



Deceitful Hints: a Meta-Analytic Review of the Part-List Cuing Impairment in Recall

Nicholas W. Pepe¹ · Anne Moyer¹ · Tori Peña¹ · Suparna Rajaram¹

Accepted: 22 February 2023 / Published online: 14 March 2023
© The Psychonomic Society, Inc. 2023

Abstract

A large body of research in the study of memory has accumulated to date on the part-list cuing impairment in recall. This phenomenon refers to the lower recall of studied information in the presence of some studied words provided as retrieval cues compared to when no cues are provided. We review the current literature on the part-list cuing impairment in recall and report a meta-analysis utilizing the procedural and statistical information obtained from 109 samples ($N = 5,605$). In each experiment, participants studied a list of words and subsequently performed a recall task either in the presence or absence of part-list cues. The meta-analysis shows that the part-list cuing impairment is a robust, medium-sized impairment (Cohen, 1988). This recall impairment was not significantly sensitive to the number of study items provided, the relationship among study items, the number of part-list cues provided, the amount of time provided for recall, or certain other factors of interest. Our analyses also demonstrate that longer retention periods between study and retrieval mitigate the part-list cuing impairment in recall. We discuss the implications of meta-analysis results for elements of experimental design, the findings of past literature, as well as the underlying theoretical mechanisms proposed to account for this impairment in recall and the applied consequences of this recall impairment.

Keywords Part-list cues · Recall · Memory impairment · Meta-analysis

Introduction

Remembering is a finicky process. We often forget things that are important to us, remember things that we would like to forget, and struggle to recollect things at a moment's notice, only to remember them later when they are not needed. The topic of this review, the *part-list cuing impairment* in recall, is one such memory phenomenon, one that is deceptive because it suggests hints or cues should benefit memory, but shows them to be detrimental for the rememberer.

It is intuitive to think that cues benefit recall of studied information. In fact, ample research has demonstrated the basic principle that a little assistance can aid recollections that might not have been otherwise remembered (Hudson & Austin, 1970; Robin & Moscovitch, 2017; Tulving, 1974). However, there is such a thing as having too many cues.

Although this situation may seem counterintuitive, it can reduce recall of remaining information. These situations arise during everyday tasks such as using a partial grocery or to-do list as well as in more consequential circumstances such as providing examples for a writing prompt on a standardized test or questioning witnesses in trials. These examples all center around a counterintuitive memory phenomenon known as the *part-list cuing impairment in recall*.

First documented by Slamecka (1968) in a series of six experiments, the part-list cuing impairment occurs when the rememberer receives an abundance of cues to aid their recall. Specifically, participants who received half of the studied items as cues to aid recall (Experiment 1) remembered fewer non-cued studied words compared to participants who received no cues. One of the first explanations for this impairment sought to determine the threshold where the amount of cue words shift from being beneficial to being detrimental (Roediger III, 1973). This study compared the recall performance of participants by providing them with a variety of words from different categories. Regardless of the number of cues provided, performance was reliably worse for those who received part-list cues.

✉ Nicholas W. Pepe
nicholas.w.pepe@stonybrook.edu

¹ Department of Psychology, Stony Brook University,
Stony Brook, NY 11794-2500, USA

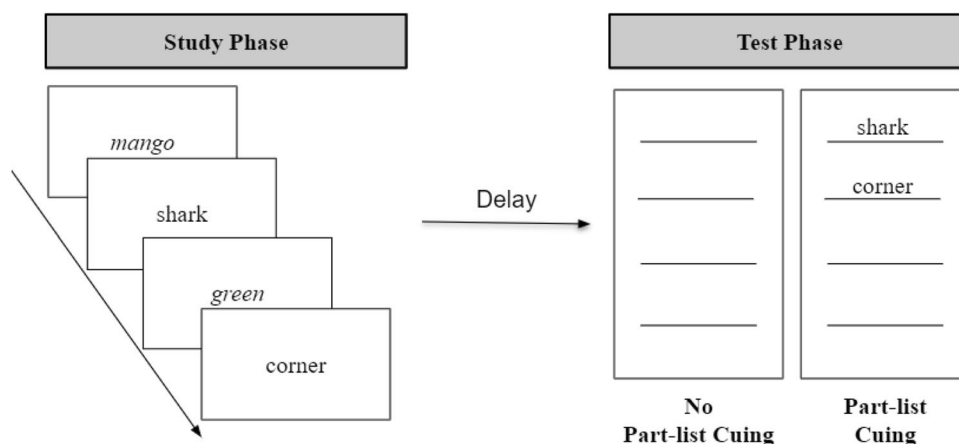


Fig. 1 A diagram of the standard experimental design used in part-list cuing experiments. For purposes of illustration, the to-be-remembered (i.e., target) words are italicized and the words to serve as cues

for participants in the part-list cuing condition are non-italicized in this example. Control (i.e., no part-list cuing condition) participants are asked to recall all studied items

These findings spurred numerous investigations into when and why the part-list cuing impairment occurs in memory. Findings that are intuitive refine our general understanding of the cognitive processes that underlie memory functions, but findings that are counterintuitive advance that understanding. When we observe counterintuitive outcomes such as the part-list cuing impairment, we must integrate these processes into our theories to account for these outcomes. Such findings provide deeper insights into how memory operates and ultimately help shape our theories on cognitive processing.

For the past 50 years, the part-list cuing impairment has been put to an array of experimental tests, and, as we describe in later sections, several theories have been developed to account for the emergence or absence of this phenomenon. For instance, across several studies, researchers have altered features of the basic experimental design and reported conditions where the size of the impairment differed (e.g., Aslan & Bäuml, 2009; Basden & Basden, 1995; Reysen & Nairne, 2002), as well as conditions where the impairment was absent (e.g., Bäuml & Samenieh, 2012; Cole et al., 2013; Serra & Oswald, 2006). As a result, a large body of empirical literature has accumulated on part-list cuing, demonstrating that a wide range of procedures can produce this memory impairment. Given the large number of manipulations applied when exploring this effect, our understanding of the impairment must account for instances when the effect is present as well as when it is not. As such, the time seems ripe to develop a quantitative and descriptive synthesis of the findings stemming from the tests of this impairment in recall.

Our research synthesis provides guidelines for when the part-list cuing impairment may occur and when it is absent. This analysis is particularly important for seeking

the generalizability of this phenomenon and the boundary conditions that constrain it. Furthermore, systematic information about what variables, study materials, and procedural conditions moderate the effect would be particularly valuable for future efforts in designing studies across populations and cultural contexts. Finally, clarity about when and how cues hurt or help accurate remembering have broad implications for the ways in which cues can be used in educational settings to help performance, in legal settings to question witnesses, and in daily life for managing day to day activities such as grocery shopping, medication routines, and handling appointments. With these goals in mind, our current research presents a meta-analytic review of past findings and offers a resource that can serve as a roadmap for future research. Where needed, we supplement the meta-analytic synthesis with a qualitative review of studies.

The Part-List Cuing Paradigm

The prototypical experimental design to test the part-list cuing impairment in recall is outlined in Fig. 1. Although studies sometimes deviate from this design and alter aspects of the procedure, this standard procedure has served as a foundation for those who seek to test this impairment in recall. As with explicit memory studies in general, in part-list cuing studies participants first receive intentional study instructions where they study a list of items for a later memory test. A few studies have also used incidental study instructions by not making a reference to the test phase (Peynircioğlu & Moro, 1995). As is customary in studies of long-term memory, a short distractor phase often follows, to prevent rehearsal of the studied material in working memory (Baddeley & Hitch, 1974).

The test phase follows next, and it typically consists of two key conditions. In the control condition, participants perform a free-recall task where they are instructed to recall all the studied items in any order. In the experimental condition, critical to the part-list cuing paradigm, participants receive a subset of the studied items as cues to aid recall of the remaining items. In some instances, participants are asked to read the cues aloud or to perform other types of checks to make sure that they are attending to the cues before they begin to recall (e.g., Mueller & Watkins, 1977; Rundus, 1973; Serra & Oswald, 2006).

The studied words of key interest are those that are not provided as cues and that participants are instructed to recall. We will refer to these non-cued studied items as *target items* in this review. These items are the main metric of comparison between the part-list cued and control conditions. Researchers use the number of target items recalled as the key measure (instead of the entire study list), because participants in the part-list cuing condition do not have the same opportunity as the control participants to recall the cued items, thus creating a fair comparison across conditions to compute the effects of part-list cues on recall.

A second, free-recall task is sometimes included as well (e.g., Basden & Basden, 1995; Bäuml & Aslan, 2006; Lehmer & Bäuml, 2018a), and this follow-up task is identical across both conditions. Placing this task at the end of the main study-test sequence allows for a comparison of the cascading effects of the part-list cues on recall that might exist even after removing the cues. As this second task does not influence performance on the first part-list cued task that is of interest in the present review, we will only discuss these results in the context of the theories where they are relevant. Thus, in this review we will focus on the results from the first set of recall tasks.

Range of Methods to Test Part-List Cuing Effects

Stimuli

In memory research, the number or the type of study and test items provided to participants constitutes one of the most common manipulations across experimental designs. In the part-list cuing literature, researchers have tested the effects of the number of cue items provided at test, the number of items at study, the extent of association between the cued items and target items, and the use of more complex stimuli compared to single words. We elaborate on each below.

The Number of Cue Words at Test Since the initial reporting of the part-list cuing impairment in recall, researchers have been interested in the effects of the number of cue words that

are provided to participants during recall (e.g., Roediger III, 1973). The number of cues provided should be thought of in two metrics - the absolute number of items being presented as cues (i.e., regardless of the study list length) and the proportion of study items presented as cues.

The part-list cuing impairment has been observed when participants receive as few as four cues (Basden et al., 2002; Bäuml & Aslan, 2004; Goernert & Larson, 1994; Peynircioğlu, 1987; Roediger III, 1973; Serra & Nairne, 2000; Watkins & Allender, 1987) to as many as 42 cues (Barber & Rajaram, 2011). In terms of the proportion of studied items, the impairment has been observed with a proportion as small as 11.11% of the studied items (Marsh et al., 2004, Experiment 3, older adult condition) to as large as 66% of the studied items (Bäuml & Samenieh, 2012; Bäuml & Schlichting, 2014; Garrido et al., 2012; Marsh et al., 2004, Experiments 1 and 2; Roediger III et al., 1977). Hence, the part-list cuing impairment occurs in the presence of a few cues, both in proportion and frequency, as well as in the presence of many.

Given that the part-list cuing impairment is reliably observed when presented with a wide range of cues, researchers have investigated whether the size of the impairment is sensitive to this variation. This question may seem straightforward, but the findings make the answer somewhat ambiguous. Some research, where the proportion of cues presented was manipulated (most commonly 33% vs. 66%), has yielded no significant differences between the conditions based solely upon the proportion of cues (Goernert & Larson, 1994; Watkins, 1975), but other studies have. The latter set includes studies using 33% versus 66% comparisons (Marsh et al., 2004, Experiments 1 and 2), as well as studies that varied the number of cues along different ranges (Roediger III, 1973; Roediger III et al., 1977; Rundus, 1973).

In addition to varying the number or the proportion of cues, altering other aspects of the stimuli such as the type of study stimuli or the overall number of study items, that is, the study list length, could account for the differences in observations. A study by Kimball and Bjork (2002) demonstrated one such interaction effect using the DRM stimuli (short for Deese, Roediger, and McDermott; Deese, 1959; Roediger III & McDermott, 1995). In studies using the DRM study lists, participants study a list of associatively related items (e.g., *pillow, bed, night, ...*), and show a high propensity for false recall of the non-presented, critical lure item (e.g., *sleep*). When participants studied items with strong inter-item associations, the number of cues presented at test had a significant impact on the size of the recall impairment; the recall of the weakest associates was lower for those provided with eight cues strongly associated to the critical lure compared to four cues strongly associated to the critical lure.

However, when the cued items had weak associations to the critical lure, the number of cues (either four or eight) did not significantly alter the recall of strong associates. In other words, a larger number of cues tend to increase the part-list cuing impairment in recall when the cues evoke strong associations.

Extra-List Cues When thinking about the reduction in recall associated with the presence of part-list cues, another fundamental question arises: Do the part-list cues necessarily have to be from the study list to lower recall? In other words, can the part-list cuing impairment occur with extra-list cues, that is, items not presented during the study phase?

To test this question, researchers have conducted experiments where the cues given at test were not present in the study list.¹ These studies have included both categorized and unrelated word lists, and while the findings are somewhat mixed across the range of studies, we do see evidence for a part-list cuing impairment when presented with extra-list cues. Findings with categorized word lists show that if the cue items were extra-list such that they were derived from a different category than the one being tested, these cues did not produce a significant impairment in recall (Mueller & Watkins, 1977, Experiment 1). Other studies show that when participants receive unrelated nouns extra-list cues can reduce recall, although the size of this impairment when using extra-list cues is not always on par with standard part-list cues (i.e., cued items that come from the study list; Andersson et al., 2006; Roediger III et al., 1977; Todres & Watkins, 1981; Watkins, 1975). For example, Andersson et al. (2006) and Roediger III et al. (1977) both reported smaller impairments for extra-list cues compared to standard, intra-list cues for study lists composed of unrelated nouns. Therefore, the part-list cuing impairment can still be present with extra-list cues, but the size of the impairment may be reduced in comparison to the cues derived from the study items.

Complex Stimuli Another extension investigating the generalizability of the part-list cuing impairment branches out from simple words to complex stimuli. These studies, bridging the findings from lab settings to more

ecologically valid information, show that the part-list cuing impairment also occurs for word-pairs (Muntean & Kimball, 2012; Riefer et al., 2002; Roediger III & Schmidt, 1980), specific words presented in sentences (Garcia-Marques et al., 2002; Garcia-Marques et al., 2012), and even passages of text (Bäumel & Schlichting, 2014; Fritz & Morris, 2015; Wallner & Bäumel, 2020).

Beyond the recollection of stimuli presented in word form, researchers have investigated the part-list cuing impairment for pictorially depicted stimuli. Bovee et al. (2009) presented words paired with related video clips (e.g., a short clip of a hand picking up seedless grapes accompanied by the text “*seedless grapes*”) as study stimuli. At test, the authors presented half the items in written word form to serve as part-list cues for recall and instructed participants to recall the remaining items. This study-test arrangement of picture-word stimuli also produced a part-list cuing impairment in recall. These findings extend the previous findings observed for words to study items that are pictorial. It should be noted, however, that the pairing of words with video clips and requiring word responses does not speak to situations where the study items are strictly pictorial. In this context, some studies have examined the impact of part-list cues in purely visual-pictorial tasks by utilizing study items and test cues that are both visual-pictorial. In these studies, participants are presented with visual arrays and then asked to reproduce these arrays at retrieval. The stimuli most frequently used in these experiments are colored snap-circuits (Cole et al., 2013; Kelley et al., 2016) and arrangements of chess pieces (Drinkwater et al., 2006; Huffman et al., 2001; Kelley et al., 2016; Watkins et al., 1984). These studies show that study material and part-list cues that are purely visuospatial do not produce the typical part-list cuing impairment on these reproduction tasks.

It is difficult to disentangle the extent to which an absence of the part-list cuing impairment in studies with visuospatial stimuli was due to the nature of stimuli or the nature of reproduction task (instead of a recall task), but it is also the case that these types of stimuli and memory tasks would be difficult to dissociate. As such, experiments involving visuospatial reconstruction tasks were not included in this meta-analysis. We next turn to the findings on the use of other memory tasks.

Other Memory Tasks

The part-list cuing impairment in memory was originally reported in recall and has been primarily tested for recall memory. Therefore, our quantitative review will focus on recall tasks. However, we briefly review other memory tasks to provide a fuller view of this literature about the extent

¹ Extra-list cues were used to examine specific hypotheses concerning the operation of principles such as cue overload (the number of targets that a cue can subsume for recall, e.g., Mueller & Watkins, 1977) and cue-target competition (competition of cues with targets at retrieval, e.g., Roediger, et al., 1977), accounts that guided early stages of research on part-list cuing. Given the small number of studies on extra-list cues, we report only on the empirical outcomes obtained with the use of extra-list cues for purposes of this review.

to which this impairment occurs in other types of memory tasks. Such undertakings have been reported for recognition memory, semantic memory, and implicit memory tasks.²

Unlike a free-recall task where participants try to remember as many items as possible in the absence of any cues, tests of recognition memory include items that were studied earlier along with items that were not studied, and participants try to identify the study status of each item. Not many part-list cuing investigations have used a recognition task. Among the few that did, the study material consisted of unrelated nouns and regardless of whether studied items or extra-list items are provided as cues during recognition, part-list cues negatively affected performance (Todres & Watkins, 1981; for similar findings on accuracy and reaction times, see Oswald et al., 2006).

