effect of font size on memory influenced confidence ratings.</p>

<p>understand the significance your work. If your article is accepted for publication, the plain language summary may be used for marketing and social media purposes by Hogrefe Publishing. Feel free to use the summary for your own purposes as well.</p> <p>If you do not wish to provide a summary, please put a placeholder in the box below and continue with your submission.</p> <p><b>Example 1:</b></p> <p>The layperson's view equates aging with cognitive decline. However, research suggests that there are both gains and losses associated with aging. We propose a framework for studying how particular ecologies - niches - can lead to failure, maintenance, or improvements in performance with increased age.</p> <p><b>Example 2:</b></p> <p>People who weigh the pros and cons of a potential goal differ in their thinking from people who plan their course of action to reach a goal. The present article shows that this difference in thinking carries over to how they see and take risks. People who plan when, where, and how to act for a goal are more eager to take risks in an entirely unrelated risk game.</p> <p>By providing this text, you consent to it being used for social media and marketing purposes. Please see our privacy policy <a href="https://us.hogrefe.com/customer-service/privacy-policy">here</a>.</p>	
<p><b>Suggested Reviewers:</b></p>	<p>Monika Undorf, PhD          Universitat Mannheim          undorf@psychologie.uni-mannheim.de          Expert in the effect of theory-based and experience-based factors in metamemory judgments.</p> <p>John Dunlosky, PhD          Kent State University          jdunlosk@kent.edu          Expert in metamemory.</p>
<p><b>Opposed Reviewers:</b></p>	

Dear Dr James R Schmidt, Action Editor,

Please find attached a revised version of the Registered Report MS-1796R3 entitled "Do beliefs about font size affect retrospective metamemory judgments in addition to prospective judgments? A registered report".

Thank you very much for your helpful and constructive reviews of the report. We found the comments very useful and made a number of changes accordingly. Specific answers to the comments are presented in the Response to comments file. The Reviewers' comments are presented in upper-case and our responses in lower-case to easily discriminate between the two.

We hope that these revisions and responses satisfactorily address yours and the Reviewers' concerns

Sincerely,

The Authors

## REVIEWERS' COMMENTS:

REVIEWER #2: I AM EXCITED TO SEE THE COMPLETION OF THIS STUDY AND CONTINUE TO APPLAUD THE AUTHORS FOR CONDUCTING A REGISTERED REPORT ON THIS SUBJECT MATTER! BELOW ARE MY COMMENTS ON THE RESULTS AND DISCUSSION.

### RESULTS

1. EXPERIMENT 1 - EXPLORATORY ANALYSIS - IS THIS ANALYSIS MEANINGFUL? FIRST, THE SAMPLE SIZE DID NOT REACH THE REQUISITE SUGGESTED IN THE POWER ANALYSIS, SO ANY RESULTS MAY NOT BE MEANINGFUL. SECOND, IDEALLY ONE WOULD ONLY TEST THE SUBSAMPLE WITH SMALL > LARGE BELIEFS AND SHOW THE OPPOSITE EFFECTS IN CONFIDENCE. INSTEAD, THE AUTHORS REPORT NULL RESULTS, WHICH IS INCONSISTENT WITH A BELIEF HYPOTHESIS. AGAIN THOUGH, GIVEN THE SMALL SAMPLE SIZE, THE ANALYSIS SIMPLY MIGHT NOT BE INTERPRETABLE. GIVEN THAT THIS ANALYSIS ALSO WAS CONDUCTED IN E2 (AND ALSO WITH LESS POWER THAN REQUIRED, I BELIEVE), I WONDER IF ONLY INCLUDING THE SMALL > LARGE BELIEFS GROUP COLLAPSED ACROSS BOTH SAMPLES WOULD YIELD A) A LARGE ENOUGH SAMPLE SIZE AND B) SIGNIFICANT CONFIDENCE EFFECTS IN THE DIRECTION OF SMALL > LARGE. GIVEN THE STRONG CONCLUSIONS BEING MADE BY THE AUTHORS (P. 28) THAT THESE EXPLORATORY ANALYSES SHOWED "CLEAR RESULTS," MAKING SURE THEY ARE WELL-POWERED AND MAKING SURE THAT THEY ARE IN A CLEAR DIRECTION OF A BELIEF-BASED HYPOTHESIS IS NECESSARY.

We thank the Reviewer for this comment. We tried to be careful in interpreting the results of the exploratory analyses (see, for example, the Discussion of Experiment 1), but we missed the reference to "clear results". We have removed that clause and are now more cautious in the interpretation of the exploratory results.

In Experiments 1 and 2, we could not test participants with small > large beliefs only because there were very few participants with that belief (3 in Experiment 1 and 10 in total in Experiment 2; collapsing participants in both experiments does not seem helpful either). Instead, we collapsed them with the participants who did not hold any belief. However, as a result of this comment and point 4 below, we replaced the exploratory analyses from Experiments 1 and 2, and now we compare participants believing and not believing that large words would be better remembered. Thus, we used the full sample of participants. In both experiments, for font-size and recognition tests, the results show that font size affected confidence only for participants in the belief group. As most of the participants in the no belief group believed that font size did not affect memory, it is reasonable that there was no font-size effect. We see this result as consistent with the beliefs explanation.

2. EXPERIMENT 2 - QUESTIONS ABOUT BELIEFS: TO SHOW THAT THEIR MANIPULATION ALTERED BELIEFS, IT WOULD BE NICE TO CALCULATE A CHI-SQUARE TEST ON THE PROPORTIONS TO SEE IF THEY SIGNIFICANTLY DIFFER BETWEEN THE TWO GROUPS. OTHERWISE, THE DIFFERENCES COULD BE DUE TO CHANCE, ESPECIALLY GIVEN THAT THE DIFFERENCES IN BELIEFS BETWEEN THOSE TWO GROUPS WERE ACTUALLY SMALLER THAN THE DIFFERENCE BETWEEN THE "LARGER IS BETTER" GROUP AND THE NO-INSTRUCTION GROUP IN EXPERIMENT 1. IN GENERAL, THE GROUP DIFFERENCES IN BELIEFS FROM E1 TO E2 ARE QUITE LARGE AND SURPRISING. DO THE AUTHORS HAVE ANY EXPLANATION? WERE

## OVERLAPPING STUDENTS TESTED IN EACH EXPERIENCE AND SO MAY HAVE BECOME "WISER" IN EXPERIMENT 2?

In our lab, we use the SONA system software for data collection. SONA software registers the experiments a given student has completed, and thus we can prevent the same student from participating in related or similar experiments. As a result, we are certain that there were no overlapping participants between the two experiments.

We agree with the Reviewer that there is a surprising difference in beliefs between Experiments 1 and 2. In line with a suggestion from Reviewer 4, we think that the instructions on the effect of font size on memory caused a sort of reactivity effect (now mentioned in the Discussion of Experiment 2). That reactivity may explain the differences in beliefs between Experiment 1 and 2 and the small effect of our beliefs manipulation.

A chi-squared showed no significant differences for beliefs depending on the instruction,  $\chi^2(2) = 1.92$ ,  $p = 0.38$ . However, other results suggest that there was, indeed, an effect of the instructions in confidence. The measure of beliefs was collected at the end of the experiment and that, as pointed previously by a Reviewer, may not have been optimal. For example, experience with the items remembered may have influenced the answer to the beliefs questions. Despite that, we preferred to keep the beliefs question at the end of the experiment to avoid interference with the main measures. In sum, the results of the beliefs question do not show an effect of the manipulation, but other results from the experiment suggest that there was, indeed, a small effect.

## 3. EXPERIMENT 2 - I'M SURPRISED AT THE DIFFERENT CONCLUSIONS FROM THE NHST AND BAYES FACTOR. SOMETIMES THEY CAN DIFFER IF THE DISTRIBUTIONS OF VALUES IS HIGHLY SKEWED OR IF OUTLIERS EXIST. DID THE AUTHORS INVESTIGATE WHETHER EITHER COULD BE DRIVING THE SLIGHTLY DIVERGING CONCLUSIONS?

We did not search for reasons for the divergent results beforehand, but we explored our data after the Reviewer's comment.

First, we are not sure of the usefulness of the outlier concept in metamemory judgments as those collected here. For example, consider a participant in a given condition with confidence  $M = 5$  (in a scale 1-7),  $SD = 1$ . If we apply the popular and straightforward criterion of plus minus  $2*SD$  to identify outliers, it would mean that judgments in the range 3-7 would be acceptable and answers 1-2 would be outliers. However, we do not think these responses should be identified as outliers and removed. In our view, responses 1-2 would be perfectly acceptable answers indicating that the participant believes that the answer is very likely incorrect. The logic of outliers applies well to reaction times and other measures, but not so well to the metamemory judgments used here. There is no good reason we can come with to consider an answer 1 (even in a participant who only answered 7 for all the other items) as an outlier that should be removed. Despite this, in our experience the high variability of metamemory measures prevents the existence of outliers. As per distribution, in our experience distributions of confidence ratings depend heavily on the difficulty of the materials. With easy materials, distributions tend to be negatively skewed, and with difficult materials distributions tend to be positively skewed, with higher skewness with more extreme materials (either easy or difficult). Thus, we are not certain of what we can learn from skewness in a confidence rating, but we acknowledge that the effects of skewness in, for example, Bayesian analyses are beyond our knowledge. Our confidence distributions were slightly negatively skewed. For example, in Experiment 2, font-size test, participants answered with confidence 1 8% of times, with confidence 2 13%, 3 16%, 4 15%, 5 16%, 6 14%, and confidence 7 19% of the times. Our interpretation of the divergence in results between NHST and Bayes Factor lies in the high variability of p-values, as shown by Cumming in the p-value dance (e.g., <https://www.youtube.com/watch?v=5OL1RqHrZQ8>). A  $p = .10$ , considering the variability of p-values, may not be totally inconsistent with a  $BF = 4$ .

## DISCUSSION

4. GIVEN THE RELATIVELY WEAK EFFECTS AND MANY NON-SIGNIFICANT INTERACTIONS IN EXPERIMENT 2, I SUPPOSE I AM NOT ENTIRELY CONVINCED THAT ONE CAN SAY BELIEFS ARE A DRIVER OF FONT SIZE EFFECTS. IT COULD STILL BE THE CASE THAT A THIRD VARIABLE (E.G., FLUENCY) IS AFFECTING CONFIDENCE JUDGMENTS BOTH AT ENCODING AND AT RETRIEVAL. TO EXPLAIN WHY DIFFERING BELIEFS IN THE EXPLORATORY ANALYSIS ALTERED THE EFFECTS, ONE WOULD FIRST NEED TO CONDUCT A TEST TO SHOW A SIGNIFICANT INTERACTION (WHICH WAS NOT DONE) AND ONE WOULD HAVE TO RULE OUT THE POSSIBILITY THAT THE TWO GROUPS WERE SIMPLY NOT DIFFERENTLY PROCESSING THE STIMULI (E.G., DIFFERED ON PERCEPTUAL THRESHOLDS OR INTERPRETATION OF FLUENCY EFFECTS, ETC.). THAT IS THE WEAKNESS OF BETWEEN SUBJECT ANALYSES.