Semantic memory is another domain of memory where researchers have investigated the part-list cuing impairment. In the standard episodic memory task, researchers provide the information at study that they test in a later memory task. In semantic memory tasks, the aim is to probe information presumably known to participants prior to their experiment participation and is not provided at study. Commonly used examples of such semantic information are the names of US states (Foos & Clark, 2000; Rhodes & Castel, 2008), celebrity names (Foos & Clark, 2000), and other content-limited categories such as the Zodiac signs (Kelley & Parihar, 2018). In most instances, since the topic material is common knowledge for those sampled (such as US states or famous celebrities), it is assumed that, on average, participants will have knowledge on these topics and differences in specific knowledge between participants will not affect the results due to random assignment. Participants exhibited impaired recall (i.e., when comparing accurate recall for non-cued items) for some categories but not for other categories in these semantic recall tasks (Foos & Clark, 2000; Kelley & Parihar, 2018); Foos and Clark (2000) did not observe a significant impairment when they tested the recall of the US states with part-list cues, but they found that part-list cues hurt the recall of celebrity names. In contrast, Rhodes and Castel (2008) found that part-list cues hurt the participants' ability to retrieve the target US states.

Varied effects of part-list cues for retrieving semantic information were similarly reported in a study that probed recall ability for semantic information from a variety of categories (Zodiac signs, campus locations, Harry Potter novels,

most recent US presidents, planets, Pixar films, Star Wars films, countries with large landmasses; Kelley & Parihar, 2018). Participants exhibited a diminished recall ability for some categories but a facilitatory effect for other categories when presented with part-list cues. Some evidence suggests that if cues are presented for prompting semantic recall, the part-list cuing impairment occurs only if the cues are from the same category as the items to be recalled and not if the cues are from a different semantic category (Watkins & Allender, 1987). Together, the findings on whether semantic memory retrieval is sensitive to disruption from part-list cues show that the effect of part-list cues on semantic memory is mixed and may vary based upon the content of what is being tested and how it is being tested.

When it comes to implicit memory tasks, the key factor that distinguishes these tasks from tests of explicit memory is the participant's lack of awareness that researchers are testing their memory for recently shown information. In explicit memory, participants are aware that the goal of the retrieval task is to try to remember as much of the information they saw earlier as possible. In implicit memory tasks, participants complete the task with the first responses that come to mind, no reference is made to the study phase, and various features are included in the procedure to reduce the likelihood that participants engage in the use of explicit memory when performing the task (Roediger & Geraci, 2003; Roediger & McDermott, 1993). Here, too, the patterns of findings are mixed. In these studies, part-list cues consisted of some of the studied words, presented either intact (Basden et al., 1991) or with a single letter missing (Peynircioğlu, 1989; Peynircioğlu & Moro, 1995), and participants were asked to complete the fragmented versions of the remaining studied words as well some nonstudied words (e.g., *_ p _ l _*; *apple*) with the first solution that comes to mind. Some studies found that part-list cues lower performance on the implicit, word fragment completion test (Peynircioğlu, 1989; Peynircioğlu & Moro, 1995, Experiment 1 and 2), whereas other studies have reported null or facilitatory effect of part-list cues (Basden et al., 1991; Peynircioğlu & Moro, 1995, Experiment 3).

In brief, the literature on the part-list cuing effect on memory tasks other than recall is modest. The few experiments that used recognition memory tasks have shown a part-list cuing impairment in recognition memory, although more research is needed to reinforce these findings. Findings for both semantic memory and implicit memory tasks have shown varied effects of part-list cues, where this effect has occurred in some studies but not in others. Furthermore, the reasons for these mixed patterns have not been obvious. As we turn to the theoretical accounts in the next section, it is pertinent to note that the leading accounts have focused on the recall task to account for the part-list cuing impairment in memory. As the vast majority of the empirical literature

² Some studies have reported part-list cuing impairment in performance on non-memory tasks such as identifying fragmented images coming into focus in the presence of some example solutions (Peynircioğlu, 1987) and solving problems with the help of some example solutions (Del Missier & Terpini, 2009). As this literature is sparse and these tasks do not readily speak to memory processes, we do not consider them further in this review.

has investigated the part-list cuing impairment in episodic recall tasks, the focus of our review will be on this recall task, both in our meta-analytic treatment and in the theoretical description.

Theoretical Accounts

As researchers have tested the range and the boundaries of the part-list cuing impairment on recall, they have also brought together this evidence to identify the mechanisms that drive this effect. Over time, some theories have gained traction over others in being able to integrate the range of available findings. A detailed treatment of these theoretical accounts can be found in Nickerson (1984), including the early hypotheses such as the editing task hypothesis (Roediger & Tulving, 1974), the increased-list-length hypothesis (Watkins, 1975), and the cue-overload hypothesis (Mueller & Watkins, 1977). We summarize here the three major and currently prevalent accounts of the part-list cuing impairment that are also directly pertinent to the current literature and this review.

Retrieval-Strategy Disruption Hypothesis

The *retrieval-strategy disruption* hypothesis is one of the leading explanations for the part-list cuing impairment and it accounts for many of the past findings for the recall task (Basden et al., 1977). This hypothesis proposes that the impairment is rooted in the disruption to the idiosyncratic manner in which people organize studied items to guide their recall. According to this hypothesis, when participants are given the part-list cues before recall the cues disrupt their ability to fully utilize their individual, intended strategy. In contrast, those who are not faced with part-list cues are free to use their preferred organizational strategy for recalling the studied items.

This hypothesis has been a focus of many experimental methods and interpretations since its inception, leading to the development of two common methodologies that provide evidence in support of this explanation: Congruency of cue order with study order and inclusion of a second free-recall task following the part-list cued recall phase. In this hypothesis, even though participants are said to form an idiosyncratic organization of the studied information it is implied that this organization will share some degree of congruency to the order in which the items were presented during study. A match between the order of items in the study list order and in recall sequence has been a subject of considerable scrutiny in the free-recall task, demonstrating a temporal contiguity effect where study order guides the recall order to a certain extent (Deese & Kaufman, 1957; Kahana, 1996). In the part-list cuing paradigm, this assumption suggests that

if researchers present the cues in an arrangement that parallels the order in which the studied items were presented, the amount of disruption to the participant's retrieval strategy should be reduced as study list order is one of the possible strategies people use to recall studied information (Basden et al., 2002). Some examples of such a presentation include presenting the part-list cue items in the same position as they appeared in the study list, with the to-be-recalled items denoted with blank slots (Basden & Basden, 1995; Experiment 1) or presenting even numbered items from study list as part-list cues (Reysen & Nairne, 2002, Experiment 2). Several studies have directly tested this assumption and found support for the retrieval-strategy disruption hypothesis since a smaller impairment is observed when the study list order and the cue list order are congruent compared to when they are incongruent (Basden & Basden, 1995; Garcia-Marques et al., 2012; Reysen & Nairne, 2002; Serra & Nairne, 2000), though some studies report the opposite with a larger impairment being observed when study list order and cue list order are congruent (Fritz & Morris, 2015; see also Wallner & Bäuml, 2020).

Additional evidence for the strategy disruption hypothesis comes from studies where researchers ask participants who previously performed the recall task with part-list cues to perform another recall task, this time without any cues, that is, using a free recall procedure. If the strategy prepared for the recall gets disrupted by the part-list cues, then when researchers remove the cues the rememberer should be able to use their originally prepared strategy. As such, one would expect that the performance of those who were previously exposed to part-list cues will no longer be hampered, and thus their ability to remember the target items will rebound. Some studies using this procedure have reported findings that show such rebounding and therefore provide support for the retrieval strategy hypothesis (D.R. Basden & Basden, 1995; B.H. Basden et al., 1991; Bäuml & Aslan, 2006; Muntean & Kimball, 2012; Roediger III et al., 1977). However, there are reports where this rebounding does not occur or it depends on other factors for its occurrence (Barber & Rajaram, 2011; Bäuml & Aslan, 2006; Del Missier & Terpini, 2009; Muntean & Kimball, 2012; Wallner & Bäuml, 2020).

Aside from instances where recall performance does not rebound on a later free-recall task for those participants who first performed the part-list cued recall, there are other challenges to the retrieval-strategy disruption hypothesis as the sole explanation for the part-list cuing impairment. For instance, it is possible to disrupt the retrieval strategy of the participants who perform free recall in the first recall phase (the control condition), and when this happens, the part-list cuing impairment is still observed. An example of such a comparison is when researchers provide participants in both the part-list cued recall and free recall (control) conditions with the unique first letter (or first two letters) of each to-be-recalled item that serve

as item-specific probes during the test phase (e.g., *sw*_____ for recalling *sweater*). In this setup, part-list cued participants receive a standard cue word display as well as item-specific probes for the to-be-recalled items and the control participants receive only item-specific probes for the target items. This procedure can create interference to the retrieval strategies in both conditions because the experimenter sets up the sequence of item-specific probes for recall in both conditions; therefore, the experimenter determines the order of recall even in the control condition rather than the participant (i.e., both experimental and control conditions receive initial-first letter cue of all non-cued target items). While some attenuation of the impairment could occur in both conditions because participants in neither condition get the opportunity to use to their preferred retrieval strategy, the question is whether the part-list cues create additional interference in the part-list cuing condition compared to the control condition and reduce recall of the target items.

In studies where this approach of providing item-specific cues to all participants was taken, researchers still consistently observed a part-list cuing impairment (Aslan et al., 2007; Aslan & Bäuml, 2009; Aslan & John, 2019; Bäuml & Aslan, 2004; Bäuml & Aslan, 2006; Bäuml & Kuhbandner, 2003; Bäuml & Samenieh, 2012; Bäuml & Schlichting, 2014; Crescentini et al., 2010; Crescentini et al., 2011; Kissler & Bäuml, 2005; Muntean & Kimball, 2012). With so many reports of the impairment being observed when using item-specific probes, researchers must ask whether disruption to the retrieval-strategy can fully explain the part-list cuing impairment.

Retrieval Inhibition Hypothesis

As noted, the retrieval-strategy disruption hypothesis cannot account for the full range of evidence for the part-list cuing impairment in recall, suggesting additional mechanisms at play. In fact, there is evidence to suggest that multiple mechanisms underlie the part-list cuing impairment (e.g., Lehmer & Bäuml, 2018a).

One alternate mechanism is known as retrieval blocking that can occur due to *competition-at-retrieval*. Here, presenting some studied items as cues during recall strengthens the memory representations of those items through covert retrieval when reviewing the cues. This strengthening increases the accessibility of those items relative to the target items. One of the potential outcomes of this strengthening is a blocking of target items during recall (Roediger III, 1973; Rundus, 1973). This outcome can be observed when instructions to recall as many of the items as possible (including the cues) lead the participants to give preference to the cued words at the beginning of recall (Roediger III et al., 1977), providing evidence that the cued items are more accessible than the target items.

Another form of competition-at-retrieval that could account for the part-list cuing impairment is *retrieval inhibition*. Whereas *retrieval blocking* results in target items being less accessible than the cued items, *retrieval inhibition* suggests that the outcome of covert retrieval is the suppression of the memory representations of the target items, making them functionally unavailable (Anderson et al., 1994; Bäuml & Aslan, 2004). Support for this mechanism can be observed in experimental research where participants are provided with either standard part-list (covert retrieval) cues to aid in their retrieval or with word stems of part-list cues (overt retrieval) to complete prior to retrieval. In these instances, both the standard part-list cues and the word-stems urge the participant to perform retrieval of the provided items, resulting in retrieval inhibition of the non-cued items which is indexed by reduced recall of these items. In another condition in this context, part-list cues are provided for the participants to use as a second re-exposure/study opportunity (i.e., “*relearning*”). In this re-exposure condition, because participants are urged to reprocess the presented material, this process does not require covert retrieval, and thus does not produce the negative impact of being presented with part-list cues on recall (Bäuml & Aslan, 2004). Support for the operation of retrieval inhibition mechanism also comes from work that focused on recall in the second recall test that following the first recall, to measure the extent to which studied items remain inaccessible following part-list cuing during the first recall (Aslan et al., 2007). If a studied item (e.g., “*ROBE*”) suffers inhibition due to part-list cuing, access to it would fail on a second recall test regardless of whether the item is cued with a probe with which it was studied (“*COTTON*”) or with an independent-probe consisting of nonstudied cues (e.g., “*CLOTHING*”) that are associated with the target items (e.g., “*ROBE*”) (see Anderson et al., 2000; Anderson & Spellman, 1995). When participants were provided with such independent-probes on the second recall test, they still exhibit part-list cuing impairment, supporting an inhibitory basis for the part-list recall impairment (Aslan et al., 2007).

Additionally, if the part-list cuing impairment is the result of retrieval blocking, then when researchers provide a participant with the unique first letters of the target items in a recall task, the impairment should no longer be present as accessibility to the target items increases (Bäuml, 2008). As noted earlier, there is a large body of literature that utilizes item-specific probes and still reports a part-list cuing impairment, which suggests the reduction in the accessibility of target items is related to inhibition of their memory representations rather than blocking (Aslan et al., 2007; Aslan & Bäuml, 2009; Aslan & John, 2019; Bäuml & Aslan, 2004; Bäuml & Aslan, 2006; Bäuml & Kuhbandner, 2003; Bäuml & Samenieh, 2012; Bäuml & Schlichting, 2014; Crescentini et al., 2010; Crescentini et al., 2011; Kissler & Bäuml, 2005; Muntean & Kimball, 2012).

The retrieval inhibition hypothesis also predicts that items that have weaker inter-item associations (e.g., *lion-zebra*) should be adversely affected by part-list cues more than those that have stronger inter-item associations (e.g., *lion-tiger*). This outcome is predicted by a form of retrieval inhibition often referred to as *feature suppression* (Anderson & Spellman, 1995). This form of retrieval inhibition suggests that when patterns (or features) of a cue are shared with a target item, the suppression of those patterns will in turn inhibit the target item less than weakly related items. Direct support for this prediction comes from a study by Aslan and Bäuml (2009), where part-list cued recall was compared for studied items with low inter-item (i.e., *lion-zebra*) and high inter-item (i.e., *lion-tiger*) associations, and the impairment was found to be greater in the recall of items with low inter-item associations.

Additional support for retrieval inhibition, in general, comes from studies that have used the DRM paradigm (Deese, 1959; Roediger III & McDermott, 1995). As noted earlier, in the DRM paradigm the false recall of the non-presented critical lure item, for example, *sleep*, increases when participants study a list of associatively related items such as *pillow*, *bed*, and *night*. Part-list cuing studies show that when the associatively related items are presented as cues the recall of critical lures is greatly diminished (Bäuml & Kuhbandner, 2003; Kimball & Bjork, 2002). These patterns are consistent with retrieval inhibition as critical lures are not studied and, therefore, should not be disrupted by the part-list cues (as the retrieval disruption hypothesis would suggest).

The retrieval inhibition explanation can also account for instances where a rebounding effect does not occur on the second recall task such that participants who first perform part-list cued recall do not show an increase in recall on a second, free-recall task (e.g., Barber & Rajaram, 2011; Bäuml & Aslan, 2006; Del Missier & Terpini, 2009; Muntean & Kimball, 2012; Wallner & Bäuml, 2020). Conversely, the retrieval inhibition hypothesis cannot explain all the data as many studies do report a rebounding effect in performance on a second, free-recall task (D.R. Basden & Basden, 1995; B.H. Basden et al., 1991; Bäuml & Aslan, 2006; Muntean & Kimball, 2012; Roediger III et al., 1977).

Multi-Mechanism Hypothesis

Most recent evidence suggests that part-list cues influence recall in multiple ways such that they impair recall in some situations, can even facilitate recall in other situations, or have little effect (Lehmer & Bäuml, 2018b). To account for this range of effects, this multi-mechanism hypothesis draws attention to the roles of the extent of study-test delay and the extent of overlap that exists between the study and test contexts. For example, part-list cues impair recall when there is a short delay between study and test, conditions that

may result in a higher degree of study-test context overlap (Lehmer & Bäuml, 2018a, Experiment 2). When the delay between study-test phases increases (e.g., 1 week), resulting in a greater degree of incongruity between the study-test context, the part-list cues may no longer impair recall. This outcome can consist of an absence of recall impairment or even facilitation in recall depending on the encoding conditions (Lehmer & Bäuml, 2018a, Experiment 2), with possibly a neutral effect of part-list cues for study-test delays in the intermediate range.