In Experiments 1 and 2, we now report analyses 2 (belief group: yes, no) x 2 (small, large) for the font-size and the recognition tests. In both experiments and tests, the interaction was significant and showed that the effect of font size on confidence was present in the belief group and absent in the no-belief group. We interpret these results as that beliefs are mostly responsible for the observed effect of font size on confidence.

5. THE DISCUSSION ON P. 30-31 ON THE RATIONALE FOR THE FONT SIZE EFFECT FOR SMALLER WORDS SOUNDS VERY SIMILAR TO A DISTINCTIVENESS HEURISTIC ARGUED BY MCDONOUGH AND GALLO (2012) AND THIS MIGHT WANT TO BE ACKNOWLEDGED IN THAT SECTION.

We thank the Reviewer for this comment. We now cite McDonough and Gallo (2012) in that section. Space constraints and the fact that memory results of the font-size test are largely secondary to our research prevented us from developing further the idea of the distinctiveness heuristic.

### REVIEWER #4: SUMMARY

THIS STUDY SOUGHT TO EXAMINE WHETHER BELIEFS ABOUT FONT SIZE AND MEMORY INFLUENCE CONFIDENCE JUDGMENTS. EXPERIMENT 1 TESTED WHETHER BELIEFS IN FONT SIZE EFFECTS INFLUENCED CONFIDENCE IN RECOGNITION MEMORY AND/OR SOURCE MEMORY. THE STUDY FOUND SMALL BUT SIGNIFICANT EFFECTS, IN CONTRAST TO PRIOR RESEARCH. EXPERIMENT 2 EXTENDED EXPERIMENT 1 BUT USED A MANIPULATION WHERE PARTICIPANTS WERE TOLD THAT FONT SIZE DID OR DID NOT INFLUENCE MEMORY. AGAIN IT WAS FOUND THAT FONT SIZE INFLUENCED CONFIDENCE. FURTHERMORE, THE EFFECT OF FONT SIZE ON CONFIDENCE WAS LARGEST FOR THOSE WHO ESPOUSED A BELIEF THAT LARGE FONT WORDS WOULD BE REMEMBERED BETTER.

### EVALUATION

OVERALL THE STUDIES WERE WELL DESIGNED AND THE MANUSCRIPT WAS WELL WRITTEN. I ESPECIALLY APPRECIATE PREREGISTRATION OF ALL ANALYSES, INCLUSION OF BOTH NHST AND BAYSIAN STATISTICS, AND CAREFUL INTERPRETATION OF RESULTS.

I HAVE GROUPED MY RECOMMENDATIONS AND CRITIQUES INTO "MAJOR" AND "MINOR" ISSUES, THOUGH I DON'T KNOW THAT ANY OF THEM ARE TRULY "MAJOR." I FEEL THAT ALL OF THEM COULD BE EASILY ADDRESSED IN A RESUBMISSION.



## "MAJOR" CONCERNS/SUGGESTIONS

THE PROPORTION OF PEOPLE BELIEVING LARGE FONT WORDS WOULD BE REMEMBERED BETTER, WHETHER THEY WERE TOLD THEY WOULD OR WOULD NOT, WAS SURPRISINGLY LOW IN EXPERIMENT 2. INDEED, THE RATES WERE HIGHER WHEN THEY WERE TOLD NOTHING (EXPERIMENT 1 AND PREVIOUS STUDIES) THAN WHEN THEY WERE TOLD LARGE FONT WOULD BE BETTER. I THINK THIS RAISES SERIOUS QUESTIONS ABOUT THE REACTIVITY OF PRIMING SUCH BELIEFS. CONTRARY TO THE EXPECTED "EXPERIMENTER EXPECTANCY EFFECTS" THE PRIME SEEMS ALMOST TO HAVE AROUSED SUSPICION. SOME DISCUSSION OF THIS IS WARRANTED.

We thank the Reviewer for this suggestion. We added a paragraph in the Discussion of Experiment 2 commenting this issue.

ALONG THE SAME LINE, IT MIGHT HELP TO DISCUSS HOW THE PROPORTION BELIEVING THAT LARGE FONT IMPROVES MEMORY COMPARES TO OTHER STUDIES. THIS WAS COMMENTED ON FOR JOLS BUT NOT BELIEFS.

To test beliefs directly, a popular way is to present a hypothetical experiment and ask participants about the likely results. For example, researchers may say that in a past experiment 40 words were presented, 20 in large font and 20 in small font, and request an estimation of how many words of each type participants in that experiment remembered. That procedure shows how many participants report higher estimates for large than small font, i.e., believe that large words would be better remembered. However, researchers use to report the estimations of the number of words remembered per condition, and the number of participants with better-large or better-small beliefs is not reported (e.g., see Hu et al., 2015; Kornell et al., 2011; or Su et al., 2017). We only found the number of participants holding better-large and better-small beliefs in Mueller et al. (2014, Experiment 3b): 88% believed better-large and 10% believed better-small. Bearing in mind the differences in the measure of beliefs in Mueller et al. (2014) and here, we preferred not to mention it in the text.

## MINOR CONCERNS/SUGGESTIONS

SUBHEADINGS FOR THE ACCURACY, CONFIDENCE, HIT RATE, ETC. ANALYSES WOULD HELP.

We thank the Reviewer for this suggestion, but we think that within each test for which we present accuracy and confidence analyses (i.e., font-size and recognition), the results are not so long as to require more subheadings.

ON A STYLISTIC NOTE, I THINK GRAPHS GENERALLY MAKE PATTERNS EASIER TO SEE THAN TABLES. BUT THIS IS JUST MY PERSONAL PREFERENCE.

We added Figure 2 with confidence in the font-size test, which is the most relevant data for the objectives of this research. We also removed the corresponding data from Table 2 to avoid repetition.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Registered Report – Stage 2

### Do Beliefs About Font Size Affect Retrospective Metamemory Judgments in Addition to Prospective Judgments? A Registered Report

Karlos Luna<sup>1</sup> and Pedro B. Albuquerque<sup>2</sup>

<sup>1</sup> Department of Psychology, Universidad Nacional de Colombia (Colombia)

<sup>2</sup> School of Psychology, University of Minho (Portugal)

#### Authors' note

Karlos Luna  <https://orcid.org/0000-0001-8592-1172>

Pedro B. Albuquerque  <https://orcid.org/0000-0002-5874-4497>

The approved registered report, data, and code to run the experiments are  
available at [https://osf.io/z62gk/?view\\_only=6497c613a6604d97a8aae05e2470cdf8](https://osf.io/z62gk/?view_only=6497c613a6604d97a8aae05e2470cdf8).

The authors have no conflicts of interest to disclose. Thanks to Nuno Fernandes for help  
with data collection.

Correspondence concerning this article should be addressed to Karlos Luna,  
Carrera 30 # 45-03, Edif. 212 Departamento de Psicología, Universidad Nacional de  
Colombia, Ciudad Universitaria, Bogotá, 111321, Colombia. E-mail:  
klunao@unal.edu.co

**Abstract**

Beliefs about how memory works explain several effects on prospective metamemory judgments (e.g., the effect of font size on JOLs). Less is known about the effect of beliefs on retrospective judgments (i.e., confidence). Here, we tested whether font size also affects confidence ratings and whether beliefs play a similar role in confidence than in JOLs. In two experiments, participants studied words in small and large size, rated JOLs, and completed a font-size test in which they indicated the font size at study and a standard old/new recognition test. The results confirmed that font size affected both JOLs and confidence ratings. The presentation of the counter-belief that memory is better for words in small font size in Experiment 2 and the analyses of confidence for participants who did not believe that large fonts improved memory suggested that the effect of font size on confidence was based on beliefs. This research shows that the debate on theory-based and experience-based factors should not be limited to prospective metamemory judgments, but also encompass retrospective judgments.

Keywords: metamemory, beliefs, confidence, font size

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## **Do Beliefs About Font Size Affect Retrospective Metamemory Judgments in Addition to Prospective Judgments? A Registered Report**

Currently, there is debate on whether metamemory judgments are based on theory-based or experience-based factors. Theory-based factors refer to the use of knowledge and beliefs about how memory works when making metamemory judgments. Experience-based factors are metamemory cues that derive from the direct experience with the items or with the task, such as retrieval fluency or retrieval speed. That debate has been informed mainly by research with prospective metamemory judgments, such as judgments of learning (JOLs, e.g., Ball et al., 2014; Besken, 2016; Besken et al., 2019; Frank & Kuhlmann, 2017; Undorf & Zimdahl, 2019; Witherby & Tauber, 2017; Yan et al., 2016) or ease-of-learning judgments (EOLs, e.g., Jemstedt et al., 2018), or with judgments about the source in which an object was presented (judgments of source or JOSs, e.g., Schaper et al., 2019).

Our aim here is to extend that debate to retrospective metamemory judgments, namely confidence (for research with the opposite objective to extend an effect from retrospective to prospective metamemory judgments, see Monds et al., 2019). Past research has shown that experience-based factors such as retrieval fluency or latency play a major role in confidence (e.g., Kelly & Lindsay, 1993; Weber & Brewer, 2006), but there is little research on the effect of theory-based factors and beliefs in confidence. Thus, our main objective is to explore the effects of metamemory beliefs on confidence. Both prospective and retrospective judgments are usually understood as stemming from the same monitoring process that examines and evaluates our memories (Nelson & Narens, 1990; Schwartz & Metcalfe, 2017). However, McDonough et al. (2021) found

1  
2 that monitoring was better conceptualized as two related sets of processes: one at  
3  
4 encoding and another at retrieval. This raises the question of whether beliefs affect  
5  
6 retrospective confidence at retrieval similarly to JOLs at encoding. The answer to this  
7  
8 question will help us identify similarities and differences between both judgments and  
9  
10 allow us to better understand the monitoring process(es) at encoding and retrieval.  
11  
12

13  
14 To study the role of beliefs on confidence, it is important to ensure that there is a  
15  
16 belief. Some manipulations of perceptual information might prove useful, namely the  
17  
18 font size and the sound volume. As per font size, participants rate materials presented in  
19  
20 larger font size with higher JOLs (Luna, Albuquerque, & Martín-Luengo, 2019; Rhodes  
21  
22 & Castel, 2008; Undorf & Zimdahl, 2019; for a meta-analysis, see Luna et al., 2018). As  
23  
24 per sound volume, words presented in louder volume are also rated with higher JOLs  
25  
26 (Frank & Kuhlmann, 2017; Peynircioğlu & Tatz, 2019; Rhodes & Castel, 2009).  
27  
28