The multi-mechanism hypothesis proposes a third, context reactivation mechanism, in addition to retrieval disruption and retrieval inhibition, to account for when part-list cues produce impairment, facilitation, or little effect (Bäuml & Samenieh, 2012; Lehmer & Bäuml, 2018b). Context reactivation refers to a process where the part-list cues reactivate the study context and improve access to targets to be recalled. The study and test conditions in any situation determine which of these three mechanisms, and in what combination, influence the recall outcome. For example, with respect to retrieval disruption, when the study-test contexts match (e.g., short study-test delay), and the encoding conditions involve study of highly associated items that facilitate the construction of a retrieval plan, the part-list cues presented at test disrupt this plan and produce recall impairment (e.g., Basden et al., 1977; Basden & Basden, 1995). With respect to context reactivation, this mechanism pertains to part-list cues activating the study context, and conditions that allow such activation, can improve recall. Conditions where both context reactivation and retrieval disruption operate, the former facilitates the activation of study items whereas the latter disrupts this process, producing a net result of no impairment. Consistent with this argument, when encoding conditions create strong associations and the study-test context mismatch (e.g., a 24-h study-test delay), part-list cues may help reactivate the study context and access to the retrieval plan, but retrieval disruption from the cues may counter this process, thus resulting in no part-list cuing impairment (Lehmer & Bäuml, 2018a).

In contrast to the encoding conditions that entail processing of high associations among study items, when the encoding conditions entail low associations among study items (e.g., a single study session, uncategorized words; Bäuml & Aslan, 2006), these conditions favor the operation of retrieval inhibition at test when part-set cues are provided for recall. Further, under conditions that include low-associative encoding and a mismatch between study and test contexts (e.g., 24-h delay), context reactivation can aid access to the study context, and given the absence of a strong retrieval plan, retrieval disruption does not counter it, and thus a facilitation in recall may be observed (Lehmer & Bäuml, 2018a). Thus, the multi-mechanism account aims to explain the reports in favor of both retrieval disruption and retrieval inhibition

observed in the literature, and offers a third mechanism, context reactivation, to reconcile the range of findings arising in response to the use of part-list cues in recall.

Current Goals

As can be surmised from our review thus far, the range of effects have been broad, and multiple theoretical explanations have been proposed for the part-list cuing impairment in recall. As with any body of literature this large, the available evidence reported over the last 50 years lends itself well to a meta-analytic synthesis. The use of systematic review techniques helps organize the process to integrate the literature review and provide quantified assessments of the effects. Since no quantitative synthesis has been conducted for the large empirical literature on the part-list cuing impairment in recall, the contributions of our review will be novel and informative. Results from our meta-analysis will assist researchers, who are looking to undertake tests on the impairment in novel domains in a way that has not previously been explored, with selecting variables and conditions that influence the size of the impairment. Our results will also assist with empirical replication efforts where findings may turn out to be theoretically or methodologically intriguing. Finally, the meta-analytic results also have the potential to inform other domains of research and applications including education and law, as well as daily life reminder lists, where the use of part-list cues is common. These overarching goals guided the objectives of the current report.

Main Goals

Rather than testing specific hypotheses, the primary goal of our meta-analysis was to assist in the planning and execution of future experiments related to the part-list cuing impairment. The results reported within this manuscript explore several aspects of the general procedure both in their prototypical execution and in their frequent deviations across experiments. When experiments have slight deviations from one another that are not the focal point of our investigation, these inconsistencies can account for some of the differences in observed effect sizes across experiments. With an emphasis on replicability and generalizability across populations in psychological research (Open Science Collaboration, 2015), our objective of assessing the robustness of the part-list cuing impairment in recall is particularly timely. To this end, we investigated eleven factors we identified in this literature as being prominently useful to examine: (1) how long each item was presented; (2) the relatedness of study items; (3) inter-item association; (4) the length of the study list; (5) the modality in which stimuli were presented; (6) the length of the distractor task; (7) the number of cue words

provided; (8) the length of the retrieval task; (9) whether or not item-specific probes were provided during retrieval; (10) the year of publication, and (11) publication status. In doing so, we examined each of these moderators individually, rather than in conjunction with one other as is sometimes done in empirical experiments, to provide estimates of their individual influence on the memory impairment phenomenon of interest and provide valuable guidance for future experiments.

A second goal of the current report is to pinpoint the weighted average effect size range of the part-list cuing impairment. This information will provide a baseline measurement for comparing the effect sizes observed in future experiments. From this information, future researchers can evaluate if the effect observed aligns with the average size of the impairment in the past literature or if it substantially deviates from the expected range. Such comparisons have the potential to provide insights into the contexts that produce or mitigate this memory impairment, and what we can infer about the theoretical reasons guiding such variations.

Auxiliary Goal

Beyond the two main goals of our report that we outlined above, our third, auxiliary goal consists of qualitatively comparing the findings from this meta-analysis to the two major theoretical accounts we outlined earlier. We acknowledge caution when directly applying our findings as a test of a cognitive theory and the view that cognitive theories are best tested through direct experimental manipulations specifically designed to test them. It is also important to keep in mind here that each moderator analysis provides insight on that moderator alone. However, our analyses will have implications for the theories discussed, and as such, it is important we consider these implications.

Method

Literature Search

The literature search for our meta-analysis was conducted through an exhaustive examination of four academic library databases: PsycINFO, PsycARTICLES, PubMed, and ProQuest. We entered the following keywords into each database in separate searches: *part-set cue*, *part set cue*, *part set cueing*, *part-set cueing*, *part set cuing*, *part-set cuing*, *part-list cue*, *part list cue*, *part-list cueing*, *part list cueing*, *part-list cuing*, and *part list cuing*. Additionally, on PsycINFO, PsycARTICLES, and PubMed we activated the filter for “peer-reviewed” to ensure only published scholarly reports that had been scrutinized by the academic community populated the search field. During the identification stage, we

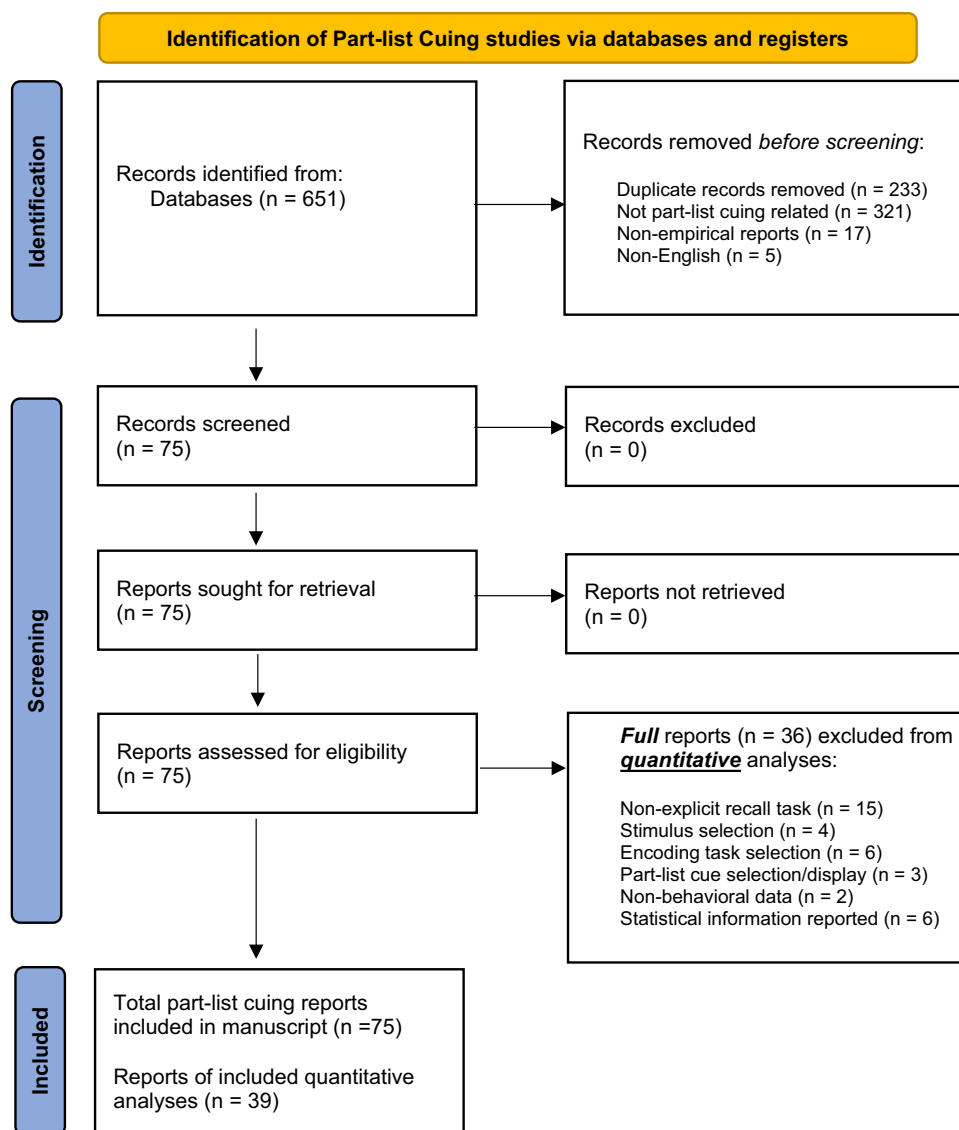


Fig. 2 A visual summary of the literature search conducted for this meta-analysis

reviewed each entry's abstract for relevance to the part-list cuing paradigm before a full assessment of its content and the extraction of pertinent procedural details. Reports that did not directly investigate the part-list cuing impairment through behavioral data analysis (such as computer simulations), literature reviews, theoretical syntheses, and those not published in the English language were excluded from the analyses. This literature search was conducted multiple times over the course of development of the current report and was conducted by the first author (a doctoral student). A detailed summary of the outcome of these searches is outlined in Fig. 2. In addition, screening of articles and study selection for inclusion was conducted by the first author in consultation with the last, faculty author. During the screening stage,

all screening was conducted by a full-text review of each individual record. No protocol was registered.

Experiment Selection and Categorization

Once reviewed, each non-overlapping sample in a given report was separated for effect size calculations. If a study reported multiple experiments or conditions where each experiment had a completely independent sample, we included each sample's effect size as a separate data point in the analysis.

All samples that met the selection criteria were then categorized as either between-subjects or within-subject measurements. Most samples excluded from analyses were

excluded for reasons related to the experimental procedures described in the next section or a lack the proper statistical information required to calculate an accurate Hedges' g effect size value.

Furthermore, in consideration that most reports included in the analyses do not report detailed demographic information, our analyses are agnostic to individual differences that may arise due to participants' demographic background. Of the reports included in the analyses, 28.21% provided the mean age of participants, 20.51% provided gender ratios of their sample, and 0% provided specific information relating to ethnicity. To note, 92.31% of reports had included undergraduate students in their samples. This lack of demographic information in the literature aligns with recent reports that fewer than 1% of publications in cognitive psychology highlight the race of participants (Roberts et al., 2020).

Experiments with Multiple Effect Sizes

Although some experiments are straightforward tests of the impairment, with one experimental condition and one control condition, others include multiple conditions compared to a single control condition. In the standard practice of conducting meta-analyses, we ensured all effect size measurements were independent. For this reason, whenever we were presented with a forced choice between effect sizes, we selected samples that most closely resembled the prototypical experimental paradigm, as previously outlined. In situations where more than one measurement met these criteria, we used the moderating factor closest to the majority of the other samples as a tiebreaker. For example, in a study where the proportion of cue items being presented was the only differing factor between two effect sizes, the one closest to 50% of the total items presented was selected.

General Exclusion Criteria

We used a general set of exclusion criteria to refine the samples that were included in our analyses.

The most common reason we excluded studies was because the necessary statistical information required to determine an accurate effect size measurement for a given condition was not available. When statistical information was lacking, we contacted authors via email for information prior to excluding the experiment. Eventually, 16 experiments were excluded for a variety of reasons (e.g., access to only the collapsed results for conditions that deviated in their procedures, only the collapsed statistics for an omnibus analysis when a significant interaction was reported, a lack of information required to calculate a within-subject effect size).

We also excluded experiments for using non-recall memory tasks or non-memory tasks as noted in the Introduction. These included recognition memory tasks (Marx, 1988; Todres &

Watkins, 1981), visual reconstruction tasks (e.g., Cole et al., 2013; Drinkwater et al., 2006; Fritz & Morris, 2015; Huffman et al., 2001; Kelley et al., 2016; Watkins et al., 1984), exemplar or option generation tasks (Del Missier & Terpini, 2009; Peynircioğlu, 1987, Experiments 1 and 4; Peynircioğlu & Gökşen-Erelç, 1988; Watkins & Allender, 1987) as well as spot-the-difference tasks (Peynircioğlu, 1987, Experiments 2 and 3). One experiment was excluded for having a sample that was not independent of another already included in the analyses (Sloman et al., 1991, Experiment 4).

Procedure-Specific Exclusion Criteria

In addition to the general exclusion criteria, the following procedure-specific criteria led to exclusion of additional studies from the main and auxiliary goals of this meta-analysis. As we noted in the *Introduction*, these exclusion criteria reflect substantial deviations from the prototypical procedures: (1) using particular types of to-be-remembered stimuli; (2) including a directed forgetting procedure; (3) using encoding tasks that involved additional tasks and varied across this subset of studies; (4) using extra-list words as part-list cues; (5) using non-episodic recall tasks; and (6) having additional procedural features during the recall task. The number of studies excluded based upon these procedure-specific criteria were not sufficient for each criterion to motivate a moderator analysis for that factor. We describe each of these criteria in more detail below and note that we still incorporated these studies in the qualitative review of the literature to capture these findings within the overall view of the part-list cuing literature. For our main and auxiliary analyses, the general and procedure-specific exclusion criteria together resulted in 47 samples being included in the between-subjects analyses and 49 samples being included in the within-subject analyses.

Turning to each of the specific exclusion criteria we listed above, the first and most straightforward exclusion criterion with respect to the procedure we adopted for our analyses was the type of to-be-remembered stimuli provided during study. To keep the metric of effect sizes obtained from each sample consistent, only measurements where single words were provided for recall were included. The number of experiments where this was not the case is small, with a total of 11 experiments using a variety of different stimuli, and thus a moderator analysis regarding this procedural manipulation could fail to capture differences related to this deviation. Experiments that met this exclusion criterion were ones that utilized full sentences (Garcia-Marques et al., 2012; Garrido et al., 2012), word pairs (Basden et al., 1991, Experiment 2; Mueller & Watkins, 1977, Experiment 4), literary prose (Bäumel & Schlichting, 2014; Fritz & Morris, 2015), and images (Bovee et al., 2009).

We excluded an additional five samples for being from conditions where directed forgetting was an experimental manipulation. In the directed-forgetting procedure, participants are provided with two consecutive sets of items to remember. After encoding the first set, they are instructed to forget that set and only remember the second set that they are presented with. At retrieval, the participants are then given surprise instructions to recall the set of items they were specifically told to forget. This instructional procedure substantively deviates from the typical instructions to remember the items for later recall. Part-list cuing studies that have used this manipulation have done so to predict an elimination or reversal of the part-list cuing effect and have consistently reported a facilitatory effect of part-list cuing rather than the normal impairment (Bäuml & Samenieh, 2012; Goernert & Larson, 1994; Lehmer & Bäuml, 2018a). For this reason, it can be inferred that this procedure likely taps into either different or additional underlying mechanisms than the typical part-list cuing procedure. To be noted, the comparison conditions utilized in these types of experiments (where participants are not instructed to forget the first set) were included in our analyses.

Encoding procedures were also scrutinized for consistency with the general paradigm. One study (Riefer et al., 2002) was excluded due to being inconsistent with the number of items being presented at one time during encoding. While included in our analyses were samples that provided multiple items at a given time (such as category blocks used in Basden & Basden, 1995), the items presented in this particular experiment were an intermix of single items as well as word-pairs.

Additionally, we excluded eight experiments because in the encoding phase participants were required to perform additional tasks that substantially extended beyond the typical procedures used for viewing the stimuli and these additional tasks also varied across the studies. While we included studies that involved processing of meaning during encoding, that is, a deep level of processing (such as in Bäuml & Aslan, 2006, and Lehmer & Bäuml, 2018a), the samples we excluded for encoding related procedures required participants to process study information in a far more involved manner than the standard procedure. For example, in these studies, participants sorted the stimulus items at their own pace into an idiosyncratic order/categories (Mueller & Watkins, 1977; Penney, 1988), indirectly generated the to-be-remembered items in response to antonyms (Muntean & Kimball, 2012), or generated the part-list cues themselves and performed a semantic retrieval task 48 h later using those cues (Brown & Hall, 1979).

Another factor we considered pertains to the part-list cues used in experiments, in particular, the use of extra-list cues (i.e., words serving as retrieval cues that were not a part of the study list). As we described in the Introduction, only

a few experiments have used extra-list cues. Furthermore, the outcome in these studies depends on the selection of to-be-studied items (unrelated nouns or categorized word lists), and its relationship to the selected extralist cues (items from studied versus non-studied categories). These variations within a small set of studies make a direct moderator analysis on this variable potentially uninformative. With that in mind, we excluded two samples from the analyses specifically for the use of extra-list cues. A similar logic was adopted for auditorily delivered part-list cues. Due to only four experiments (three between-subjects and one within-subject) presenting cues auditorily (Andersson et al., 2006; Zellner & Bäuml, 2005), we only included studies that visually presented part-list cues.

When it comes to the types of recall memory tasks, we excluded experiments that involved semantic and implicit memory from our analyses (four and six experiments, respectively). Semantic memory experiments lack an encoding phase as the researchers are probing participants' knowledge, such as states in the U.S., that the researchers do not provide to participants (Brown & Hall, 1979; Foos & Clark, 2000; Kelley & Parihar, 2018; Rhodes & Castel, 2008). Although these experiments can be informative about the generalizability of the impairment, the design of these investigations deviates considerably from the standard part-list cuing paradigm and, in turn, from the main goal of our analyses.

Similar reasons guided our exclusion of implicit memory experiments. Implicit memory experiments do include a study phase, but their procedure requires the participants to perform memory tasks such as word fragment completion that do not elicit explicit memory (Peynircioğlu, 1989; Peynircioğlu & Moro, 1995). Other procedural features, like deception, are often included to disguise the fact that the task taps into memory for recently viewed information (e.g., Peynircioğlu & Moro, 1995, Experiment 3). Although these implicit retrieval conditions investigate the boundaries of the part-list cuing impairment, they also depart from the typical recall procedure in which our analyses are grounded.

Finally, we excluded eight experiments for deviating, in several ways, from the prototypical procedures used during retrieval in tests of part-list cuing impairment in recall. This set included conditions where participants were asked to produce the cue words using word stems (Peynircioğlu, 1989) or word fragments (Peynircioğlu & Moro, 1995), where participants completed multiple free recall attempts before performing part-list cued recall while receiving feedback on their performance on the free recall retrievals (Mueller & Watkins, 1977), where part-list cues were presented gradually throughout the retrieval phase rather than all being presented at the start of the recall phase (Garrido et al., 2012), and where the distractor task came after the presentation of the part-list cues (Bäuml & Aslan, 2004).

These criteria were set to ensure that the effect sizes used in the analyses captured a homogenous retrieval task with the key manipulation consisting of simply providing participants with a set of cues prior to recall. For these substantial deviations from the prototypical procedures, a total of 41 samples across 29 experiments were excluded due to our procedural criteria with 18 samples deriving from between-subject designs and 23 deriving from within-subject designs. As noted, we include this set of studies in our discussion later for full consideration.

Calculation of Effect Sizes

We calculated effect sizes in a variety of ways to enable us to include as many effect size measurements as possible. All measures were then converted by Comprehensive Meta-Analysis (CMA) 2 to be in Hedges' g (Borenstein et al., 2005). To assure accuracy in these calculations, all effect size measurements were double coded using the method previously described.

Direct reports of Cohen's d were the preferred methods of obtaining the Hedges' g value. If the effect size was not directly reported, the t -score was weighted with the size of the sample in each condition to produce an effect size. If the results were only reported as the main effect F -value, this value was converted to a t -score by taking the square-root of the F -value. Lastly, if the above information was not reported but measures of central tendency were and the design was between-subjects, we used those values to derive Hedges' g . For within-subject designs, this approach does not meet the criteria required to calculate the effect size as it ignores the pair-wise differences within a given participant.

Importantly, for reports where a null effect was observed the above steps were still followed whenever possible. In many instances, researchers do not provide the full breadth of statistical information for the null effects they observed. Rather than excluding these samples and adding bias to our analyses, we still included these reports. If the report only stated " $p < .05$ ", " $t < 1$ " or " $F < 1$," as is common practice, we entered the effect size as zero as the most conservative representation of the size of the effect available.

Analysis Classification

Main Analyses The main analyses in our meta-analysis consisted of measurements from samples tested using the standard part-list cuing recall paradigm. As summarized in our Introduction, there are several distinct features that characterize the standard part-list cuing paradigm. Many conditions across experiments vary these features to investigate the boundaries of the phenomenon.

We defined the prototypical paradigm based upon the most common methodology that researchers have utilized

when examining the impairment (see Fig. 1) and that has served as the standard design that researchers have manipulated to test potential moderators. To summarize, the samples matching these criteria were from conditions that all presented single words during the encoding period rather than more complex stimuli. If the distractor task was employed between encoding and retrieval, it had to be presented before the part-list cues were presented to the participants. Additionally, cues had to be presented visually at the beginning of the retrieval task and come from the initial study list. Finally, participants' interaction with the cues could not exceed a standard attention check (such as reading the cue aloud or placing checkmarks next to each word), and the retrieval task had to be an explicit and episodic recall task.

After these exclusions, our final analyses consisted of 47 samples ($N = 2,574$) for the between-subjects analyses and 49 samples ($N = 2,102$) for the within-subject analyses.

Conservative Analyses In addition to our main set of analyses, we conducted a narrower set of analyses where we applied additional selection criteria. The goal of these analyses was two-fold: (1) to assess how robust the impairment is when the criteria are tightened (i.e., to assess if the effect size deviates compared to the main analysis) and (2) to assess if the results of our planned moderator analyses in our main set of analyses would change when methods that diverged from the prototypical paradigm to some extent (though not as much as the exclusion criteria described in the previous sections) were removed. Here, effect sizes from conditions where a specific organization of the items was imposed during encoding (such as blocked presentation of items like in Basden & Basden, 1995, and Sloman, 1991, or cue words being congruent with the encoding order in Sloman et al., 1991, and Serra & Oswald, 2006) rather than randomization of both presentation and cue order were not included in this set of analyses. We made this decision because these types of methodological procedures have the potential to change the organizational approach and retrieval-strategy that participants adopt and are often the reason why researchers use them, thus on purpose diverging from the standard procedure.

An additional restriction concerned the length of the retention period between the encoding and retrieval phases. If the retention period exceeded 5 min, we did not include the study in this more stringent analysis. Research has shown that increasing the retention period between the two phases can extinguish the impairment (Lehmer & Bäuml, 2018a) and, at longer intervals (e.g., 30 min and beyond), can sometimes produce a facilitatory effect in recall (Bäuml & Schlichting, 2014; Lehmer & Bäuml, 2018a). This outcome is predicted based on theory where the part-list cues reinstate the initial study context, and, in turn, also serve as cues that reactivate the earlier, study context, a factor that does not

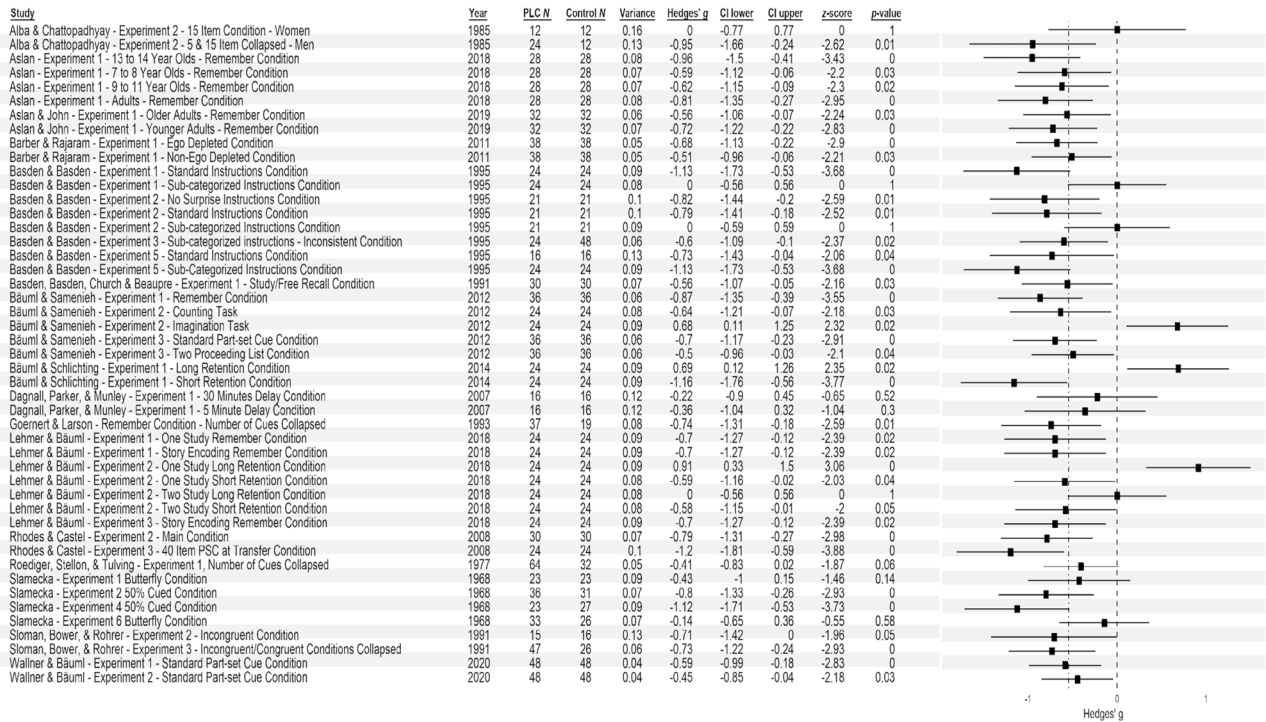


Fig. 3 Forest plot of effect sizes for between-subjects designs included in Part-list Cuing Impairment in Recall meta-analysis. Each effect size included in the meta-analysis has its own row which depicts the study information and statistics. On the right side of the

diagram, the Hedges' g value is represented as a rectangle with the confidence intervals represented as the error bars and the vertical lines representing the mean effect size (the left line) and zero (the right line)

occur during the shorter retention periods of the standard part-list cuing design (Lehmer & Bäuml, 2018a).

After these exclusions, our final analyses consisted of 30 samples ($N = 1,727$) for the between-subjects analyses and 48 samples ($N = 2,074$) for the within-subject analyses.

Lenient Analyses We conducted a final set of sensitivity analyses to focus on the robustness of our findings by using a broader range of samples. The goal of these analyses was to incorporate unpublished experiments in doctoral dissertations that had gone through a committee review but never underwent the full peer-review process. This set of analyses added only within-subject designs, as we did not locate any samples that met our general inclusion criteria that were between-subject designs. After these inclusions, our final analyses consisted of 62 samples ($N = 3,031$) for the within-subject lenient analyses.

For the sake of efficiency, we will report results from the main analyses and will report the conservative and lenient analyses only in reference to the weighted overall effect size or in cases where the findings differ from the main analyses. A full list of the samples we included can be found in Figs. 3 and 4 for between-subjects and within-subject designs, respectively. Articles in which at least one condition was included in the meta-analysis are denoted with an asterisk in our reference list.

Coding Moderators and Effect Sizes

All procedural moderators were coded by the first author for all studies. All effect sizes were calculated by the first and third authors independently, both of whom were doctoral students at the time. We then compared these values and reconciled. In any instance where the two authors differed, we reviewed the paper in question together, identified the source of the difference, and reached an agreement on the correct effect size. This process was taken for every effect size included in the analyses until the authors reached 100% agreement. The full dataset of coded moderators and effect sizes will be made available upon request, please direct inquiries to the corresponding author.

Study Presentation Time In episodic memory tasks, each item on the study list is typically presented for a specific amount of time per item during the study phase. While this duration often does not exceed the range of 2–5 s per item, we wanted to provide insight into whether presentation time had any predictive value in terms of the size of the impairment. This factor is informative for studies where the experimental design usually calls for longer presentation times, such as in fMRI studies or studies with children, older adults, and memory-impaired individuals. This

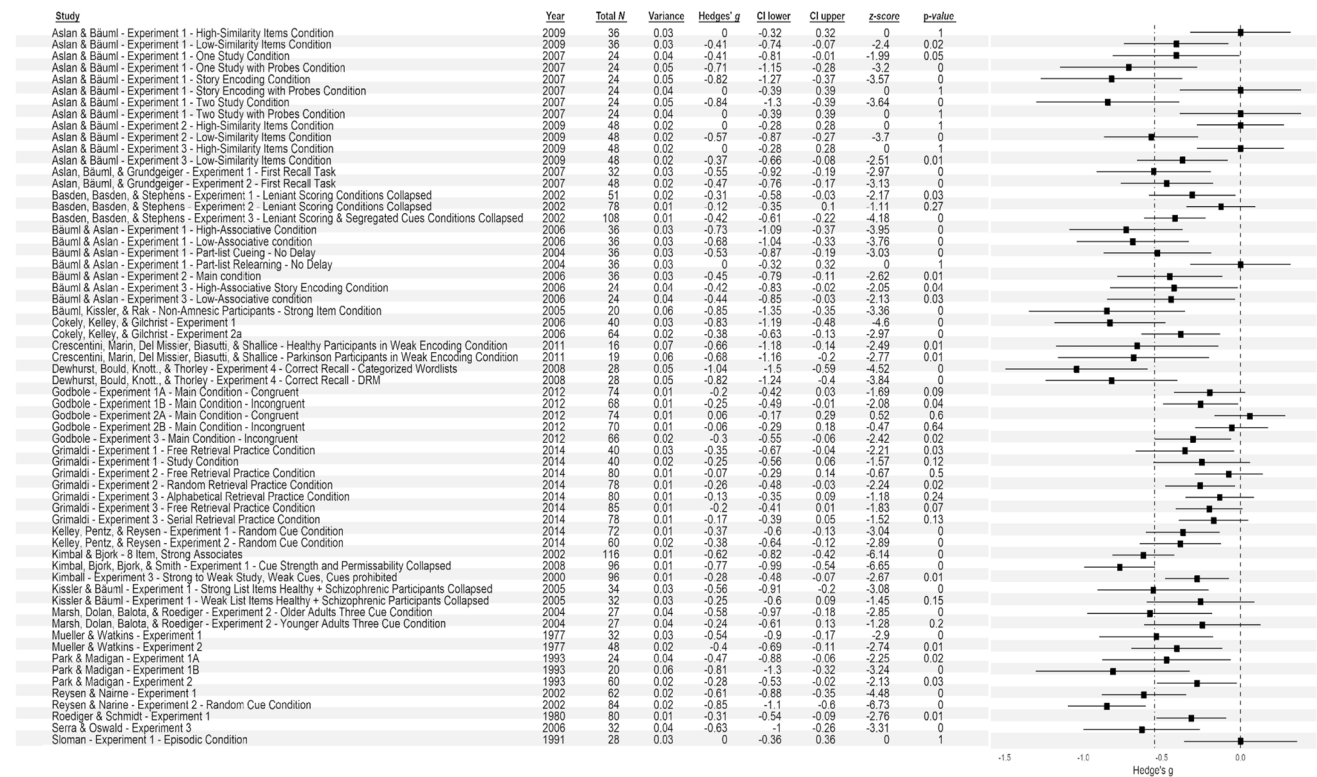


Fig. 4 Forest plot of effect sizes for within-subject designs included in Part-list Cuing Impairment in Recall meta-analysis. Each effect size included in the meta-analysis has its own row which depicts the study information and statistics. On the right side of the diagram,

the hedges' g value is represented as a rectangle with the confidence intervals represented as the error bars and the vertical lines representing the mean effect size (left line) and zero (right line)

moderator was coded for the number of seconds each item was presented.