29  
30 Current knowledge strongly suggests that these results are mostly due to people's  
31  
32 beliefs (Blake & Castel, 2018; Luna, Nogueira, & Albuquerque, 2019; Undorf &  
33  
34 Zimdahl, 2019):<sup>1</sup> People believe that words in large font or louder volume are better  
35  
36 remembered than words in small font or lower volume (Frank & Kuhlmann, 2017;  
37  
38 Kornell et al., 2011; Mueller et al., 2014). Thus, perceptual information provides the  
39  
40 perfect opportunity to test the effects of beliefs on confidence. In this research, we focus  
41  
42 on font size for practical reasons, but we also review relevant studies about volume.  
43  
44  
45  
46  
47

48  
49 In the remaining of this introduction, we first describe studies suggesting that font  
50  
51 size and volume, in general, do not affect confidence or that the effect is, at best, small.  
52

53  
54 Then, we argue that there are theoretical and empirical reasons to expect an effect of

55  
56 1 There was a debate on whether experience-based factors (e.g., fluency) could also account for  
57  
58 that effect, but Luna, Nogueira, and Albuquerque (2019) and Undorf and Zimdahl (2019) found that with  
59  
60 very large font sizes the effect of font size on JOLs was higher, but processing fluency was lower. These  
61  
62 results are incompatible with the idea that experience-based factors are mostly responsible for the font-  
63  
64 size effect on JOLs.  
65

1  
2 beliefs on confidence and offer an explanation for the results from previous studies.  
3

4 Finally, we present the current research and the hypotheses described in the registered  
5 report (available at  
6  
7 [https://osf.io/z62gk/?view\\_only=6497c613a6604d97a8aae05e2470cdf8](https://osf.io/z62gk/?view_only=6497c613a6604d97a8aae05e2470cdf8)).  
8  
9

### 10 11 **Does Perceptual Information Affect Confidence Ratings?** 12

13  
14 To our knowledge, only seven experiments in three papers manipulated font size  
15 (six experiments) or volume (one experiment) and collected confidence ratings. All the  
16 experiments collected JOLs and showed the expected effect, i.e., higher JOLs for large  
17 than small font and for loud than quiet volume. All the experiments but two showed no  
18 effect of perceptual information on confidence.  
19  
20  
21  
22  
23  
24

25  
26 Three of the experiments that collected confidence ratings were presented in Luna  
27 et al. (2018; see the Supplemental Materials of that article for a full description). In  
28 Study 1, participants read words presented in either small or large font size, completed a  
29 recognition test with confidence ratings, and font size did not affect confidence.  
30  
31 However, there was no information about font size at test, so participants may have just  
32 ignored that cue. Thus, the information that could be obtained from that experiment is  
33 limited. In Studies 2 and 5, participants read word pairs in small and large font and  
34 completed a cued recall test in which the cues were presented in the same font as during  
35 the study phase. Again, font size did not affect confidence. However, in a cued recall  
36 test participants have to generate the answer, which may produce a set of cues that may  
37 be either more diagnostic or weigh more in the confidence rating (e.g., retrieval fluency  
38 or speed of retrieval, Kelley & Lindsay, 1993), thus reducing the need of other cues  
39 such as font size.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Maybe in a recognition memory test participants will not have the cues stemming from the process of generating the answers and, thus, may rely on other cues for confidence ratings, including beliefs. For example, McDonough and Gallo (2012) presented words twice, in either small, large, or both font sizes. Then, participants completed several yes/no recognition tests with studied and new words presented in intermediate font size. In the “large” test, participants had to answer “yes” if the word was studied in large font and in the “small” test they had to answer “yes” if the word was studied in small font. In both tests, participants also rated their confidence. Across three experiments, confidence was higher for hits in the “large” test (i.e., answers “large” to words studied in large size) than in the “small” test (i.e., answers “small” to words studied in small size), although these effects were small ( $d$  between 0.10<sup>2</sup> and 0.43) and were not statistically significant in Experiment 2. The authors concluded that “These differences are consistent with the idea that participants carried a false expectation that large words were better remembered than small words” (p. 395). In other words, the results from McDonough and Gallo (2012) suggest that, under certain conditions, beliefs about perceptual information may affect confidence.

In particular, three procedural details in McDonough and Gallo (2012) may have helped beliefs affect confidence. First, items were presented twice, which may have increased participants’ ability to recollect font size at test. Second, the difference between small (25 points) and large font size (125 points) was larger than in other studies. When the difference between small and large font size increases, the effect on JOLs is larger (Halamish, 2018; Undorf & Zimdahl, 2019), and, thus, a larger font size

2 McDonough and Gallo (2012) did not report Cohen’s  $d$  for non-significant results. We computed this effect size from the reported descriptive statistics, and thus this number should be considered an approximation to the actual figure.

1  
2 difference may have increased the chances of detecting an effect on confidence. And  
3  
4 third, at test participants were requested to recognize only words in large or in small  
5  
6 font, which may have made font size salient at test. If font size affects metamemory  
7  
8 judgments via beliefs, those beliefs must be active to affect metamemory (Kornell et al.,  
9  
10 2011). Thus, it may be that by making font size salient, the task activated the belief that  
11  
12 font size affects memory. These procedural details may have increased the sensitivity of  
13  
14 the test to detect the effect of beliefs on confidence. However, some other procedural  
15  
16 details may have worked in the opposite direction. For example, McDonough and Gallo  
17  
18 presented 144 words twice at study. The high number of words at study may have made  
19  
20 font size recollection for a specific item more difficult.  
21  
22  
23  
24  
25

26  
27 Finally, Frank and Kuhlmann (2017) asked participants whether they thought they  
28  
29 were going to remember better words in quiet or in loud volume. Seventy-five percent  
30  
31 of participants reported believing that loud words were going to be remembered better,  
32  
33 thus ensuring that they had a belief in the expected direction. Then, participants listened  
34  
35 to words in quiet or loud volume and completed a written old/new recognition test with  
36  
37 confidence. Importantly, for the confidence rating the authors provided information  
38  
39 about the volume at study. If participants paid attention to that information, that should  
40  
41 have activated the pre-existing belief as in McDonough and Gallo (2012). Despite that,  
42  
43 the results did not show an effect of volume on confidence.  
44  
45  
46  
47

48  
49 In sum, the above results suggest that, in general, perceptual information, either  
50  
51 font size or sound volume, has a very limited effect on confidence ratings. Most of the  
52  
53 experiments did not find such effect, but under certain conditions it can be detected.  
54  
55 One explanation for this pattern of results is that theory-based factors do not affect  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



1  
2 confidence and the positive results were Type-I errors. We now provide some evidence  
3  
4 that suggests otherwise.  
5

### 6 7 **Do Beliefs Affect Confidence Ratings?** 8

9  
10 Metamemory judgments are inferences made from cues available at the time of  
11 the judgment (e.g., Koriat, 1997), and many researchers have included metacognitive  
12 beliefs in their general understanding of the cues for confidence ratings (e.g., Buratti &  
13 Allwood, 2015; Koriat et al., 2008). In addition, there are also empirical examples of  
14 beliefs affecting confidence.  
15  
16  
17  
18  
19  
20

21  
22 For example, Leippe et al. (2009, p. 195) proposed the cue-belief model of  
23 retrospective confidence, based on Koriat's (1997) cue-utilization approach to JOLs.  
24 The cue-belief model proposes that confidence is an inference made from cues  
25 emanating from three sources: one's beliefs, one's mental processes, and the testing  
26 situation. In line with their model and the proposed relevant role of beliefs, Leippe et al.  
27 found that the belief that if memory is good for one aspect of an event, then memory  
28 will also be good for other aspects of the same event explained why feedback increased  
29 confidence. Perfect (2004) offered another example of beliefs affecting retrospective  
30 confidence. Before the experiment, participants stated that they would perform better at  
31 recognizing faces than answering questions about sports, and although there were no  
32 differences in actual performance, they still rated faces with higher confidence,  
33 consistently with their initial belief.  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

50  
51 Similarly, Brewer et al. (2005) also proposed that retrospective confidence was  
52 affected by metamemory beliefs. In particular, they tested the effect of the belief that a  
53 complete recall is an accurate recall on confidence. The authors presented sentences that  
54 promoted a synonym substitution, e.g., from "Russian is difficult to pronounce" to  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 “Russian is hard to pronounce”, and later measured memory, confidence, and whether  
3  
4 participants thought that they recalled the sentence completely, partially, or not at all. In  
5  
6 the cued recall test, participants committed many synonym substitutions and, despite the  
7  
8 warning of the strict correction criteria, they believed that their recall was complete.  
9  
10  
11 Importantly, participants rated the synonym substitutions with similar confidence than  
12  
13 actually complete responses and sentences that did not promote a synonym substitution.  
14  
15  
16 These results were interpreted as indicating that participants hold the metamemory  
17  
18 belief that a complete recall is a good indicator of accuracy and thus rated confidence  
19  
20 accordingly (for a related study with similar results and conclusions, see Brewer &  
21  
22 Sampaio, 2006).  
23  
24

25  
26  
27 Other examples of retrospective confidence models include beliefs or theory-  
28  
29 based factors in the form of previous knowledge used as a cue to confidence ratings.  
30  
31 The theory of probabilistic mental models (Gigerenzer et al., 1991) prominently features  
32  
33 prior knowledge's effect on memory and confidence ratings. Similarly, the consensuality  
34  
35 principle (Koriat, 2008), the empirical result that confidence in two-alternative forced-  
36  
37 choice questions is correlated with the proportion of people endorsing one of the  
38  
39 alternatives and not with the accuracy of the answer, is also consistent with theory-  
40  
41 based factors.  
42  
43  
44

45  
46 In sum, the distinction between theory-based and experience-based factors seems  
47  
48 transversal to metamemory judgments, and thus it appears that beliefs should influence  
49  
50 any type of metamemory judgment, including confidence. Then, why does the belief  
51  
52 that memory is better for words presented in large font have a robust and consistent  
53  
54 effect on prospective metamemory judgments but not on confidence?  
55  
56

### 57 58 **The Present Research** 59 60 61 62 63 64 65

1  
2 We propose that the conditions in previous studies were not optimal to find the  
3  
4 expected effect of beliefs on confidence. In particular, all the experiments included a  
5  
6 memory task in which participants had to retrieve or recognize the items. Experience-  
7  
8 based cues stemming from item retrieval, such as familiarity (Yonelinas, 1994),  
9  
10 subjective vividness (Robinson et al., 2000), or retrieval fluency (Kelley & Lindsay,  
11  
12 1993), might be more relevant than theory-based factors for retrospective judgments  
13  
14 (Frank & Kuhlmann, 2017). This may explain why previous studies mostly failed to  
15  
16 show an effect of beliefs on retrospective confidence. To overcome this limitation, we  
17  
18 propose a memory test focused on font size to highlight theory-based cues (i.e., a font-  
19  
20 size test), which might provide more room to detect an effect of beliefs on confidence.  
21  
22 In the font-size test, participants indicate the font size in which words previously  
23  
24 studied were presented.  
25  
26  
27  
28  
29  
30