Relatedness of Study Items Although only studies utilizing single words were included in our analyses, these too can be broken down into different categories. As such, each stimulus list was coded as consisting of either semantically related or unrelated words. Types of stimuli identified as semantically related included any set of items that were related to one or more categories such as a single category (e.g., all items on the study list were fruits), several categories (e.g., all items on the study list were instruments, vegetables, and sports), DRM (Deese, 1959; Roediger III & McDermott, 1995), or associative chains (Serra & Oswald, 2006). The unrelated word sets consisted of unrelated nouns (e.g., Rhodes & Castel, 2008) or concrete nouns that were also unrelated to one another (e.g., Lehmer & Bäuml, 2018a).

Inter-Item Association In the part-list cuing literature, the degree to which participants are encouraged to create associations among study items during the study phase has been a variable of theoretical interest (e.g., Bäuml & Aslan, 2006; Lehmer & Bäuml, 2018a). Of note, this notion of inter-item associations (as referenced here and elsewhere in the paper)

pertains to sequential associations established across the list of study items presented to the participant rather than associations that may be formed between the elements of a single presentation of an item (e.g., a word pair.) While these varying encoding conditions are expected to activate different cognitive mechanisms, their consequences on the extent of part-list cuing impairment are not expected to differ. In other words, differential inter-item associations created at study are expected to produce similar recall impairment albeit for different reasons. To assess the possibility if the part-list cuing impairment is different across studies that involved low versus high inter-item associations generated at study, we followed the conceptualization in the literature to code the studies. Thus, unrelated items and one single study cycle were coded as “low” inter-item association (e.g., Lehmer & Bäuml, 2018a) and related stimuli (e.g., categorized words), unrelated words with deep processing instructions (e.g., creating a story with the words), and/or multiple study phases were coded as “high” inter-item association.

Study List Length We coded this moderator as a continuous moderator representing the number of items in a set of the to-be-remembered items provided to participants during the encoding phase (i.e., the number of study items). For

between-subject designs, this represented the entirety of the study list, and for within-subject designs, this consisted of the number of items in a study list provided for each retrieval phase, that is, the control phase (free recall) and the part-list cued recall phase.

Study Modality While most of the experiments we investigated presented the items visually for study, some samples were exposed to items auditorily. To examine whether the presentation of items via tape-recordings influences the findings, we coded each sample on this categorical moderator. Importantly, in all the studies included in this moderator analyses (as well as all other analyses we report) the part-list cues at test were presented in the visual modality.

Distractor Length The distractor length used in each experiment was recorded in seconds to investigate the potential influence that the timing between encoding and retrieval phases may exert. In samples that had no distractor task, we coded the length of the distractor as zero seconds. When it came to effect size measurements originating from directed-forgetting experiments, the amount of time the second list (the one that participants did not recall in their critical retrieval task) took to encode was calculated as the distractor length.

The Number of Part-List Cues We also coded the number of cues provided during part-list cuing recall on a continuous scale. This moderator provides insight into whether or not the sheer volume of cue words moderates the size of the observed effect. It does not provide information regarding the influence of the proportion of cue words relative to the number of studied items. The latter relationship is also of interest, but the majority of experiments used 50% of the set as cues rendering a meta-regression on those proportions considerably less informative about the true influence the proportion may exert. Nonetheless, in addition to the number of part-list cues provided we conducted a mini meta-analysis comparing studies that presented over 50% of the study list as cues to those that presented less than 50%.

Recall Time Allotted Pragmatically speaking, if the amount allotted for retrieval does not influence the effect of part-list cues on recall, then this provides flexibility for allocating the time provided; for example, one might allocate less time to reduce the overall length of the experiment or allocate as much as time as the participants might take to complete the retrieval task. We coded this procedural factor in seconds. If a study failed to provide a specific amount of time or had a self-paced retrieval task, we did not include it in this meta-regression.

Item-Specific Probes As we described earlier, in studies that used item-specific probes participants in both the experimental and control conditions were provided with the unique first letter of each studied item during the retrieval phase. These probes induce a similar form of retrieval disruption in the control condition as well as the experimental condition. We coded each study as either including these probes or not.

Year of Publication Each sample in our analyses was coded for the year in which it was published. This moderator provides insight regarding publication bias and if a decline effect can be observed in this body of literature. Publication bias refers to the likelihood for statistically significant outcomes to be accepted for publication over nonsignificant findings (Rosenthal, 1979). A decline effect is characterized as a linear decline of the size of published effect sizes since the first publication of an effect (Schooler, 2011).

Statistical Approach

As mentioned earlier, we used Hedges' g to measure effect sizes, which corrects for a bias in the standardized mean difference effect size represented by Cohen's d , particularly in small samples (Borenstein et al., 2009). A value of 0.20 is considered to be a small effect size, 0.5 is considered moderate, and 0.8 is considered large (Cohen, 1988). All samples we included were required to provide sample size information per condition so that their effect sizes could be inverse-variance weighted in aggregations using CMA. Categorical moderators were tested with mixed-effects analogue-to-ANOVA and continuous moderators were tested with mixed-effects meta-regression using CMA. Hedges' g and their associated confidence intervals were determined using random-effects models, and we examined heterogeneity of these studies using Q and I^2 . We note that $I^2 < 25%$ is considered low, 50% is considered average, and $>75%$ is considered high. All our models assume normality. Publication bias was assessed by a fail-safe N analysis and a trim and fill analysis along with their corresponding funnel plots. The funnel plot is akin to a scatterplot with the observed effect sizes on the horizontal axis and the standard error on the vertical axis. Notably, when there is no publication bias then we should observe a symmetrical, upside down funnel. Lastly, we adopted a strict alpha cutoff of $p < .01$ to determine statistical significance in view of the number of comparisons reported.

Results

Due to differences in the calculation and interpretation of effect size across between-subjects and within-subject experimental designs, we report each design separately. As noted

earlier, we will report the main analyses and report the conservative and lenient analyses only for the overall weighted effect size and for where results deviate from the main analysis. The approach to separate between-subjects and within-subject designs is supported by two random effects mixed-model analog-to-ANOVAs with design type (Between vs Within) used as a moderator. Results indicate the effect size for the 47 samples that utilized between-subjects designs in the main analysis, $g = -0.55$, $CI (-0.67) - (-0.43)$, $p < .001$, was not significantly different than the effect size for the 49 samples that utilized within-subject designs, $g = -0.45$, $CI (-0.53) - (-0.38)$, $p < .001$, $Q(1) = 1.67$, $p = .196$.

The results from the conservative analysis once again revealed a strong part-list cuing impairment in both the between-subjects and within-subject designs. The difference between the two failed to reach statistical significance in this comparison, as the effect size for the 30 samples that utilized between-subjects designs in the conservative analysis, $g = -0.6$, $CI (-0.72) - (-0.49)$, $p < .001$, was not significantly different than the effect size for the 48 samples that utilized within-subject designs, $g = -0.46$, $CI (-0.54) - (-0.39)$, $p < .001$, $Q(1) = 3.93$, $p = .047$.

For the lenient analysis that included unpublished experiments, adding 13 additional measurements, the magnitude of the part-list cuing impairment in studies that utilized within-subject designs was once again present, $g = -0.39$, $CI (-0.45) - (-0.32)$, $p < .001$. As can be seen from the g values, this effect appeared to be smaller, and this decrease is in line with research on publication bias that suggests unpublished reports often have weaker or null effects (Rosenthal, 1979). Nonetheless, the part-list cuing impairment was significant in this analysis as well. As we previously noted, no unpublished experiments that used between-subjects designs met our inclusion criteria.

In brief, a robust part-list cuing impairment in recall was observed overall. We now turn to the effects of moderators on this main phenomenon.

Main Analyses

Part-List Cuing Impairment Effect Size Forty-seven effect size measurements were included in the between-subjects main analysis, which ranged from -1.2 to 0.92. Two facilitatory effects are included in this range that derive from studies with long retention periods (Bäuml & Schlichting, 2014; Lehmer & Bäuml, 2018a). These predicted findings will be discussed in greater detail later where applicable. Of these effect sizes, 6.38% reported a positive effect, 8.51% reported a null effect, and 85.11% reported a negative effect of part-list cues with a mean effect of $g = -0.55$. The effect size was significantly lower than zero, $z = -8.76$, $p < .001$. Due to the likelihood of existing but unidentified studies reporting small or no effects, a classic fail-safe N was calculated

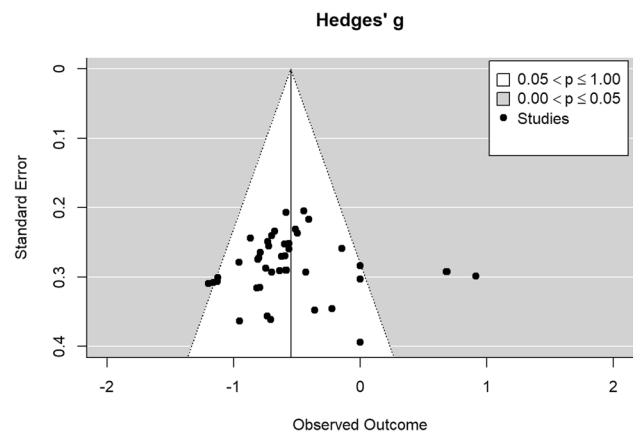


Fig. 5 Duval and Tweedie Trim and Fill Imputed Funnel Plot for Between-subjects Designs. Black circles represent observed studies in the meta-analysis. There are no white circles in this figure to represent imputed studies that are thought to be missing due to publication bias

(Rosenthal, 1979). This value represents the number of missing studies with effect sizes of zero that would have to exist to eliminate the significant difference observed, which turned out to be 2,170 missing studies, thus representing a very large number. For a Duval and Tweedie trim and fill imputed funnel plot for publication bias, see Fig. 5 (Viechtbauer, 2010). Notably, our trim and fill analysis suggests that there are no studies missing due to publication bias.

The analysis for the 49 samples included in the within-subject main analysis showed a similar outcome. The effect sizes in the analysis ranged from -1.04 to 0. Of these effect sizes 14.29% reported a null effect and 85.71% reported a negative effect of part-list cues with a mean effect of $g = -0.45$. This effect size was significantly lower than zero, $z = -11.99$, $p < .001$. The classic fail-safe N suggests that there would need to be 4,665 unidentified studies with an effect size of zero to extinguish this significant effect, again demonstrating a very large number and exceeding the fail-safe N of between-subject designs. Even though the within-subject analysis has a smaller average effect size than the between-subjects analysis, in both samples an equivalent proportion of studies (85%) reported a negative effect, and the fail-safe N for the within-subject designs was even larger than that observed for the between-subjects designs. In other words, the presence of a part-list cuing impairment and the number of null studies required to extinguish this effect is larger due to the consistency of the data in the sample and indexed by these differences in Z-scores (-11.99 for within-subject compared to -8.76 for between-subjects). For a Duval and Tweedie trim and fill imputed funnel plot for publication bias, see Fig. 6 (Viechtbauer, 2010). Our trim and fill analysis suggests that 11 studies are thought to be missing on the right side of the funnel plot due to publication bias. With

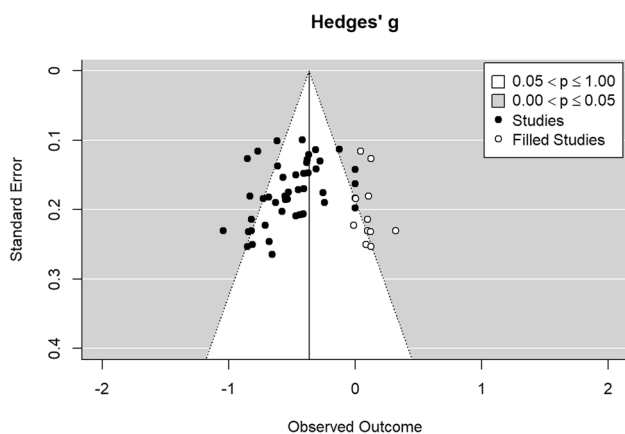


Fig. 6 Duval and Tweedie Trim and Fill Imputed Funnel Plot for Within-subject Designs. Black circles represent observed studies in the meta-analysis. White circles represent 11 imputed studies that are thought to be missing due to publication bias

these studies included, the adjusted effect size is estimated to be slightly smaller albeit still statistically significant, $g = -0.36$, CI (-0.44) - (-0.28), $p < .001$.

In both analyses, there was significant heterogeneity in the effect sizes. This pattern suggests the possibility that there are factors that moderate the size of the effect and could be potentially explained by the moderator analyses, $Q(46) = 108.59$, $p < .001$, $I^2 = 57.64$; $Q(48) = 123.89$, $p < .001$, $I^2 = 61.26$, for between-subjects and within-subject designs, respectively. We provide a summary of the findings of the moderator analyses in Tables 1 and 3.

Study Presentation Time Although the amount of time an item was presented for study did not vary widely across samples, with all but one sample being presented with items between 1.5 and 5 s per item, a mixed-effects meta-regression was conducted to examine if the length of the exposure period nonetheless had a significant influence on the size of the impairment reported. The item presentation time of the between-subjects samples ranged from 2 to 17 s and the within-subject samples ranged from 1.5 to 5 s.

The amount of time during which an item was presented did not significantly predict the size of the impairment in the between-subjects analysis of 45 samples (two samples from Basden and Basden (1995) Experiment 1 were excluded for using varying exposure time between conditions), $Q(1) = 0.72$, $p = .398$, $\beta = 0.0148$. This effect was also not significant for the within-subject analysis of 47 samples (two samples from Experiment 3 of Aslan and Bäuml (2009) were excluded for having varying exposure time between conditions), in the main analyses, $Q(1) = 5$, $p = .025$, $\beta = 0.0366$. We note that the one study to exceed the study presentation rate of 5 s per item, with 17 s per item used as the presentation rate (Basden et al., 1991), also revealed a moderate part-list cuing impairment ($g = -0.561$). This finding suggests

that item presentation rates ranging from 1.5 to 17 s result in the part-list cuing impairment in recall.

Relatedness of Study Items We also explored the influence of whether the relatedness of stimuli influence the part-list cuing impairment. In all the analyses we report, the items were classified as either having some inter-relatedness between the items or as being unrelated items.

Forty-seven effect size measurements were included in the between-subjects random effects mixed-model analog-to-ANOVA, with 36.17% of the samples being exposed to semantically related items (mean $g = -0.61$; $z = -6.74$, $p < .001$) and 63.83% to items that were not semantically related (mean $g = -0.51$; $z = -6.17$, $p < .001$). The difference between these stimulus types was not significant, $Q(1) = 0.62$, $p = .432$.

In the within-subject experiments, out of the 49 samples included, 67.35% of samples were exposed to semantically related items (mean $g = -0.47$; $z = -9.69$, $p < .001$) and 32.65% of samples were exposed to unrelated items (mean $g = -0.42$; $z = -7.07$, $p < .001$). The difference in these observations was once again not significant, $Q(1) = 0.53$, $p = .468$.

This pattern of findings held for the conservative analyses, but in the lenient analyses we observed a significant difference in the size of the effect for within-subject designs, $Q(1) = 7.65$, $p = .006$, with a greater part-list cuing impairment observed for related ($g = -0.46$) than unrelated ($g = -0.29$) study stimuli.

In summary, the related nature of the items on the study list had a non-significant influence on the size of the impairment in our main and conservative analyses, though numerical trends across all analyses share directionality of a larger part-list cuing impairment for related than unrelated study items that we observed in the lenient analyses.

Study List Length We tested the potential moderating effect of the number of items provided for participants to remember (i.e., the length of the study list) using a mixed-effects meta-regression. The study list length of the between-subjects samples ranged from 12 to 84 words and the within-subject samples ranged from 8 to 70 words. Neither the 47 samples included in the between-subjects analysis, $Q(1) = 1.19$, $p = .276$, $\beta = -0.0023$, nor the 49 samples of the within-subject analysis, $Q(1) = 2.85$, $p = .091$, $\beta = 0.0029$, yielded a significant influence of the number of study items on the size of the recall impairment.