31 The proposed font-size test is similar to a source memory test, in which  
32  
33 participants are requested to indicate the source of an item or to indicate some  
34  
35 characteristic present at study. However, note that in a typical source memory test, item  
36  
37 retrieval is also mixed with the source task, that is, participants had to indicate whether  
38  
39 an item was presented by source 1, source 2, or it was new. In the font-size test, we only  
40  
41 presented studied words and asked participants to identify font size, which simplifies  
42  
43 the task. Despite this difference, past studies on source memory can inform us whether  
44  
45 participants remember details such as the voice that read a statement or the colour in  
46  
47 which it was presented. In general, the results confirm that memory for characteristics  
48  
49 of the items is good, even after a delay or in a surprise test. For example, in an  
50  
51 immediate test young adults correctly identified 77% of the time that the voice that read  
52  
53 a sentence was male or female, and after 24 hours they still identified it correctly 64%  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 of the time (Dodson, Bawa, & Slotnick, 2007; see also Dodson, Bawa, & Krueger,  
3  
4 2007, for similar results). Likewise, in a surprise test participants correctly recognized  
5  
6 the colour in which a sentence was studied 69% of the time and the voice that read it  
7  
8 68% of the time (Fazio & Marsh, 2009). These results show that memory for item  
9  
10 characteristics is good in general, and thus we expect that during the font-size test  
11  
12 participants will retrieve the font size at study.  
13  
14  
15

16  
17 As mentioned, with the font-size test we expect to elicit cues related to font size.  
18  
19 Some of these cues will be experience-based, but we expect that beliefs about the effect  
20  
21 of font size will also be activated and used to rate confidence. If beliefs affect  
22  
23 retrospective confidence in a similar way than JOLs, then we expect that participants  
24  
25 responding that a word was studied in large font size will access the belief that these  
26  
27 words are better remembered, which should increase confidence. The underlying  
28  
29 process could be exemplified by a participant thinking “I remember that this word was  
30  
31 presented in large font size; of course I remember, words in large font are better  
32  
33 remembered”. Such thinking will not happen with words in small font.  
34  
35  
36  
37  
38

### 39 **Hypotheses**

40  
41 To test the effect of font size and beliefs in confidence, we conducted two  
42  
43 experiments approved by the local Ethics Committee. In both experiments, we  
44  
45 presented participants with words in either small or large font and requested JOLs. We  
46  
47 expected higher JOLs for words in large than in small font. After that, participants  
48  
49 completed the font-size test, in which they had to identify the font size in which words  
50  
51 were studied and rated their confidence in the response. In this test, we expected no  
52  
53 differences in font-size identification. The effect of beliefs on confidence may be shown  
54  
55 in two non-exclusive ways. First, confidence may be higher for responses “large” than  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 for responses “small”. Second, the effect may be limited to correct responses, that is,  
3  
4 there may be higher confidence for “large” correct responses than for “small” correct  
5  
6 responses. This may be the case if participants access the actual font size at study. In  
7  
8 that case, beliefs may boost confidence mainly for correct responses to words in large  
9  
10 font, with no effect on incorrect “large” responses.  
11  
12

13  
14 Finally, participants also completed a standard yes/no recognition test with  
15  
16 confidence to provide a comparison with previous studies. Most studies on font size  
17  
18 have found no effect of font size on memory (Kornell et al., 2011; Mueller et al., 2014;  
19  
20 Rhodes & Castel, 2008), although others did (Halamish, 2018; Luna et al., 2018; Price  
21  
22 & Harrison, 2017; Undorf & Zimdahl, 2019). When present, the effect of font size on  
23  
24 memory is usually small and limited to studies with free recall test (Chang & Brainerd,  
25  
26 in press; Luna et al., 2018). Thus, we do not expect any effect of font size on memory.  
27  
28 As per confidence, the results may depend on the strength of the effect of beliefs in  
29  
30 confidence. If the effect is strong, then all the analyses could show higher confidence  
31  
32 for words in large than in small font size, but if the effect is weak it may be limited to  
33  
34 some types of response, particularly hits.  
35  
36  
37  
38  
39  
40

41 Finally, participants answered a question to test their beliefs directly. Following  
42  
43 previous research (e.g., Mueller et al., 2014), we expect that most participants will  
44  
45 believe that words in large font are better remembered.  
46  
47

## 48 **Experiment 1**

### 49 **Method**

#### 50 *Participants*

51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

1  
2 confidence, and aimed at a sample size able to detect an effect size within that range,  
3  
4 *Cohen's d* = 0.35. A power analysis with the package pwr (Champely, 2018) in R (R  
5  
6 Core Team, 2018) showed that 66 participants were necessary to obtain a power = .80  
7  
8 with alpha = .05. A total of 72 university students took part in the experiment (55  
9  
10 females, age  $M = 20.76$ ,  $SD = 3.38$ ) in exchange for course credits. The only requisite  
11  
12 was that participants were native speakers of the language in which the experiment was  
13  
14 conducted.  
15  
16  
17

### 18 *Design*

19  
20 The only variable was font size (small, 18 points; large, 100 points), manipulated  
21  
22 within subject. We used a 100 points large font size instead of the usual 48 points  
23  
24 because the font-size effect is stronger with larger font sizes (Luna, Nogueira, &  
25  
26 Albuquerque, 2019; Undorf & Zimdahl, 2019). With a larger-than-usual font size, we  
27  
28 expected to increase the chances of finding an effect on confidence.  
29  
30  
31  
32

### 33 *Materials*

34  
35 One hundred and four singular nouns were selected from a linguistic database.  
36  
37 For counterbalancing purposes, these words were separated into six lists of 16 words,  
38  
39 matched in number of letters and word frequency. The remaining eight words were used  
40  
41 as primacy and recency buffers during study. Four lists of 16 words were used during  
42  
43 the study phase, two of them were later used in the font-size test and the other two in the  
44  
45 recognition test. The two final lists were used as non-studied words for the recognition  
46  
47 test (see below).  
48  
49  
50  
51

### 52 *Procedure*

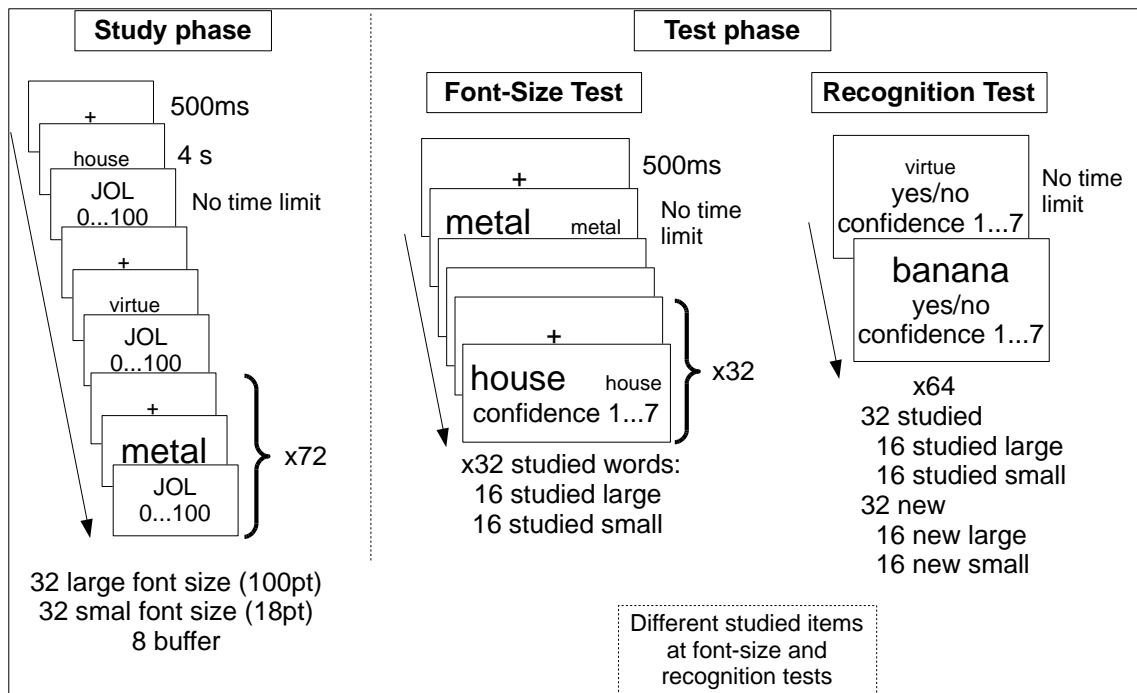
53  
54 The experiment was programmed in LiveCode (version 9.5.1), and participants  
55  
56 completed the experiment in individual soundproof booths. Executables and code from  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 the original experiments are available at the OSF. All responses were provided by  
3  
4 selecting the appropriate radio buttons. First, participants provided informed consent  
5  
6 and basic demographic information. Then, they read the instructions that encouraged  
7  
8 them to pay attention to words because later there would be a memory test. To increase  
9  
10 the chances that participants encoded font size and later recollected it during the font-  
11  
12 size test, instructions mentioned that they would have to remember the item and the font  
13  
14 size in which it was presented.  
15  
16  
17

18  
19 The experiment included a study phase and a test phase with two memory tests.  
20  
21 Figure 1 shows a summary of the procedure. In the study phase, each trial started with a  
22  
23 fixation cross presented for 500 milliseconds, followed by the word for four seconds.  
24  
25 Words were presented in the centre of the screen, half in small (18 points) and half in  
26  
27 large font size (100 points), counterbalanced. After each word, participants rated the  
28  
29 chances that they would recall the word later on a scale from 0 to 100 in deciles. The  
30  
31 first and last four words were buffers and were not analysed. Words were presented in a  
32  
33 pseudo-random order with no more than two words in the same font size in a row. After  
34  
35 72 trials, participants completed a filler task of writing European cities for two minutes.  
36  
37  
38  
39  
40  
41  
42

### 43 **Figure 1**

44 *Summary of the procedure used in Experiments 1 and 2*  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



Note: Filler tasks and final question after the recognition test not shown. See text for details.