Interitem Association We then examined if the encoding conditions associated with high versus low inter-item associations influenced the magnitude of part-list cuing impairment. Such encoding conditions have been manipulated to experimentally test if each of these conditions activates a

Table 1 Potential moderator variables of the part-list cuing impairment in recall: Main analyses

Variable	Design	Number of effect sizes	Mean effect size (<i>g</i>)	Regression slope estimation	<i>Q</i> -test
Study presentation time	Between	45	-	0.01479	0.715
	Within	47	-	0.03660	4.997
Relatedness of Study Items	Between	47			0.43
		Related	17	-0.61	-
Unrelated	Within	30	-0.51	-	-
		49			0.53
Related	Between	33	-0.47	-	
		Unrelated	16	-0.42	-
Study list length	Between	47	-	-0.00229	1.187
	Within	49	-	0.00287	2.854
Study modality	Between	47			0.161
		Visually	40	-0.537	-
Auditorily	Within	7	-0.593	-	
		49			0.003
Visually	Between	41	-0.455	-	
		Auditorily	8	-0.451	-
Distractor task length	Between	44	-	0.0005	40.16***
	Within	48	-	0.0014	21.93***
Number of Part-list Cues	Between	44	-	-0.0062	1.93
	Within	49	-	0.0041	1.88
Retrieval time allotted	Between	40	-	<-0.00001	<0.001
	Within	37	-	0.0004	3.11
Presence of Item-Specific Probes	Between	47			0.06
		Item-specific probes present	15	-0.53	-
Item-specific probes not present	Within	32	-0.56	-	
		49			2.61
Item-specific probes present	Between	17	-0.37	-	
		Item-specific probes not present	32	-0.5	-
Inter-item Association	Between	47			1.78
		High	29	-0.62	-
Low	Within	18	-0.44	-	
		49			0.11
High	Between	38	-0.46	-	
		Low	11	-0.43	-
Publication year	Between	47	-	0.00193	0.59
	Within	49	-	0.00198	0.56
Publication status	Between	45			
		Published	45	-	-
Unpublished	Within	0	-	-	-
		62			29.89***
Published	Between	49	-0.45	-	
		Unpublished	13	-0.18	-

p* < .01*p* < .001

different mechanism. We predict no difference between high and low associative encoding conditions as both types of encoding are expected to produce part-list cuing impairment

in the literature, albeit due to the activation of different mechanisms, namely retrieval disruption or retrieval inhibition, respectively.

Forty-seven effect size samples were included in the between-subjects random effects mixed-model analog-to-ANOVA, with 61.70% of the samples having high associative encoding (mean $g = -0.62$; $z = -7.69$, $p < .001$) and 38.30% having low associative encoding conditions (mean $g = -0.44$; $z = -4.45$, $p < .001$). The heterogeneity of observed effects between these methods was not significant, $Q(1) = 1.78$, $p = .182$.

The random effects mixed-model analog-to-ANOVA reported similar findings when it came to the 49 samples included in the within-subject analysis. In this analysis, 77.55% of the samples were experienced high associative encoding (mean $g = -0.46$; $z = -10.52$, $p < .001$), while 22.45% of the samples experienced low associative encoding (mean $g = -0.43$; $z = -5.58$, $p < .001$). This moderator once again did not significantly explain the heterogeneity among the effect sizes, $Q(1) = 0.11$, $p = .74$.

In other words, as expected, both high and low associative encoding conditions produced part-list cuing impairment and the magnitude of the impairment did not differ in the meta-analysis.

Study Modality This analysis tested whether the part-list cuing impairment was moderated by the modality in which the items were presented during study. In the case of the samples we selected, the items on the study list were always either presented visually to participants (either on a computer screen or on index cards) or presented auditorily for set durations using a tape recorder. The part-list cues at test were presented in the visual modality in all cases.

Forty-seven effect size measurements were included in the between-subjects random effects mixed-model analog-to-ANOVA, with 85.11% of the samples having items on the study list presented visually (mean $g = -0.54$; $z = -7.56$, $p < .001$) and 14.89% providing the items auditorily (mean $g = -0.59$; $z = -4.94$, $p < .001$). The difference between modalities was not significant, $Q(1) = 0.16$, $p = .689$.

The mixed-model analog-to-ANOVA had similar results when it came to the 49 samples included in the within-subject analysis. In this analysis, 83.67% of the samples were presented stimuli in the visual modality (mean $g = -0.46$; $z = -10.7$, $p < .001$) while 16.33% of the samples received exposure to the stimuli through an auditory medium (mean $g = -0.45$; $z = -5.25$, $p < .001$). This moderator did not significantly explain heterogeneity among the effect sizes, $Q(1) = 0.003$, $p = .960$.

In brief, the modality of study presentation does not influence the presence or the size of part-list cuing impairment in recall as far as visual versus auditory modalities are concerned. We note here that a majority of the studies have used visual presentation of study stimuli, with only a rather small sample being available to readily detect a potentially different influence of auditory encoding.

Distractor Length We also examined the length of the distractor task that occurs between the study and test phases. Recent research suggests that the direction of the effect of part-list cues could be mediated by the length of time elapsing between encoding and retrieval (Bäuml & Schlichting, 2014; Lehmer & Bäuml, 2018a). If an experiment has a short retention period between encoding and retrieval, we should anticipate the standard part-list cuing impairment. However, when a longer retention period is inserted, the study and recall test contexts differ from each other rather than overlap as would be the case in a short retention period. Due to this non-overlap between study and test contexts with longer intervals between the two, we hypothesized that part-list cues may reactivate the study context and that reactivation of context can serve as an additional cue that results in a facilitatory effect that is absent for the controls (Lehmer & Bäuml, 2018a). As the distractor task length increases, so does the retention period between the two experimental phases, making this moderator the focus of the present analysis.

The predictive value of the length of the distractor task was tested using a mixed-effect meta-regression on the continuous predictor of distractor length in seconds. The distractor length of the between-subjects samples ranged from 0 to 2,880 s and the within-subject samples ranged from 0 to 480 s. Three samples in the between-subject analysis (Rhodes & Castel, 2008, Experiments 2 and 3; Slamecka, 1968, Experiment 6) and one sample in the within-subject analysis (Cokely et al., 2006, Experiment 2a) were excluded from this analysis for having varying distractor task length between the key conditions of the experiment.

Length of distractor task was a significant positive predictor of effect sizes for 44 samples that utilized between-subjects designs, $Q(1) = 40.18$, $p < .001$, $\beta = 0.0005$, such that as the length of the distractor period increased, the size of the part-list cuing impairment decreased. This predictor was not significant in our conservative analysis; $Q(1) = 0.0003$, $p = .987$, $\beta = -0.00001$. This result is consistent with the context reactivation account (Lehmer & Bäuml, 2018a) because one criterion of our conservative analysis is that the distractor length cannot exceed 5 min, which resulted in exclusions of the longer retention intervals in the between-subjects samples. Turning to the within-subject analysis, the length of the distractor task was once again a significant positive predictor for the effect sizes of the 48 samples, $Q(1) = 21.93$, $p < .001$, $\beta = 0.0014$. Thus, this moderator's effect on the size of the part-list cuing impairment was in line with the theoretical expectations.

The Number of Part-List Cues We tested the potential moderating effect of the number of part-list cues presented during recall using a mixed-effects meta-regression. Across studies, the number of part-list cues ranged from five to 42 cues in the between-subject samples and from three to 40 cues in

the within-subject samples. Three samples from the between-subjects analysis (Alba & Chattopadhyay, 1985, Experiment 2 Men's Condition; Goernert & Larson, 1994; Roediger III et al., 1977, Experiment 1) were excluded from the analysis for providing statistics that were collapsed across conditions that received a varying number of cues. Neither the 44 samples included in the between-subject analysis, $Q(1) = 1.9$, $p = .165$, $\beta = -0.0062$, nor the 49 samples of the within-subject analysis, $Q(1) = 1.88$, $p = .170$, $\beta = 0.0041$, yielded a significant influence of the number of cues provided during retrieval.

This result is rather surprising because single cues can facilitate recall (Hudson & Austin, 1970; Robin & Moscovitch, 2017; Tulving, 1974). The explanation for a lack of an effect for the number of part-list cues could be rooted in the ratio between the number of items provided during encoding and the number of cues provided during recall. However, due to our sample criteria favoring selecting samples where 50% of the items are provided as cues (a procedure that represents the standard paradigm), the regressions reported here may be skewed to an over-representation of this ratio. For this reason, we conducted a mini-meta-analysis, with a subset of studies from our main analyses, comparing the size of the impairment in studies that provided above 50% of items as cues to those that provided below 50% of items as cues.

Sixteen effect size measurements were included in the between-subjects random effects mixed-model analog-to-ANOVA, with 87.50% of the samples presented with greater than 50% of items as part-list cues (mean $g = -0.45$; $z = -3.56$, $p < .001$) and 12.50% of the samples received less than 50% of items as part-list cues (mean $g = -0.29$; $z = -1.19$, $p = .234$). The difference between proportions was not significant, $Q(1) = 0.53$, $p = .468$.

The mixed-model analog-to-ANOVA had similar results when it came to the ten samples included in the within-subject analysis. In this analysis, 40% of the samples were presented with greater than 50% of items as part-list cues (mean $g = -0.53$; $z = -5.03$, $p < .001$), while 60% of the samples received less than 60% of items as part-list cues (mean $g = -0.36$; $z = -4.30$, $p < .001$). This moderator did not significantly explain heterogeneity among the effect sizes, $Q(1) = 1.55$, $p = .212$.

As can be seen in the above analyses, as well as Table 2, when the proportion of cues was above 50%, the effect sizes were numerically greater yet not significantly different from those below 50%. This pattern was consistent across designs. However, due to the small number of samples that deviated from 50%, and disproportionate numbers of samples in each bin, these shifts in effect sizes should be interpreted with caution and call for systematic empirical tests.

Recall Time Allotted Another factor that we examined was the number of seconds participants were given to complete the recall task.

Table 2 Mean effect size of the proportion of cues to studied items

Proportion	Between	Within
Above 50%	-0.50 (14)	-0.53 (4)
Exactly 50%	-0.57 (28)	-0.46 (39)
Below 50%	-0.29 (2)	-0.36 (6)

Note. The first number represents the mean Hedges' g , whereas the number in the parentheses indicates how many samples were included in that mean

We excluded seven of the samples in the between-subjects analysis (both samples in Basden & Basden, 1995, Experiment 5; both samples in Dagnall et al., 2007, Experiment 1; Goernert & Larson, 1994; and both samples in Sloman et al., 1991, Experiment 2 and 3) and 12 samples in the within-subject analysis (both conditions in Aslan & Bäuml, 2009, Experiments 2 and 3; Basden et al., 2002 Experiments 1, 2, and 3; both conditions Experiment 4 of Dewhurst et al., 2009; Reysen & Nairne, 2002, Experiments 1 and 2; and Serra & Oswald, 2006, Experiment 3) because the recall portion of the experiment was self-paced and/or time taken was not reported, and, in turn, could not be coded into seconds.

A mixed-effect meta-regression on the continuous predictor of retrieval time allotted was tested to examine whether the number of seconds provided for the recall was a predictor of the size of the part-list cuing impairment. The allotted time varied from 24 to 600 s across the between-subjects samples and from 17 to 540 ss across the within-subject samples. The time allotted was not a significant predictor of the size of the impairment observed either for the 40 samples examined in the between-subjects analysis, $Q(1) < 0.001$, $p = .988$, $\beta < -0.00001$, or for the 37 samples included in the within-subject analysis, $Q(1) = 3.11$, $p = .078$, $\beta = 0.0004$.

Item-Specific Probes Next, we investigated whether the item-specific probes that the experimenter provided to participants during recall influenced the size of the impairment. This recall method has been utilized in many studies in the literature and is hypothesized to minimize the disparity of retrieval disruption between the experimental and control conditions. This modification is applied by making every item in the study set start with a unique first-letter or unique first-two-letter combination and providing these letters to participants in both conditions during recall. We predict a smaller difference between the control and part-list cued participants when item-specific probes are provided.

Forty-seven effect size samples were included in the between-subjects random effects mixed-model analog-to-ANOVA, with 31.91% of the samples having item-specific probes (mean $g = -0.53$; $z = -4.38$, $p < .001$) and 68.09% not providing item-specific probes (mean $g = -0.56$; $z = -7.63$,

$p < .001$). The heterogeneity of observed effects between these methods was not significant, $Q(1) = 0.06, p = .812$.

The random effects mixed-model analog-to-ANOVA reported similar findings when it came to the 49 samples included in the within-subject analysis. In this analysis, 34.69% of the samples were presented with item-specific probes during recall (mean $g = -0.37; z = -5.33, p < .001$) while 65.31% of the samples did not receive item-specific probes during recall (mean $g = -0.5; z = -11.29, p < .001$). This moderator once again did not significantly explain the heterogeneity among the effect sizes, $Q(1) = 2.61, p = .106$.

When these results are put into a theoretical context, they conflict with the retrieval-strategy disruption account of the part-list cuing impairment (Baden et al., 1977). Based on the retrieval-strategy disruption hypothesis, part-list cues interfere with the rememberer's retrieval-strategy. As such, it is expected that there will be a significant difference in the impairment when participants in both conditions have their retrieval strategies disrupted. While the numerical patterns observed align with the direction of this prediction, the analyses suggest that the use of item-specific probes does not significantly moderate the impairment. We will return to this topic in more detail in the *General discussion*.

Year of Publication We tested the decline effect, which is the systematic decrease in the effect size relative to the year in which a study on a specific phenomenon was published (Schooler, 2011), using a mixed-effect meta-regression using the continuous predictor of the year of publication. The publication year of the between-subjects samples ranged from 1968 to 2020 and the within-subject samples ranged from 1977 to 2014. The year of publication was not a significant predictor of effect sizes for between-subject designs, $Q(1) = 0.59, p = .444, \beta = 0.0019$, or for the effect sizes included in the within-subject designs, $Q(1) = 0.56, p = .455, \beta = -0.002$.

This pattern of findings held for the conservative analyses, but in the lenient analyses we observe a significant decline effect for within-subject designs, $Q(1) = 7.44, p = .006, \beta = 0.006$. This outcome is not wholly surprising as aside from Kimball (2000, Experiment 3) the unpublished reports included in this analysis are more recent than all the other studies included in the analyses. This confounds the publication year and the publication status within the regression. As noted earlier, a predictable decrease in the size of the effect size is expected in unpublished reports (Rosenthal, 1979). Nonetheless, as noted earlier, even with the inclusion of unpublished studies (in this case, all within-subject designs) the analyses revealed a significant part-list cuing impairment in within-subject experiments and the fail-safe N remained very large.

Publication Status Finally, for inclusion in our lenient sample an effect size had to originate from either a published paper that had gone through the peer-review process or a

dissertation/thesis that went through a committee of academics. We ran publication status as a moderator to gain insight on how this may account for the differences in moderator findings between our main and lenient analyses.

The random-effects mixed-model analog-to-ANOVA of the 62 samples included in the lenient analysis demonstrate a significant effect of publication status in within-subject designs. In this analysis, 20.97% of the samples were from unpublished studies (mean $g = -0.18; z = -5.48, p < .001$) while 70.03% of the samples were from published samples (mean $g = -0.45; z = -11.99, p < .001$). This moderator was a significant source of heterogeneity in this analysis, $Q(1) = 29.89, p < .001$. These patterns suggest while both unpublished and published studies report a part-list cuing impairment, published studies have significantly larger effect sizes than unpublished. No unpublished studies met inclusion criteria to be included in the between-subjects analyses.

Discussion

Our analyses show that the part-list cuing impairment in recall is a robust and consistent phenomenon. This effect is resilient in the context of a majority of experimental factors that differ across the coded studies. Regardless of whether investigators use a between-subjects or a within-subject design or whether they use a narrower or a broader set of criteria for sample selection, the part-list cuing impairment in recall remains stable. Based on our analyses, part-list cuing produces a medium-sized effect (Cohen, 1988) of impairment in recalling target items that the participants studied earlier.

Before diving into our results, it is important to highlight the approach we took to interpret our findings. Since each factor we explored was tested in five different analyses (between-subjects, within-subject, conservative between-subjects, conservative within-subject, and lenient-range within-subject [no between-subjects studies were available for inclusion in this last analysis]), there will be instances where the findings are largely but not entirely consistent across all analyses. Our recommendations will be guided by the preponderance of the evidence we observed as well as by the prior literature in instances where reconciling the differences would require an undue amount of speculation. The reader should also bear in mind that all moderator analyses are independent and do not account for the influence of complex interactions in the experimental designs.

Main Goal

Overall Design Based on the overall weighted effect size across the samples included in our main analysis, between-subjects designs did not significantly differ from

within-subject designs. As within-subject designs tend to mitigate noise in the data produced by individual differences, and typically require a smaller sample, we recommend using a within-subject design unless the research goals necessitate a between-subjects design. Additionally, according to our fail-safe N analyses, within-subject designs would require substantially more unidentified reports than the between-subject designs to extinguish the effect (i.e., 4,665 compared to 2,170 reports). At the same time, if a researcher is looking to maximize the size of the impairment, a between-subjects design should be considered as the average weight effect size was numerically larger in our main analysis and the conservative analysis. Regardless, the impairment reliably occurs in both design types, which provides flexibility to investigators for selecting the more suitable design for their purposes.

Stimulus Selection Relatedness of study items is a common experimental feature that differs across studies in prior literature. We coded it as a dichotomous categorical variable and found that both related and unrelated word lists at study produce the part-list cuing impairment in recall. Therefore, the broader goals of the experiment can guide the selection of stimuli. For example, related stimuli are useful because these word lists allow computation of subjective organization of recall in more ways than do unrelated words (Roediger et al., 1971), and can provide additional insights into the recall process as relevant to the motivation of the experiment. In terms of replication goals, we recommend keeping the stimuli constant across experiments due to potential variations that the related/unrelated stimuli can produce in recall impairment considering the significant result being observed in the lenient and similar numerical trend in all analyses.

As for the length of the study list, according to all of our analyses this factor does not have a significant predictive value on the part-list cuing impairment. Considering these results, researchers need not be overly concerned with the overall number of items presented at study when aiming to observe the part-list cuing impairment. However, it should be noted that the studies included in our analyses only cover a range of 8–84 study items, and thus we recommend caution when using a study list outside of this range as our analyses may not capture substantial deviations from this range. Further, different hypotheses may also take into account different study list length outside this range as may be suitable to test the limits of the part-list cuing impairment.

Number of Part-List Cues Our meta-analysis also suggests that the number of part-list cues provided during recall does not moderate the size of the impairment. This finding supports some reports in the prior literature (Goernert & Larson, 1994; Watkins, 1975) but contrasts others (Marsh et al., 2004, Experiments 1 and 2; Roediger III, 1973; Roediger III et al., 1977; Rundus, 1973).

We followed up this somewhat surprising finding with a mini-meta-analysis that we conducted on a subset of the samples that deviated from 50% of the items provided as part-list cues to elucidate our null findings. As a reminder, we chose not to conduct an overall analysis on this proportion due to our selection criteria favoring conditions where 50% of the studied items were provided as part-list cues and would unevenly weight a meta-regression at 50%. We made this selection so that when multiple conditions compared to a single control, we could select a condition to include in a systematic fashion and still maintain sample independence. Our review of the literature indicated that presenting 50% of the studied items as cues most closely aligned with the standard part-list cuing paradigm researchers employed. This criterion created consistency in the selection process but resulted in the number of samples exposed to more or less than 50% being under-represented. The under-representation of these conditions calls for caution in assessing the outcome of both the meta-regression pertaining to the number of cues as well as the supplementary mini-meta-analysis we conducted.

Our mini-meta-analysis did not yield significant differences regarding cue proportion when contrasting studies that presented above 50% of studied items as cues to those that presented less than 50% of studied items as cues. This outcome is generally in line with the numerical patterns in the samples included in our main analysis (see Table 2), providing participants with more than 50% of the study items as cues does not substantially increase the size of the impairment. Conversely, there does appear to be a consistent numerical pattern of a smaller impairment when providing less than 50% of the study items as cues.

To further clarify the effects of the number of cues provided for designing future experiments, we also provide numerical patterns from prior studies (Table 2). When designing an experiment, if a researcher is looking for a reliable size of the impairment, then providing 50% of the study list as cues should be considered a conservative design choice. If a study design requires cue proportions that are higher than 50%, the numerical patterns of the effect sizes suggest that this choice will not have a substantial impact on the size of the impairment. However, with respect to lower than 50% cues, while the impairment can be found, the size of the impairment has the potential to be sensitive to this proportion as suggested by the numerical patterns we observed (Table 2). We emphasize that future empirical exploration is needed to reliably determine whether the number of part-list cues provided changes the size of the impairment (Table 3).

Encoding Procedures Our results for the study presentation time are straightforward to interpret. As long as researchers stay within the normative range of 2–5 s of presentation

Table 3 Moderator analyses summary by analysis

Moderator	Between-subjects		Within-subject		
	Main	Conservative	Main	Conservative	Lenient
Study Presentation Time	☒	☒	☒	☒	☒
Relatedness of Study Items	☒	☒	☒	☒	✓
Study List Length	☒	☒	☒	☒	☒
Study Modality	☒	☒	☒	☒	☒
Distractor Length	✓	☒	✓	✓	✓
The Number of Part-list Cues	☒	☒	☒	☒	☒
Recall Time Allotted	☒	☒	☒	☒	☒
Interitem Association	☒	☒	☒	☒	☒
Item-Specific Probes	☒	☒	☒	☒	☒
Publication Year	☒	☒	☒	☒	✓
Publication Status	-	-	-	-	✓

Note. A “☒” indicates that the moderator was found to be significant below a .01 threshold for an analysis, whereas a “✓” indicates that the moderator exceeded a .01 threshold, and a “-” indicates that the moderator was not able to be tested for that analysis

rate per item, this procedural detail should bear no influence on the impairment. However, if researchers want to select presentation times that substantially exceed this range, the literature is not extensive because only three samples in our analyses exceeded presentation times of 5 s per item, and none has presented items for less than 1.5 s each. In other words, while we did not find the study time per item to be a significant predictor of the impairment, we also did not find information that speaks to the impact of large changes in the presentation times on part-list cued recall and thus recommend caution when departing from the bound of the range of our analyses.

Another procedural detail of interest at study is the modality of item presentation at study. We found that the study modality did not have a significant impact on the part-list cuing impairment as a function of auditory (e.g., Roediger III & Schmidt, 1980; Slamecka, 1968; Sloman, 1991) or visual presentations at study (e.g., Barber & Rajaram, 2011; Basden & Basden, 1995; Reysen & Nairne, 2002). Thus, if an auditory presentation is preferred due to other requirements of the experimental procedure (e.g., children or visually impaired populations), this modification is not expected to significantly impact the impairment. We note here that very little information is available on the presentation modality of the part-list cues themselves, the reason why we did not include this factor in our analyses. To the extent that modality of cue presentation can be relevant (e.g., as noted for special populations above), this question awaits future research for answers.

The last encoding procedure of interest is the degree of inter-item association (i.e., low vs. high associative encoding). Our analyses suggest that both encoding conditions will elicit the part-list cuing impairment, and this is consistent with previous work (Bäuml & Aslan, 2006; Lehmer

& Bäuml, 2018a). Therefore, based on our analyses, part-list cuing impairment is likely to be observed regardless of whether the participants study unrelated words only once or higher associative items (e.g., categorized words, asking participants to form a story with unrelated words), or are exposed to repeated study sessions that allow formation of greater inter-item associations.

Distractor Length While we did not find a significant influence of several encoding procedures or stimulus selection on the size of the part-list cuing impairment, the results of our analyses suggest that the length of the distractor can be an important factor. Our findings align with prior research that increased retention periods reduces the size of the part-list cuing impairment in recall (Bäuml & Schlichting, 2014; Lehmer & Bäuml, 2018a). In a majority of our analyses, the length of time between study and test had a significant negative linear relationship with the size of the part-list cuing impairment. Furthermore, consistent with this pattern, in our conservative, between-subjects analyses where studies with distractor phases longer than 5 min were excluded, distractor length was no longer a significant predictor. As noted earlier, this pattern supports the context reactivation account whereby this effect of longer distractor periods could be related to the relationship between context at study and context at test (Lehmer & Bäuml, 2018a). The context between the two phases overlaps to a greater extent when a short distractor period is used. After a long distractor period, part-list cues are presumed to take on the role of an additional contextual cue. Under these conditions, part-list cuing can assist in recall.

In short, a long retention period between study and recall will likely reduce the size of an impairment in a part-list cuing experiment. This prediction is guided by our analyses as well as the prior literature (Bäuml & Schlichting, 2014;

Lehmer & Bäuml, 2018a). Therefore, the retention period should not greatly exceed 5 min unless it is related to the a priori research goals of an experiment.

Retrieval Procedures We now turn our attention to the details of the retrieval task. Based on the variations we found in the literature, we evaluated the time given to participants to perform the recall task. We found that the amount of task time was not a significant predictor in any of our analyses. These patterns suggest that the part-list cuing impairment occurs regardless of the length of time provided for the task. This conclusion is further supported by studies that reported a part-list cuing impairment under conditions that allowed an unlimited amount of time for recall (e.g., Bovee et al., 2009; Brown & Hall, 1979; Sloman et al., 1991). Consequently, we conclude that the amount of time for recall task completion does not significantly impact the size of the impairment.

Other Procedural Features The final set of analyses to discuss is the use of item-specific probes in the part-list cued recall task. To recap, item-specific probes procedures are those when we provide participants with the unique first letter or first two letters of each studied item during retrieval. As the use of such cues sets up a specific order for recall in both the part-list cuing and control conditions, this sequence is expected to interfere with the idiosyncratic retrieval sequence that people use, and, consequently, it should provide some disruption to the retrieval-strategy for participants in both conditions.

Our meta-analyses suggest that item-specific probes do not significantly influence the size of the part-list cuing impairment despite the possibility that participants in the free-recall condition also experience a form of disruption when faced with item-specific probes. Due to the restrictive demands this procedure places on stimulus selection and the theoretical underpinnings of this procedure, we recommend that researchers refrain from using item-specific probes unless the motivation of the experiment requires the use of this feature. With respect to stimulus selection, item-specific probes require that each item on the study list has a unique first letter (or a first two-letter combination) which limits the study stimuli and potentially poses challenges when drawing stimuli from established norms (such as Battig & Montague, 1969 and Van Overschelde et al., 2004). With respect to the theoretical considerations, we discuss this process in more detail in the section on the retrieval-strategy disruption hypothesis.

Theoretical Implications

In the discussion so far, we have interpreted our meta-analyses in the context of certain procedural details influencing the size of the impairment and how to prevent the undue

influence of these procedural details when these procedures may be unrelated to the goal of an experiment. We now turn to the prevalent theoretical accounts for the part-list cuing impairment in recall that we described in our Introduction. While the methodological comparisons guided our meta-analyses, the results reported here have implications that inherently call for placing them in a proper theoretical context. We also remind the reader that the moderator analyses were not intended to, and do not, address potential interactions of design elements.

Retrieval-Strategy Disruption Hypothesis Retrieval-strategy disruption is the proposal that part-list cuing interferes with the idiosyncratic strategy, such as the sequence in which to recall the studied items, that the rememberer develops for the studied items. The part-list cues are assumed to disrupt this strategy and thereby lower recall. This hypothesis has received strong support in the literature from two lines of research. One, the finding that on subsequent free-recall tasks, where no cues are present, those who previously exhibited a part-list cuing impairment have a rebound in performance that matches control performance (e.g., D.R. Basden & Basden, 1995; B.H. Basden et al., 1991; Bäuml & Aslan, 2006). Two, experiments where the order in which the cues were presented at test aligned with the study order for the items (see Deese & Kaufman, 1957; Kahana, 1996), thus minimizing disruption to the retrieval strategy when presented with part-list cues, the impairment was reduced (e.g., Basden & Basden, 1995; Fritz & Morris, 2015; Garcia-Marques et al., 2012; Reysen & Nairne, 2002; Serra & Nairne, 2000). Although the experimenter-determined test order is not necessarily the idiosyncratic order that a participant might develop, it is a better match for the studied information than a subset of cues taken from different parts of the study list. While we did not test this design feature in our analyses due to a limited sample of available measurements as well as the selection criteria we needed to set, our analyses can speak to other lines of research that support retrieval-strategy disruption as a mechanism involved in the part-list cuing impairment.

Our results for the use of item-specific probes mainly relate to the retrieval-strategy disruption hypothesis. Item-specific probes reduce the disparity in the disruption to the retrieval-strategy across the experimental and control conditions in a part-list cuing experiment.³ When provided with item-specific probes, we expect that the part-list cuing impairment would

³ According to the previous literature, this outcome is expected to occur when serial retrieval plans have been built up and emphasized and is not necessarily the case when serial retrieval plans are not strengthened. Thus, outcome of studies may depend on these possibilities.

decrease if retrieval-strategy disruption is the sole explanation for the effect. That is, prompting participants to recall the items in a specific order intuitively should be disruptive to their planned retrieval-strategy. Consistent with this reasoning, when Aslan and Bäuml (2007) directly compared conditions provided with and without item-specific probes, and in some instances, item-specific probes significantly reduced the size of the impairment. In our meta-analysis, however, we did not find a significant reduction in the part-list cuing impairment based on the presence of item-specific probes. Taken together, a plausible interpretation of these findings is that in the studies sampled for our meta-analysis, the reduction in disparity between the disruption of the two conditions was not enough to capture a significant relationship. Together the direct comparisons in the empirical evidence described above (Aslan & Bäuml, 2007), and our findings suggest that item-specific probes likely impact recall performance in the part-list cuing paradigm but that the size of this impact is not substantial.

Thus, aligning with the previous literature on the accounts of cognitive mechanisms responsible for the impairment, it seems likely that retrieval strategy disruption plays a role in the occurrence of the impairment, but only when the strategy is bolstered during the encoding process and the study-test context matches to enable part-list cues to produce disruption (Lehmer & Bäuml, 2018a).

Retrieval Inhibition The second major hypothesis we considered falls under the umbrella of the competition-at-retrieval hypothesis, specifically *retrieval inhibition*. This hypothesis proposes that when participants receive part-list cues, the reading and covert retrieval of the cues increases the accessibility of the cued items and, in turn, decreases the probability of recalling the target items. This lack of accessibility to the target items can have two main consequences on memory. The cue items can block the target items such that the rememberer cannot access these items during recall. However, on a recognition test where the studied target item is provided, the participant can recognize it. Two, the cue items block and inhibit the non-cued items such that the rememberer can fail to recall and also fail to recognize the items even when presented with them. In other words, part-list cue exposure can have a long-lasting, inhibiting effect on memory (Bäuml, 2008; Rundus, 1973).

The retrieval inhibition mechanism is thought to occur alongside the disruption to retrieval-strategy, particularly when the encoding situation does not bolster strong retrieval strategies, and accounts for the lines of evidence that the retrieval-strategy disruption hypothesis cannot explain. As previously noted, when provided with a second, free-recall task, participants exhibit a rebound in the recall of target studied items, and this effect supports the retrieval disruption hypothesis (D.R. Basden & Basden, 1995; B.H. Basden et al., 1991; Bäuml & Aslan, 2006; Bäuml & Schlichting, 2014; Muntean & Kimball, 2012; Roediger III et al., 1977).

However, when participants reliably fail to recall the target studied items on the second, free-recall task, the retrieval inhibition hypothesis can account for this long-lasting inhibition that is sometimes present on the second, free-recall task (Barber & Rajaram, 2011; Bäuml & Aslan, 2006; Del Missier & Terpini, 2009; Muntean & Kimball, 2012).

The finding from our meta-analyses that provides support for retrieval inhibition is the effect of item-specific probes. As we noted in our discussion of the retrieval-strategy disruption hypothesis, we did not find item-specific probes to have a significant impact so as to eliminate the part-list cuing effect. According to Aslan and Bäuml (2007), observing a part-list cuing impairment when providing item-specific probes challenges the retrieval strategy account for the impairment, as both cued and control conditions face disruption to their preferred retrieval strategy. In our meta-analysis, we observed a significant part-list cuing impairment across studies that utilized item-specific probes. This finding supports the retrieval inhibition hypothesis since the impairment is consistently observed even when the disparity to the disruption of retrieval strategies is minimized across conditions (Lehmer & Bäuml, 2018a).