The test phase included a font-size and a standard recognition test, both without time limit. In the font-size test, instructions mentioned that 32 studied words would be presented again and that participants had to indicate the size in which they were presented. Words were presented in a different order than at study. After a fixation point of 500 ms, each word was presented in both the small (18 pt) and large (100 pt) font sizes on the same screen; radio buttons below allowed participants to indicate the font size at study. The position of the large and small words was counterbalanced across participants. Immediately below, participants indicated the confidence that they had selected the correct font size on a scale from 1 (extremely low) to 7 (extremely high), following McDonough and Gallo (2012). This scale allowed a clear separation of confidence and JOLs (on a percentage scale). After the font-size test, participants



1  
2 completed another filler task in which they listed fruits, mammals, trees, and fishes for  
3  
4 one minute each.  
5

6  
7 Then, participants completed a standard recognition test. The 32 studied words  
8  
9 not used in the font-size test and 32 new words (counterbalanced) were presented in  
10  
11 pseudo-random order, one at a time, with no more than two words of the same type in a  
12  
13 row. Studied words were presented in the same font size as during the study phase, and  
14  
15 half of the new words were presented in small (18 pt) and half in large font size (100  
16  
17 pt). Instructions mentioned that font size of studied words corresponded to that during  
18  
19 the study phase. This served to avoid participants selecting “no” for a word they  
20  
21 recognized but they thought it was studied in a different font. Participants answered  
22  
23 yes/no the word was studied. On the same screen, participants also rated their  
24  
25 confidence that they selected the correct answer on the same scale from 1 to 7 as in the  
26  
27 font-size test.  
28  
29  
30  
31  
32

33  
34 The order of the tests was not counterbalanced for two reasons. First, our main  
35  
36 objective was better tested with the font-size test and the recognition test was included  
37  
38 only for comparison purposes. Thus, it was important to prevent the secondary  
39  
40 recognition test from interfering with the main font-size test, which could happen in the  
41  
42 recognition-test-first order (see McDonough & Gallo, 2012, for a similar justification).  
43  
44 And, second, administering the recognition-test-first order would increase the delay  
45  
46 between the study phase and the font-size test, which may make recollection of the font  
47  
48 size too difficult due to forgetting or interference from the new items in the recognition  
49  
50 test.  
51  
52  
53

54  
55 Finally, to check participants’ beliefs about the effect of font size on memory,  
56  
57 they answered the following question, partially based on Mueller and Dunlosky (2017,  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 Experiment 5): “In this experiment, you read words in small and large font size. Which  
3  
4 font size do you think leads to a better memory?”. The three alternative responses were:  
5  
6

7 - I think that, in general, words in large font size are better remembered.

8  
9 - I think that, in general, words in small font size are better remembered.

10  
11  
12 - I think that, in general, they will be similarly remembered.  
13

## 14 **Results**

15  
16 We expected that font size would affect confidence, but it may not. To get  
17  
18 evidence in support or against our hypotheses, the analyses of confidence directly  
19  
20 related to our key hypotheses were conducted with both the NHST and Bayesian  
21  
22 perspectives. For the NHST analyses, alpha was set to .05. For the Bayesian analyses,  
23  
24 we compared the hypothesis of differences between conditions (H1) against the  
25  
26 hypothesis of no differences (H2). When the reported *BFs* are higher than 1 they support  
27  
28 H1, and when lower they support H2. We used the default Cauchy prior and the  
29  
30 standard criteria of 0.33, 1, 3, and so on to determine the strength of the evidence in  
31  
32 favour of H1 or H2. All the analyses were conducted with R (R Core Team, 2018) and a  
33  
34 copy of the data is available at the OSF.  
35  
36  
37  
38  
39  
40

### 41 ***Question about Beliefs***

42  
43 Seventy-four percent of participants indicated that they thought they would  
44  
45 remember better words in large font, 4% that they would remember better words in  
46  
47 small font, and 22% that they would remember both equally. These results confirmed  
48  
49 that most participants held the belief that words in large font are better remembered.  
50  
51

### 52 ***Judgments of Learning***

53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Participants attributed higher JOLs to words in large than in small font,  $t(71) = 9.43, p < .001, d_{av} = 0.87$  (see Table 1). This result replicates the typical font size effect in JOLs extensively reported before.

**Table 1**

*Mean (SD) for JOLs in Experiments 1 and 2*

	Small	Large	Total
E1	37.94 (16.81)	52.72 (17.07)	
E2 – Better Small	38.52 (15.47)	43.79 (15.73)	41.16 (14.67)
E2 – Better Large	32.23 (16.11)	51.07 (16.01)	41.65 (14.10)
E2 – Total	35.35 (16.06)	47.46 (16.23)	

*Note:* E1: Experiment 1; E2: Experiment 2; JOLs on a scale from 0 to 100%.

### **Font-Size Test**

We conducted two analyses, one with the proportion of correct identifications of the font size at study and another with confidence in the response. First, participants identified better the font size when the words were studied in small font ( $M = .78, SD = .13$ ) than in large font ( $M = .66, SD = .18$ ),  $t(71) = 4.36, p < .001, d_{av} = 0.70$ . Performance in both conditions was better than chance, for small  $t(71) = 18.13, p < .001, d_{av} = 2.14$ , and for large,  $t(71) = 7.59, p < .001, d_{av} = 0.89$ .

**Table 2**

*Mean (SD) for Confidence in the Font Size Test in Experiments 1 and 2*

	Answer “Small”	Answer “Large”
<b>Experiment 1</b>		
Correct Responses	4.44 (0.99)	5.14 (0.99)

Incorrect Responses	3.68 (1.04)	4.34 (1.30)
<b>Experiment 2 – Better Large</b>		
Correct Responses	4.43 (1.10)	5.07 (1.17)
Incorrect Responses	3.53 (1.06)	3.95 (1.29)
<b>Experiment 2 – Better Small</b>		
Correct Responses	4.27 (1.19)	4.62 (1.17)
Incorrect Responses	3.40 (1.21)	3.60 (1.34)

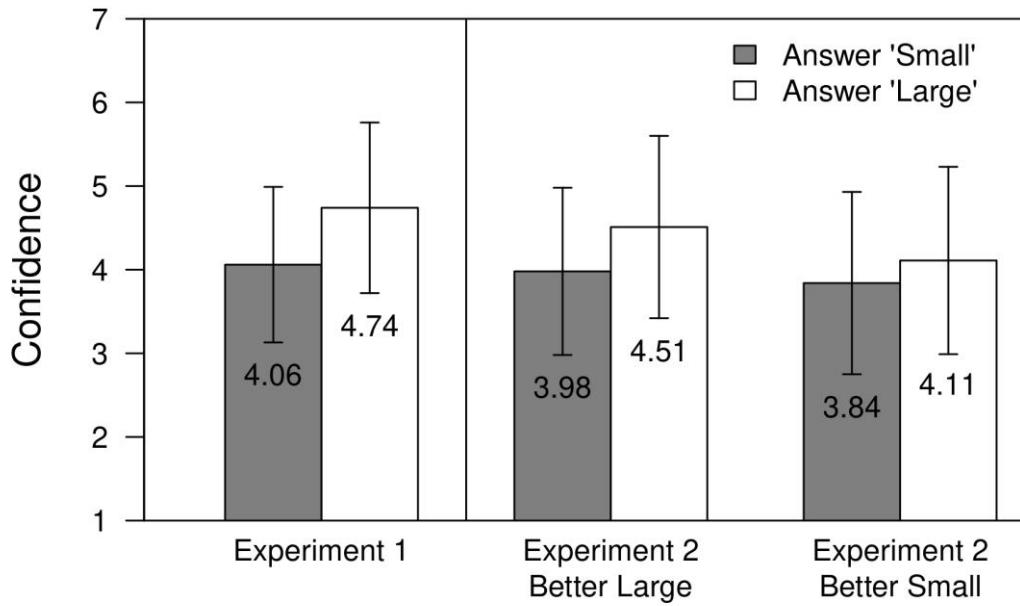
---

*Note:* Confidence on a scale from 1 to 7.

Second, for confidence we conducted an analysis of variance 2 (type of answer: “small”, “large”) x 2 (type of response: correct, incorrect). See Figure 2 and Table 2. Six participants were lost for this analysis because they did not have data in all the cells. Confidence was higher for answers “large” than “small”,  $F(1, 65) = 30.29, p < .001, \eta_p^2 = .32, BF_{inc} = 7.71 \times 10^7$ , and for correct ( $M = 4.79, SD = 0.83$ ) than incorrect responses ( $M = 4.01, SD = 0.99$ ),  $F(1, 65) = 73.11, p < .001, \eta_p^2 = .53, BF_{inc} = 1.73 \times 10^{10}$ . The interaction was not significant,  $F(1, 65) = 0.04, p = .839, \eta_p^2 < .01$ , with an inconclusive  $BF_{inc} = 0.763$ . In sum, both NHST and Bayesian analyses were consistent in that font size affected confidence in the responses in the font-size test.

## Figure 2

*Confidence in the Font Size Test in Experiments 1 and 2. Error bars indicate SD*



**Recognition Memory Test**

We also conducted analysis of memory performance and confidence. The main descriptive statistics for measures of memory performance are presented in Table 3. There were more hits with items in large than in small font,  $t(71) = 3.29, p = .002, d_{av} = 0.34$ , no differences in false alarms,  $t(71) = 0.85, p = .396, d_{av} = 0.10$ , and better accuracy, measured with  $d'$ , for words in large than small font,  $t(71) = 2.77, p = .007, d_{av} = 0.28$ .

**Table 3**

*Mean (SD) for Hits, False Alarms, and  $d'$  in the Recognition Test in Experiments 1 and*

*2*

	Hits		False Alarms		$d'$	
	Small	Large	Small	Large	Small	Large
E1	.78 (.16)	.83 (.13)	.28 (.15)	.26 (.14)	1.55 (0.74)	1.76 (0.74)

E2 – Better Small	.77 (.17)	.80 (.14)	.28 (.17)	.25 (.16)	1.50 (0.74)	1.70 (0.73)
E2 – Better Large	.80 (.14)	.82 (.12)	.26 (.16)	.26 (.17)	1.69 (0.76)	1.73 (0.77)

Note: E1: Experiment 1; E2: Experiment 2.

For confidence, we conducted an analysis of variance 2 (font size: small, large) x 2 (type of response: correct, incorrect).<sup>3</sup> Two participants were lost because they did not have data in all the cells. Results suggested that confidence was higher for words studied in large than in small font size, but the result was not particularly clear,  $F(1, 69) = 5.68, p = .020, \eta_p^2 = .08$ , with an anecdotal  $BF_{inc} = 2.04$  (see Table 4). Confidence was higher for correct than incorrect responses,  $F(1, 69) = 178.61, p < .001, \eta_p^2 = .72, BF_{inc} = 2.91 \times 10^{31}$ , and the interaction was not significant,  $F(1, 69) = 2.50, p = .119, \eta_p^2 = .03$ , with an inconclusive  $BF_{inc} = 1.08$ . In sum, the effect of font size on confidence in the recognition test was small and barely noticeable, but the results were in the same direction of those from the font-size test.

**Table 4**

*Mean (SD) for Confidence in the Recognition Test in Experiments 1 and 2*

	Small Font Size	Large Font Size	Total
<b>Experiment 1</b>			
Correct Responses	5.02 (0.81)	5.24 (0.82)	5.13 (0.79)
Incorrect Responses	4.20 (0.99)	4.27 (1.00)	4.23 (0.91)
Total	4.61 (0.85)	4.76 (0.83)	
<b>Experiment 2 – Better Large</b>			
Correct Responses	4.99 (1.00)	5.20 (0.94)	5.10 (0.94)

<sup>3</sup> We registered a 2 (type of response: “small”, “large”) x 2 (type of response: correct, incorrect) similar to that of the font-size test, but later realized that there are no responses “small” or “large” in this test. To follow the analyses in previous research, we replaced type of response by font size.

Incorrect Responses	4.19 (1.14)	4.21 (1.13)	4.20 (1.07)
Total	4.59 (1.01)	4.71 (0.98)	4.65 (0.97)
<b>Experiment 2 – Better Small</b>			
Correct Responses	5.12 (0.90)	5.24 (0.86)	5.18 (0.86)
Incorrect Responses	4.03 (1.25)	4.23 (1.20)	4.13 (1.12)
Total	4.57 (1.00)	4.74 (0.95)	4.65 (0.93)
<b>Experiment 2 – Total</b>	4.58 (1.00)	4.72 (0.96)	

---

*Note:* Confidence on a scale from 1 to 7.

To further test the effect of font size in retrospective confidence, we also analysed confidence for hits and false alarms. For hits, results suggested a higher confidence for words in large ( $M = 6.06$ ,  $SD = 0.88$ ) than small font size ( $M = 5.90$ ,  $SD = 0.85$ ),  $t(71) = 2.07$ ,  $p = .042$ ,  $d_{av} = 0.18$ , but with an inconclusive  $BF = 0.96$ . For false alarms, both analyses were consistent in showing no effect of font size on confidence, with moderate evidence in support of no differences,  $t(71) = 1.05$ ,  $p = .297$ ,  $d_{av} = 0.08$ ,  $BF = 0.22$  (large  $M = 4.21$ ,  $SD = 1.69$ ; small  $M = 4.09$ ,  $SD = 1.69$ ).

### ***Exploratory Analyses***

The analyses in this section were not registered. The results above suggest that font size affects retrospective confidence in the font-size test and, to a lesser degree, also in the recognition test. However, these effects could be driven by beliefs or by other factors. One way to get information on the role of beliefs is to compare the effect of font size in participants believing and not believing that large words would be better remembered. If beliefs drive the observed effects of font size on confidence, then the effects should be reduced or eliminated for participants without the belief. We conducted a 2 (belief: yes, no) x 2 (type of answer: “small”, “large”) analysis of variance with confidence in the font-size test and found a significant interaction,  $F(1,$

1  
2 70) = 9.52,  $p = .003$ ,  $\eta_p^2 = .12$ ,  $BF_{inc} = 18.23$ . In the belief group, the effect of font size  
3  
4 on confidence was significant (large  $M = 5.11$ ,  $SD = 1.05$ ; small  $M = 4.22$ ,  $SD = 1.05$ ),  
5  
6  $t(52) = 6.69$ ,  $p < .001$ ,  $d_{av} = 0.85$ ,  $BF = 8.28 \times 10^5$ , but it was not for participants in the  
7  
8 no-belief group (large  $M = 4.77$ ,  $SD = 0.84$ ; small  $M = 4.67$ ,  $SD = 0.71$ ),  $t(18) = 0.47$ ,  $p$   
9  
10 = .645,  $d_{av} = 0.13$ ,  $BF = 0.26$ .  
11  
12

13  
14 Also, in the recognition test a 2 (belief: yes, no) x 2 (font size: small, large)  
15  
16 analysis with confidence showed the same results. The interaction was significant,  $F(1,$   
17  
18  $70) = 5.42$ ,  $p = .023$ ,  $\eta_p^2 = .07$ ,  $BF_{inc} = 6.10$ , and the effect of font size on confidence  
19  
20 was significant in the belief group (large  $M = 5.01$ ,  $SD = 0.86$ ; small  $M = 4.73$ ,  $SD =$   
21  
22  $0.90$ ),  $t(52) = 4.84$ ,  $p < .001$ ,  $d_{av} = 0.33$ ,  $BF = 1.58 \times 10^3$ , and not significant in the no-  
23  
24 belief group (large  $M = 5.27$ ,  $SD = 0.77$ ; small  $M = 5.24$ ,  $SD = 0.64$ ),  $t(18) = 0.28$ ,  $p =$   
25  
26  $.785$ ,  $d_{av} = 0.03$ ,  $BF = 0.25$ . In sum, these analyses suggest that font size affects  
27  
28 confidence only for participants believing that large words are better remembered.  
29  
30

### 31 32 33 34 **Discussion**

35  
36 Experiment 1 replicated the effect of font size on JOLs and showed an effect of  
37  
38 font size on confidence in the font-size test. In the recognition test, results were in the  
39  
40 same line but were less clear. Existing metamemory beliefs could explain these results,  
41  
42 an interpretation supported by the exploratory analyses. However, font size could also  
43  
44 affect confidence through experience-based factors. With the font-size test, we tried to  
45  
46 reduce experiential cues, but they may still provide a reasonable explanation. We tested  
47  
48 the influence of beliefs in the effect of font size on confidence in Experiment 2.  
49  
50  
51

### 52 53 54 **Experiment 2**

55  
56 In Experiment 2, we manipulated the strength of participants' beliefs by  
57  
58 presenting information either in support of a better memory for words in large or in  
59  
60  
61  
62  
63  
64  
65



1  
2 small font size (a counter-belief). If beliefs drive the effect of font size on confidence,  
3  
4 then we expected that the counter-belief would reduce, eliminate, or revert that effect.  
5  
6

## 7 **Method**

### 8 *Participants and Design*

9  
10 Experiment 2 was a 2 (font size: small, 18 points; large, 100 points) 2  
11  
12 (instructions: better memory for small font size, better memory for large font size)  
13  
14 mixed design, with instructions manipulated between participants. In Experiment 1, we  
15  
16 found moderate to large effects of font size in confidence (font-size test  $\eta_p^2 = .32$ , and  
17  
18 recognition  $\eta_p^2 = .08$ ), but our hypotheses pertained to the interaction. Thus, we  
19  
20 determined the sample size to find a medium effect in the interaction. We registered the  
21  
22 use of GPower (Faul et al., 2007) to compute sample size, but after reading Brysbaert  
23  
24 (2019), we were unconvinced by the result that a total of 34 participants were needed to  
25  
26 get power = .80 in the interaction in a 2x2 mixed design. Brysbaert (2019) suggests 130  
27  
28 participants (see Table 7 of Brysbaert, 2019). This result is consistent with other power  
29  
30 analysis software, MorePower (Campbell & Thompson, 2012) and WebPower (Zhang &  
31  
32 Yuan, 2018). Thus, we aimed at a sample size of 130 with 65 participants in each group,  
33  
34 but we ended up collecting more participants to match the number of participants by  
35  
36 counterbalance group. A total of 145 participants completed the experiment in exchange  
37  
38 for course credits (117 female, age  $M = 20.75$ ,  $SD = 4.15$ ), with participants randomly  
39  
40 allocated to each instruction (72 to the better small and 73 to the better large).  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

### 51 *Materials and Procedure*

52  
53 Experiment 2 used the same materials and procedure as in Experiment 1, with the  
54  
55 exception that the instructions at the start of the experiment and before the font-size test  
56  
57 presented a brief justification for the better memory for words in small or large font.  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 The instructions were based on Blake and Castel (2018, Experiment 2): “Research has  
3 shown that, for college-age participants, words in [larger/smaller] fonts are easier to  
4 recall than words in [smaller/larger] fonts”.

## 5 6 7 8 9 **Results**

### 10 ***Questions about Beliefs***

11  
12 In the group with better large instructions, 55% of the participants indicated that  
13 they thought they would remember better words in large font, 4% that they would  
14 remember better words in small font, and 41% that they would remember them equally  
15 well. In the group with better small instructions, 47% thought that they would remember  
16 better words in large font, 10% that they would remember better words in small font,  
17 and 42% that they would remember them equally. These results suggest that our  
18 manipulation of beliefs had an effect, albeit small, on participants’ beliefs.

### 19 ***Judgments of Learning***

20  
21 An analysis of variance 2 (font size: small, large) x 2 (instructions: better small,  
22 better large) showed higher JOLs for words in large than small font size,  $F(1, 143) =$   
23  $121.24, p < .001, \eta_p^2 = .46$ , no effect of instructions,  $F(1, 143) = 0.04, p = .838, \eta_p^2 <$   
24  $.01$ , and a significant interaction,  $F(1, 143) = 38.13, p < .001, \eta_p^2 = .21$  (see Table 1).  
25 The effect of font size on JOLs was present in both groups, for the better small  $t(71) =$   
26  $4.21, p < .001, d_{av} = 0.34$  and for the better large  $t(72) = 10.46, p < .001, d_{av} = 1.17$ , and  
27 the effect was smaller in the former than in the latter,  $t(143) = 6.17, p < .001, d_{av} = 1.03$ .  
28 Consistent with the results above, JOLs showed that the manipulation of the beliefs was  
29 successful, but that the effect was small because the counter-belief did not eliminate the  
30 font size effect.

### 31 ***Font-Size Test***

32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 For memory performance, we found better memory for the font size of words  
3  
4 presented in small ( $M = .76$ ,  $SD = .14$ ) than in large font size ( $M = .63$ ,  $SD = .16$ ),  $F(1,$   
5  $143) = 60.33$ ,  $p < .001$ ,  $\eta_p^2 = .30$ . We found no effect of instructions,  $F(1, 143) = 0.24$ ,  $p$   
6  $= .623$ ,  $\eta_p^2 < .01$ , nor the interaction,  $F(1, 143) = 0.84$ ,  $p = .361$ ,  $\eta_p^2 = .01$ . Performance  
7 was better than chance for both small  $t(144) = 22.68$ ,  $p < .001$ ,  $d_{av} = 1.88$ , and large  
8 words,  $t(144) = 9.47$ ,  $p < .001$ ,  $d_{av} = 0.79$ .

9  
10  
11  
12  
13  
14  
15  
16  
17 As per confidence, we conducted a 2 (type of answer: “small”, “large”) x 2 (type  
18 of response: correct, incorrect) x 2 (instructions: better small, better large). Twelve  
19 participants were lost for this analysis. Confidence was higher for answers “large” ( $M =$   
20  $4.31$ ,  $SD = 1.12$ ) than “small” ( $M = 3.91$ ,  $SD = 1.04$ ),  $F(1, 131) = 27.24$ ,  $p < .001$ ,  $\eta_p^2 =$   
21  $.17$ ,  $BF_{inc} = 3.93 \times 10^{15}$ , and for correct ( $M = 4.60$ ,  $SD = 1.07$ ) than incorrect responses  
22 ( $M = 3.62$ ,  $SD = 1.05$ ),  $F(1, 131) = 204.69$ ,  $p < .001$ ,  $\eta_p^2 = .61$ ,  $BF_{inc} = 4.74 \times 10^{82}$  (see  
23 Table 2). There was no effect of instructions,  $F(1, 131) = 2.51$ ,  $p = .115$ ,  $\eta_p^2 = .02$ ,  $BF_{inc}$   
24  $= 1.55$ . These results replicate those from Experiment 1 in which font size affected  
25 confidence.

26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39 The interaction between font size and instructions was not significant  $F(1, 131) =$   
40  $2.67$ ,  $p = .105$ ,  $\eta_p^2 = .02$ , but the Bayesian analysis showed moderate evidence in  
41 support of differences,  $BF_{inc} = 4.44$  (see Figure 2). This interaction tested our main  
42 hypothesis, so we conducted further analyses to explore it. With both instructions,  
43 confidence was higher for words in large than small font size. With the better large  
44 instructions the effect size was medium and the  $BF$  showed extreme evidence in support  
45 of differences,  $t(66) = 6.52$ ,  $p < .001$ ,  $d_{av} = 0.58$ ,  $BF = 1.06 \times 10^6$ . With the better small  
46 instructions the effect size and  $BF$  were smaller at a descriptive level,  $t(65) = 3.38$ ,  $p =$   
47  $.001$ ,  $d_{av} = 0.32$ ,  $BF = 21.24$ . However, the difference between instructions was not large  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 enough as to show in the statistical comparisons,  $t(131) = 1.87$ ,  $p = .063$ ,  $d_{av} = 0.33$ ,  
3  
4 including an inconclusive  $BF = 0.91$ . These results hint that the counter-belief may have  
5  
6 reduced the effect of font size on confidence, but that effect was not particularly clear.  
7  
8  
9 This goes in line with the small effect of the beliefs manipulation.  
10

### 11 ***Recognition Memory Test***

12  
13 For memory performance, there were no differences in hits and false alarms, but  
14  
15 accuracy measured with  $d'$  was higher with words in large ( $M = 1.71$ ,  $SD = 0.75$ ) than  
16  
17 in small font size ( $M = 1.59$ ,  $SD = 0.76$ ),  $F(1, 143) = 4.38$ ,  $p = .038$ ,  $\eta_p^2 = .03$  (see Table  
18  
19 3).  
20  
21  
22  
23

24 For confidence, we conducted an analysis of variance 2 (font size: small, large) x  
25  
26 2 (type of response: correct, incorrect) x 2 (instructions: better small, better large). Four  
27  
28 participants were lost for this analysis. Confidence was higher for words in large ( $M =$   
29  
30  $4.72$ ,  $SD = 0.96$ ) than in small font ( $M = 4.58$ ,  $SD = 1.00$ ),  $F(1, 139) = 10.87$ ,  $p = .001$ ,  
31  
32  $\eta_p^2 = .07$ ,  $BF_{inc} = 1.07 \times 10^3$ , and also higher for correct ( $M = 5.14$ ,  $SD = 0.90$ ) than  
33  
34 incorrect responses ( $M = 4.17$ ,  $SD = 1.09$ ),  $F(1, 139) = 316.22$ ,  $p < .001$ ,  $\eta_p^2 = .69$ ,  $BF_{inc}$   
35  
36  $= 6.10 \times 10^{150}$  (see Table 4). There was no effect of instructions,  $F(1, 139) < 0.01$ ,  $p =$   
37  
38  $.983$ ,  $\eta_p^2 < .01$ ,  $BF_{inc} = 0.38$ , and no interactions.  
39  
40  
41  
42  
43

44 We also analysed confidence for hits and false alarms. A 2 (font size: small, large)  
45  
46 x 2 (instructions: better small, better large) analysis of variance showed higher  
47  
48 confidence for hits with words in large font size ( $M = 6.01$ ,  $SD = 0.85$ ) than small ( $M =$   
49  
50  $5.85$ ,  $SD = 0.93$ ),  $F(1, 143) = 11.82$ ,  $p = .001$ ,  $\eta_p^2 = .08$ ,  $BF_{inc} = 20.89$ , and no other  
51  
52 effects. A similar analysis with confidence in false alarms showed no effects. In  
53  
54 particular, font size did not affect confidence in false alarms with strong evidence in  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 support of no differences,  $F(1, 143) = 0.09, p = .760, \eta_p^2 < .01, BF_{inc} = 0.09$  (for words  
3  
4 in large font,  $M = 4.23, SD = 1.60$ , and for words in small font,  $M = 4.18, SD = 1.78$ ).  
5  
6

### 7 *Exploratory analyses*

8  
9 The analyses in this section were not registered. We replicated here the analysis  
10  
11 of Experiment 1. Seventy-five participants believed that memory would be better for  
12  
13 words in large font size and 70 believed otherwise. For the font-size test, we conducted  
14  
15 a 2 (belief: yes, no) x 2 (type of answer: “small”, “large”) analysis of variance with  
16  
17 confidence and found a significant interaction,  $F(1, 143) = 22.14, p < .001, \eta_p^2 = .13$ ,  
18  
19  $BF_{inc} = 9.11 \times 10^3$ . The effect of font size on confidence was significant in the belief  
20  
21 group (large  $M = 4.99, SD = 0.93$ ; small  $M = 4.19, SD = 1.01$ ),  $t(74) = 7.89, p < .001$ ,  
22  
23  $d_{av} = 0.83, BF = 4.31 \times 10^8$ , and not significant in the no-belief group (large  $M = 4.25, SD$   
24  
25  $= 1.20$ ; small  $M = 4.08, SD = 1.15$ ),  $t(69) = 1.79, p = .079, d_{av} = 0.14, BF = 0.59$ . The  
26  
27 difference in confidence for answers large and small was higher in the belief group than  
28  
29 in the no-belief group,  $t(143) = 4.70, p < .001, d_{av} = 0.78, BF = 2.75 \times 10^3$ .  
30  
31  
32  
33  
34  
35

36 For the recognition test, a 2 (belief: yes, no) x 2 (font size: small, large) also  
37  
38 showed an interaction,  $F(1, 143) = 6.18, p = .014, \eta_p^2 = .04, BF_{inc} = 3.11$ , with similar  
39  
40 results. The effect of font size on confidence was significant in the belief group (large  $M$   
41  
42  $= 5.08, SD = 0.86$ ; small  $M = 4.85, SD = 0.88$ ),  $t(74) = 5.41, p < .001, d_{av} = 0.27, BF =$   
43  
44  $2.00 \times 10^4$ , and not significant in the no belief group (large  $M = 4.99, SD = 1.02$ ; small  $M$   
45  
46  $= 4.91, SD = 1.07$ ),  $t(69) = 1.77, p = .081, d_{av} = 0.08, BF = 0.58$ . Also, the difference in  
47  
48 confidence between large and small words was higher in the belief group than in the no-  
49  
50 belief group,  $t(143) = 2.49, p = .014, d_{av} = 0.41, BF = 2.93$ .  
51  
52  
53  
54  
55

### 56 **Discussion**

57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 The results from Experiment 2 were consistent with those from Experiment 1:  
3  
4 Font size influenced confidence in both the font-size and the recognition tests. The  
5  
6 results also suggest that beliefs play a role in the effect, in line with past research with  
7  
8 JOLs (Luna, Nogueira, & Albuquerque, 2019). When we introduced the counter-belief  
9  
10 that words in small font size were recalled better, the effect of font size on confidence  
11  
12 was smaller, although, in general, the effect of the counter-belief was small. Also, the  
13  
14 results from the exploratory analyses showed that only participants who believe that  
15  
16 words in large font are better remembered rate words in large font with higher  
17  
18 confidence. The effect vanishes when there is no such belief.  
19  
20  
21  
22  
23

24 In Experiment 2, the proportion of participants believing that words in large font  
25  
26 are better remembered was numerically lower (55% and 47%) than in Experiment 1  
27  
28 (74%). This may be due to an unexpected reactivity to the information in Experiment 2  
29  
30 about the effect of font size on memory. That information may have drawn attention to  
31  
32 font size and made participants reflect on its likely impact on memory. Upon reflection,  
33  
34 some participants may have concluded that font size does not affect memory. To our  
35  
36 knowledge, this is the first time that reactivity effects to instructions manipulating  
37  
38 metamemory beliefs have been reported. Future research should try to replicate and  
39  
40 study that effect.  
41  
42  
43  
44  
45

### 46 **General Discussion**

47  
48 This research tested whether metamemory beliefs, known to affect prospective  
49  
50 judgments, also affect retrospective judgments. In particular, we tested whether the  
51  
52 belief that words in larger font size are better remembered also affected confidence.  
53  
54 The main results were that (1) font size affected confidence and that (2) the effect  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 depended on participants' beliefs. We discuss each result in turn and then comment on  
3  
4 memory results.  
5

6  
7 This research was the first to comprehensively study the effect of font size on  
8  
9 confidence. Metamemory judgments are inferential and based on cues, and we should  
10  
11 not take for granted that a cue affecting one type of judgment (e.g., JOL) would affect  
12  
13 another (e.g., confidence). Similarly, there is no good theoretical reason to assume that  
14  
15 cue utilisation and cue validity are similar between prospective and retrospective  
16  
17 metamemory judgments. In support of differences between prospective and  
18  
19 retrospective judgments, McDonough et al. (2021) found that monitoring at encoding  
20  
21 and retrieval may be different. Despite these arguments, we found that font size affects  
22  
23 both JOLs and confidence. This should not be interpreted as supporting that similar  
24  
25 cues, in general, affect both judgments, but instead that the specific manipulation here,  
26  
27 i.e., font size, is used as a cue to both JOLs and confidence. Other cues may have  
28  
29 different effects, and we join McDonough et al. (2021) in suggesting that future research  
30  
31 should include both prospective and retrospective metamemory measures to better  
32  
33 understand the monitoring process(es).  
34  
35  
36  
37  
38  
39  
40

41 More relevant, this research showed that beliefs also affect confidence.  
42  
43 Participants who did not believe that large fonts were better remembered did not rate  
44  
45 words in large font size with higher confidence than words in small font size. Similarly,  
46  
47 the presentation of a counter-belief seemingly reduced the effect of font size on  
48  
49 confidence, although this effect was not particularly clear because, in general, the  
50  
51 counter-belief had a small effect on confidence. These results suggest that the debate on  
52  
53 the role of theory-based and experience-based factors should be extended and  
54  
55 encompass retrospective judgments. We make no claims on the contribution of  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 experiential factors in the font-size effect on confidence because the experiments were  
3  
4 not designed to test them (but see below for some speculation). Instead, we would like  
5  
6 to stress the effect of beliefs, which suggests that theory-based factors are transversal to  
7  
8 both JOLs and confidence. Previous studies seeking an effect of font size on confidence  
9  
10 were not sensitive enough to find it, either because the belief may not have been active  
11  
12 at retrieval or because the font size at study may not have been retrieved and, thus,  
13  
14 ignored as a cue for the judgment. All these factors should be taken into account in  
15  
16 future research studying the effect of theory-based factors on confidence.  
17  
18  
19  
20  
21

22         Although we did not study nor measured experiential cues, some results may give  
23  
24 us a first insight into the weights of beliefs and experiential factors in our experiments.  
25  
26 In the font-size test, participants identified better words in small font size (i.e., answer  
27  
28 “small”) but rated their answers “large” with higher confidence. These results suggest  
29  
30 that, in this particular research setting, experiential factors during retrieval may have  
31  
32 been less relevant as cues for confidence than metamemory beliefs. However, the  
33  
34 counter-belief in Experiment 2 had a clear effect on JOLs, a barely noticeable effect on  
35  
36 the font-size test, and no effect on recognition, which suggests a higher role of beliefs  
37  
38 on JOLs and a lesser role on confidence. These results are apparently contradictory, but  
39  
40 are consistent with the idea that the belief must be active to influence metamemory  
41  
42 judgments. Participants read the counter-belief before the study phase and again before  
43  
44 the font-size test, but not before the recognition test. By then, most participants may  
45  
46 have forgotten about the counter-belief (which may also show in the small effect of the  
47  
48 counter-belief in the final question about beliefs). Our conclusion, largely speculative, is  
49  
50 that when beliefs are active, they may have a stronger role in the font-size effect on  
51  
52 confidence than experiential factors, but future research should test this idea.  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



1  
2 Memory was secondary to this research, and as font size usually has a small  
3  
4 effect on memory, we expected no effect of font size on memory. However, we  
5  
6 consistently found better accuracy with large words in the recognition test. Past meta-  
7  
8 analyses have shown that font size also affects memory (Chang & Brainerd, in press;  
9  
10 Luna et al., 2018), but it seems that the effect of font size on memory is no longer part  
11  
12 of the meta-analysis realm. Several individual experiments have reported similar effects  
13  
14 (e.g., Halamish, 2018; Undorf & Zimdahl, 2019). Thus, maybe it is time to reconsider  
15  
16 the repeated claim that font size does not affect memory and start acknowledging that it  
17  
18 has a consistent effect when statistical power is large enough.  
19  
20  
21  
22  
23

24 In a similar line, we did not expect font size to affect correct identifications in the  
25  
26 font-size test. However, we found better memory for words presented in small than  
27  
28 large font. We have two tentative explanations for this result. First, it may be due to  
29  
30 participants remembering better the font size for small words, with particular  
31  
32 mechanisms yet to be identified. And, second, it may be the consequence of some  
33  
34 compensatory strategy. If participants cannot retrieve that a word was presented in large  
35  
36 font, they could consider that they should have remembered the font size for a large  
37  
38 word because large words are better remembered. Then, they could turn and answer  
39  
40 “small”. Similar compensatory mechanisms have been proposed before in related  
41  
42 research (e.g., see the distinctiveness heuristic in McDonough & Gallo, 2012). In  
43  
44 support of this interpretation, participants reported more answers “small” than “large”  
45  
46 (Experiment 1: 56% vs. 44%; Experiment 2: 57% vs. 43%). These explanations are  
47  
48 highly speculative and should be tested in future research.  
49  
50  
51  
52  
53  
54

55 In sum, this research found that beliefs affect confidence in a similar way to  
56  
57 JOLs. This may be specific to the font size manipulation, but it suggests that research  
58  
59  
60  
61  
62  
63  
64  
65

1  
2 should pay more attention to theory-based factors in retrospective metamemory  
3  
4 judgments.  
5  
6  
7  
8  
9

### 10 **References**

- 11 Ball, B. H., Klein, K. N., & Brewer, G. A. (2014). Processing fluency mediates the  
12 influence of perceptual information on monitoring learning of educationally  
13 relevant materials. *Journal of Experimental Psychology: Applied*, *20*, 336-348.  
14  
15 Besken, M. (2016). Picture-perfect is not perfect for metamemory: Testing the  
16 perceptual fluency hypothesis with degraded images. *Journal of Experimental*  
17 *Psychology: Learning, Memory, and Cognition*, *42*, 1417-1433.  
18  
19 Besken, M., Solmaz, E. C., Karaca, M., & Atılgan, N. (2019). Not all perceptual  
20 difficulties lower memory predictions: Testing the perceptual fluency hypothesis  
21 with rotated and inverted object images. *Memory & Cognition*, *47*(5), 906-922.  
22  
23 Blake, A. B., & Castel, A. D. (2018). On belief and fluency in the construction of  
24 judgments of learning: Assessing and altering the direct effects of belief. *Acta*  
25 *Psychologica*, *186*, 27-38.  
26  
27 Brewer, W. F., & Sampaio, C. (2006). Processes leading to confidence and accuracy in  
28 sentence recognition: A metamemory approach. *Memory*, *14*, 540-552.  
29  
30 Brewer, W. F., Sampaio, C., & Barlow, M. R. (2005). Confidence and accuracy in the  
31 recall of deceptive and nondeceptive sentences. *Journal of Memory and*  
32 *Language*, *52*, 618-627.  
33  
34 Brysbaert, M. (2019). How many participants do we have to include in properly  
35 powered experiments? A tutorial of power analysis with reference tables. *Journal*  
36 *of Cognition*, *2*(1), 1-38.  
37  
38 Buratti, S., & Allwood, C. M. (2015). Regulating metacognitive processes—Support for  
39 a meta-metacognitive ability. In A. Peña-Ayala (Ed.), *Metacognition:*  
40 *Fundamentals, applications, and trends* (pp. 17-38). Springer.  
41  
42 Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with  
43 relational confidence intervals and Bayesian analysis. *Behavior Research*  
44 *Methods*, *44*(4), 1255-1265.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

- 1  
2 Champely, S. (2018). pwr: Basic functions for power analysis. R package version 1.2-2.  
3  
4 <https://CRAN.R-project.org/package=pwr>  
5
- 6 Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible  
7  
8 statistical power analysis program for the social, behavioral, and biomedical  
9  
10 sciences. *Behavior Research Methods*, *39*, 175-191.
- 11 Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs  
12  
13 jointly contribute to volume effects on metacognitive judgments. *Journal of*  
14  
15 *Experimental Psychology: Learning, Memory, and Cognition*, *43*, 680-693.
- 16 Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A  
17  
18 Brunswikian theory of confidence. *Psychological Review*, *98*, 506-528.
- 19 Halamish, V. (2018). Can very small size enhance memory? *Memory & Cognition*, *46*,  
20  
21 979-993.
- 22  
23  
24 JASP Team (2018). JASP (Version 0.9) [Computer software]. Available at: <https://jasp->  
25  
26 [stats.org](https://jasp-stats.org)
- 27 Jemstedt, A., Schwartz, B. L., & Jönsson, F. U. (2018). Ease-of-learning judgments are  
28  
29 based on both processing fluency and beliefs. *Memory*, *26*, 807-815.
- 30 Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2016). The effect of word  
31  
32 frequency on judgments of learning: Contributions of beliefs and processing  
33  
34 fluency. *Frontiers in Psychology*, *6*, 1995.
- 35  
36 Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of  
37  
38 retrieval as a basis for confidence in answers to general knowledge questions.  
39  
40 *Journal of Memory and Language*, *32*, 1-24.
- 41  
42 Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization  
43  
44 approach to judgments of learning. *Journal of Experimental Psychology: General*,  
45  
46 *126*, 349-370.
- 47  
48 Koriat, A. (2008). Subjective confidence in one's answers: The consensuality principle.  
49  
50 *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 945-  
51  
52 959.
- 53  
54 Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and  
55  
56 experience-based metacognitive judgments: Evidence from subjective confidence.  
57  
58 In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp.  
59  
60 117-135). Psychology Press.
- 61  
62  
63  
64  
65

- 1  
2 Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-  
3 processing heuristic and the stability bias: Dissociating memory, memory beliefs,  
4 and memory judgments. *Psychological Science*, *22*, 787-794.  
5  
6  
7  
8 Leippe, M. R., Eisenstadt, D., & Rausch, S. M. (2009). Cueing confidence in  
9 eyewitness identifications: Influence of biased lineup instructions and pre-  
10 identification memory feedback under varying lineup conditions. *Law and Human*  
11 *Behaviour*, *33*, 194-212.  
12  
13  
14  
15 Li, P., Jia, X., Li, X., & Li, W. (2016). The effect of animacy on metamemory. *Memory*  
16 *& Cognition*, *44*, 696-705.  
17  
18  
19 LiveCode (2017). *LiveCode (Version 8.1.5) [Computer Software]*. Edinburgh, UK:  
20 LiveCode.  
21  
22  
23 Luna, K., Albuquerque, P. B., & Martín-Luengo, B. (2019). Cognitive load eliminates  
24 the effect of perceptual information on judgments of learning with sentences.  
25 *Memory & Cognition*, *47*, 106-116.  
26  
27  
28 Luna, K., Martín-Luengo, B., & Albuquerque, P. B. (2018). Do delayed judgments of  
29 learning reduce metamemory illusions? A meta-analysis. *Quarterly Journal of*  
30 *Experimental Psychology*, *71*, 1626-1636.  
31  
32  
33  
34 Luna, K., Nogueira, M., & Albuquerque, P. B. (2019). Words in larger font are  
35 perceived as more important: Explaining the belief that font size affects memory.  
36 *Memory*, *27*, 555-560.  
37  
38  
39 McDonough, I. M., & Gallo, D. A. (2012). Illusory expectations can affect retrieval-  
40 monitoring accuracy. *Journal of Experimental Psychology: Learning, Memory,*  
41 *and Cognition*, *38*, 391-404.  
42  
43  
44 McDonough, I. M., Enam, T., Kraemer, K. R., Eakin, D. K., & Kim, M. (2021). Is there  
45 more to metamemory? An argument for two specialized monitoring abilities.  
46 *Psychonomic Bulletin & Review*, *28*(5), 1657-1667.  
47  
48  
49 Monds, L. A., Kloft, L., Sauer, J. D., Honan, C. A., & Palmer, M. A. (2019). No  
50 evidence that alcohol intoxication impairs judgments of learning in face  
51 recognition. *Applied Cognitive Psychology*, *33*(3), 325-333.  
52  
53  
54  
55 Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning:  
56 Evaluating analytic processing theory with beliefs about fluency. *Journal of*  
57 *Memory and Language*, *93*, 245-458.  
58  
59  
60  
61  
62  
63  
64  
65

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65
- 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 of reflect people's beliefs about memory? *Journal of Memory and Language, 70*, 1-12.
- Perfect, T. J. (2004). The role of self-rated ability in the accuracy of confidence judgements in eyewitness memory and general knowledge. *Applied Cognitive Psychology, 18*, 157-168.
- Peynircioğlu, Z. F., & Tatz, J. R. (2019). Intensifying the intensity illusion in judgments of learning: Modality and cue combinations. *Memory & Cognition, 47*, 412-419.
- 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.
- R Core Team (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*, 615-625.
- Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review, 16*, 550-554.
- Robinson, M. D., Johnson, J. T., & Robertson, D. A. (2000). Process versus content in eyewitness metamemory monitoring. *Journal of Experimental Psychology: Applied, 6*, 207-221.
- Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019). Metacognitive expectancy effects in source monitoring: Beliefs, in-the-moment experiences, or both? *Journal of Memory and Language, 107*, 95-110.
- 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*, 97-109.
- Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. *Applied Cognitive Psychology, 20*, 17-31.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65
- Witherby, A. E., & Tauber, S. K. (2017). The concreteness effect on judgments of learning: Evaluating the contributions of fluency and beliefs. *Memory & Cognition*, *45*, 639-650.
- Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priory theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*, 918-933.
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1341-1354.
- Zhang, Z., & Yuan, K.-H. (Eds.). (2018). *Practical statistical power analysis using WebPower and R*. ISDSA Press. <https://webpower.psychstat.org>