Multi-Mechanism Hypothesis In our findings, the length of the distractor period, that is, the delay between study and test turned out to be an influential procedural factor for observing the part-list cuing impairment in the standard part-list cuing procedure. This outcome lends support to the multi-mechanism hypothesis that includes the operation of the context reactivation mechanism, in addition to the retrieval disruption and retrieval inhibition mechanisms, to account for the part-list cuing impairment in recall. Our findings show that as the overlap in the contexts between the study and retrieval phases increases (as would the case with shorter study-test delays), so does the magnitude of the recall impairment. Further, when access to the study context is impaired at test, as is the case with longer distractor periods, part-list cues help reinstate the study context. In this situation, if the study conditions do not encourage development of an idiosyncratic retrieval plan, part-list cues can actually facilitate recall (Lehmer & Bäuml, 2018a, Experiment 2). Together, such evidence supports the notion that the time difference between study and recall can have a moderating effect on impairment, such that the benefits of having the study context reactivated at test can mitigate and supersede the impairment caused by retrieval strategy disruption and retrieval inhibition, respectively (Lehmer & Bäuml, 2018a, 2018b).

Conclusion

Our meta-analysis examined a counterintuitive phenomenon in memory. When individuals recall studied information in the presence of a subset of those study items intended to

serve as retrieval cues, their recall is reduced compared to a control condition where no cues are presented for recalling the studied information. This counterintuitive phenomenon where retrieval cues hurt rather than aid recall is known as the part-list cuing impairment in recall. In our meta-analysis, we undertook a thorough survey of the past literature. We considered individual design elements relevant for designing a part-list cuing experiment. Overall, this set of meta-analyses shows that the part-list cuing impairment in recall is robust, it occurs in both between-subjects and within-subject designs, and it is resilient in response to many procedural deviations that have been implemented across studies.

Our quantitative review also showed that we can consistently expect a negative medium effect (Cohen, 1988) for part-list cuing in recall tasks. This effect appears to be relatively stable regardless of: (1) whether the to-be-recalled items are related or unrelated to one another; (2) the number of cues given at retrieval; (3) the modality – visual versus auditory – in which items are presented for study; (4) the length of presentation of study items; (5) the amount of time allotted for recall; and (6) the presence of item-specific probes. The most influential moderator factor of the part-list cuing impairment we found was the retention period between study and retrieval. This factor produced a stable influence on the impairment when considerably long retention periods, for example, 30 min, were used.

From a procedural standpoint, our meta-analysis provides several options for investigators to consider when designing their experiments, as outlined above. From a theoretical standpoint, the meta-analytic findings based on over 90 samples reinforces a multiple mechanism account that researchers have proposed and investigated in individual empirical studies.

The counterintuitive phenomenon of a part-list cuing impairment has been a prevalent and recurring topic in a large number of reports in cognitive psychological research over the last five decades. Partial retrieval cues can be relevant, in a harmful way, to remembering even though their use may be well-intentioned in real-world situations such as context given to witnesses by law enforcement or examples provided by an instructor on an exam. We provide this meta-analytic review to serve as an anchor for studies that may be designed to replicate this memory impairment, test its nature, examine its theoretical accounts, extend its boundaries across a broad array of contexts and populations, and examine its applications to real-world scenarios. In this context, it is worth noting that while the current meta-analyses focused on simple stimuli (e.g., word lists), it is reasonable to assume that many of the findings will generalize to more ecologically valid stimuli such as prose and other more complex stimuli where the impairment has been observed in the prior literature (Bäuml & Schlichting, 2014; Fritz & Morris, 2015; Wallner & Bäuml, 2020). Such an analytical foundation is important to have because, after all, memory is imperfect, and at times, remembering can be a finicky process.

Author Note Preparation of this research was in part supported by the National Science Foundation Grant 1456928 and National Science Foundation Graduate Research Fellowship 1839287. The authors do not have any conflicts of interest.

Code Availability The software used for this manuscript was Comprehensive Meta-Analysis 2.0 (CMA 2.0). This software is purchasable from Biostat Inc. (<https://www.meta-analysis.com/>).

Authors' Contributions N.P. and S.R. conceived this review and planned it based on substantive discussions. A.M. joined the project during development of the statistical approach. The three authors collaborated to design and refine the method of the protocol. In consultation with S.R., N.P. led the implementation and execution of the project, data coding and extraction, and data analyses. T.P. joined the project to assist with double coding of pertinent information extracted from the literature and double-checking of all statistical analyses for accuracy. The four authors worked together to interpret the findings. N.P. drafted the first version of the manuscript in consultation with S.R., after which all four authors contributed to the manuscript preparation and submission process in a collaborative fashion. All authors approved the final draft of the manuscript.

Funding Preparation of this research was in part supported by the National Science Foundation Grant 1456928 and National Science Foundation Graduate Research Fellowship 1839287.

Data Availability The full dataset of coded moderators and effect sizes will be made available upon request, please direct inquiries to the corresponding author.

Declarations

Conflicts of Interest/Competing Interests The authors do not have any conflicts of interest.

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

References

Note. References Indicated with an Asterisk Were Included in the Meta-Analytic Analyses.

- *Alba, J. W., & Chattopadhyay, A. (1985). Effects of context and part-category cues on recall of competing brands. *Journal of Marketing Research*, 22(3), 340–349.
- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1063.
- Anderson, M. C., Green, C., & McCulloch, K. C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1141–1159.
- Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review*, 102, 68–100.

- Andersson, J., Hitch, G., & Meudell, P. (2006). Effects of the timing and identity of retrieval cues in individual recall: An attempt to mimic cross-cueing in collaborative recall. *Memory, 14*(1), 94–103.
- *Aslan, A., & Bäuml, K. H. T. (2007). Part-list cuing with and without item-specific probes: The role of encoding. *Psychonomic Bulletin & Review, 14*(3), 489–494.
- *Aslan, A., & Bäuml, K. H. T. (2009). The role of item similarity in part-list cueing impairment. *Memory, 17*(7), 697–707.
- *Aslan, A., Bäuml, K. H. T., & Grundgeiger, T. (2007). The role of inhibitory processes in part-list cuing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(2), 335–341.
- *Aslan, A., & John, T. (2019). Part-list cuing effects in younger and older adults' episodic recall. *Psychology and Aging, 34*(2), 262–267.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press.
- *Barber, S. J., & Rajaram, S. (2011). Collaborative memory and part-set cueing impairments: The role of executive depletion in modulating retrieval disruption. *Memory, 19*(4), 378–397.
- *Basden, D. R., & Basden, B. H. (1995). Some tests of the strategy disruption interpretation of part-list cuing inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(6), 1656–1669.
- *Basden, B. H., Basden, D. R., Church, B. A., & Beapre, P. (1991). Setting boundary conditions on the part-set cuing effect. *Bulletin of the Psychonomic Society, 29*(2), 213–216.
- *Basden, D. R., Basden, B. H., & Galloway, B. C. (1977). Inhibition with part-list cuing: Some tests of the item strength hypothesis. *Journal of Experimental Psychology: Human Learning and Memory, 3*(1), 100–108.
- *Basden, B. H., Basden, D. R., & Stephens, J. P. (2002). Part-set cuing of order information in recall tests. *Journal of Memory and Language, 47*(4), 517–529.
- Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80*(3, Pt.2), 1–46.
- Bäuml, K. H. (2008). Inhibitory processes. In H. L. Roediger III (Ed.), *Cognitive psychology of memory* (pp. 195–220). Elsevier.
- *Bäuml, K. H. T., & Aslan, A. (2004). Part-list cuing as instructed retrieval inhibition. *Memory & Cognition, 32*(4), 610–617.
- *Bäuml, K. H. T., & Aslan, A. (2006). Part-list cuing can be transient and lasting: The role of encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 33–43.
- *Bäuml, K. H. T., Kissler, J., & Rak, A. (2002). Part-list cuing in amnesic patients: Evidence for a retrieval deficit. *Memory & Cognition, 30*(6), 862–870.
- Bäuml, K. H. T., & Kuhbandner, C. (2003). Retrieval-induced forgetting and part-list cuing in associatively structured lists. *Memory & Cognition, 31*(8), 1188–1197.
- *Bäuml, K. H. T., & Samenieh, A. (2012). Influences of part-list cuing on different forms of episodic forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(2), 366–375.
- *Bäuml, K. H. T., & Schlichting, A. (2014). Memory retrieval as a self-propagating process. *Cognition, 132*(1), 16–21.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. *Comprehensive meta-analysis version 2*. Englewood, NJ, Biostat; 2005.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Wiley.
- Bovee, J. C., Fitz, C., Yehl, G., Parrott, S., & Kelley, M. R. (2009). Applied part-set cuing. In *Applied memory, 73–87*. Publishers.
- Brown, A. S., & Hall, L. A. (1979). Part-list cueing inhibition in semantic memory structures. *The American Journal of Psychology, 351–362*. <https://doi.org/10.2307/1421929>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge Academic.
- *Cokely, E. T., Kelley, C. M., & Gilchrist, A. L. (2006). Sources of individual differences in working memory: Contributions of strategy to capacity. *Psychonomic Bulletin & Review, 13*(6), 991–997.
- Cole, S. M., Reysen, M. B., & Kelley, M. R. (2013). Part-set cuing facilitation for spatial information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(5), 1615–1620.
- *Crescentini, C., Marin, D., Del Missier, F., Biasutti, E., & Shallice, T. (2011). Interference from retrieval cues in parkinson's disease. *Neuropsychology, 25*(6), 720–733.
- Crescentini, C., Shallice, T., & Macaluso, E. (2010). Item retrieval and competition in noun and verb generation: An fMRI study. *Journal of Cognitive Neuroscience, 22*(6), 1140–1157.
- *Dagnall, N., Parker, A., & Munley, G. (2007). Effects of part-set cuing on reminiscence. *Perceptual and Motor Skills, 105*(3), 1015–1022. [10.2466/2Fpms.105.3.1015-1022](https://doi.org/10.2466/2Fpms.105.3.1015-1022)
- Deese, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports, 5*(3), 305–312.
- Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology, 54*(3), 180–187.
- Del Missier, F., & Terpini, C. (2009). Part-set cuing in option generation. *Memory & Cognition, 37*(3), 265–276.
- *Dewhurst, S. A., Bould, E., Knott, L. M., & Thorley, C. (2009). The roles of encoding and retrieval processes in associative and categorical memory illusions. *Journal of Memory and Language, 60*(1), 154–164.
- Drinkwater, K., Dagnall, N., & Parker, A. (2006). Effects of part-set cuing on experienced and novice chess players' reconstruction of a typical chess midgame position. *Perceptual and Motor Skills, 102*(3), 645–653.
- Fritz, C. O., & Morris, P. E. (2015). Part-set cuing of texts, scenes, and matrices. *British Journal of Psychology, 106*(1), 1–21.
- Foos, P. W., & Clark, M. C. (2000). Old age, inhibition, and the part-set cuing effect. *Educational Gerontology, 26*(2), 155–160.
- Garcia-Marques, L., Garrido, M. V., Hamilton, D. L., & Ferreira, M. B. (2012). Effects of correspondence between encoding and retrieval organization in social memory. *Journal of Experimental Social Psychology, 48*(1), 200–206.
- Garcia-Marques, L., Hamilton, D. L., & Maddox, K. B. (2002). Exhaustive and heuristic retrieval processes in person cognition: Further tests of the TRAP model. *Journal of Personality and Social Psychology, 82*(2), 193–207.
- Garrido, M., Garcia-Marques, L., & Hamilton, D. (2012). Enhancing the comparability between part-list cueing and collaborative recall: A gradual part-list cueing paradigm. *Experimental Psychology, 59*(4), 199–205.
- *Godbole, N. (2012). Assessing temporal associations in recall: A new take on the strategy disruption account of the part-set cuing effect. *The University of North Carolina at Greensboro*.
- *Goernert, P. N., & Larson, M. E. (1994). The initiation and release of retrieval inhibition. *The Journal of General Psychology, 121*(1), 61–66.
- *Grimaldi, P. J. (2014). *Retrieval enhanced learning: Exploring the role of retrieval strategies* (Doctoral dissertation, Purdue University).
- Hudson, R. L., & Austin, J. B. (1970). Effect of context and category name on the recall of categorized word lists. *Journal of Experimental Psychology, 86*(1), 43–47.
- Huffman, C. J., Matthews, T. D., & Gagne, P. E. (2001). The role of part-set cuing in the recall of chess positions: Influence of chunking in memory. *North American Journal of Psychology, 3*(3), 535–542.
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition, 24*(1), 103–109.

- Kelley, M. R., Parasiuk, Y., Salgado-Benz, J., & Crocco, M. (2016). Spatial part-set cuing facilitation. *Memory*, 24(6), 737–745.
- Kelley, M. R., & Parihar, S. A. (2018). Part-set cuing impairment & facilitation in semantic memory. *Memory*, 26(7), 1008–1018.
- *Kelley, M. R., Pentz, C., & Reysen, M. B. (2014). The joint influence of collaboration and part-set cuing. *Quarterly Journal of Experimental Psychology*, 67(10), 1977–1985.
- Kimball, D. R. (2000). *The role of retrieval inhibition in the occurrence of false memories*. University of California.
- *Kimball, D. R., & Bjork, R. A. (2002). Influences of intentional and unintentional forgetting on false memories. *Journal of Experimental Psychology: General*, 131(1), 116–130.
- *Kimball, D. R., Bjork, E. L., Bjork, R. A., & Smith, T. A. (2008). Part-list cuing and the dynamics of false recall. *Psychonomic Bulletin & Review*, 15(2), 296–301.
- *Kissler, J., & Bäuml, K. H. T. (2005). Memory retrieval in schizophrenia: Evidence from part-list cuing. *Journal of the International Neuropsychological Society*, 11(3), 273–280.
- *Lehmer, E.-M., & Bäuml, K.H. T. (2018a). Part-list cuing can impair, improve, or not influence recall performance: The critical roles of encoding and access to study context at test. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(8), 1186–1200.
- Lehmer, E.-M., & Bäuml, K.-H. T. (2018b). The many faces of part-list cuing - evidence for the interplay between detrimental and beneficial mechanisms. *Frontiers in Psychology*, 9, 701–708.
- *Marsh, E. J., Dolan, P. O., Balota, D. A., & Roediger, H. L. (2004). Part-set cuing effects in younger and older adults. *Psychology and Aging*, 19(1), 134–144.
- Marx, M. H. (1988). Facilitation of free recall of category names and instances by indirect part-set cuing. *Bulletin of the Psychonomic Society*, 26(3), 195–196.
- *Mueller, C. W., & Watkins, M. J. (1977). Inhibition from part-set cuing: A cue-overload interpretation. *Journal of Verbal Learning and Verbal Behavior*, 16(6), 699–709.
- Muntean, W. J., & Kimball, D. R. (2012). Part-set cuing and the generation effect: An evaluation of a two-mechanism account of part-set cuing. *Journal of Cognitive Psychology*, 24(8), 957–964.
- Nickerson, R. S. (1984). Retrieval inhibition from part-set cuing: A persisting enigma in memory research. *Memory & Cognition*, 12(6), 531–552.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251).
- Oswald, K. M., Serra, M., & Krishna, A. (2006). Part-list cuing in speeded recognition and free recall. *Memory & Cognition*, 34(3), 518–526.
- *Park, H., & Madigan, S. (1993). Negative effects of part-set cues: Are they reversible?. *Bulletin of the Psychonomic Society*, 31(4), 311–313.
- Penney, C. G. (1988). A beneficial effect of part-list cuing with unrelated words. *Bulletin of the Psychonomic Society*, 26(4), 297–300.
- Peynircioğlu, Z. F. (1987). On the generality of the part-set cuing effect: Evidence from nonmemory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3), 437–442.
- Peynircioğlu, Z. F. (1989). The generation effect with pictures and nonsense figures. *Acta Psychologica*, 70(2), 153–160.
- Peynircioğlu, Z. F., & Gökşen-erelç, F. (1988). Part-set cuing across languages: Evidence for both word-and concept-mediated inhibition depending on language dominance. *Acta Psychologica*, 67(1), 19–32.
- Peynircioğlu, Z. F., & Moro, C. (1995). Part-set cuing in incidental and implicit memory. *The American Journal of Psychology*, 1-11. <https://doi.org/10.2307/1423098>
- Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, 14(2), 184–201. 0.1037/1040-3590.14.2.184.
- *Reysen, M. B., & Nairne, J. S. (2002). Part-set cuing of false memories. *Psychonomic Bulletin & Review*, 9(2), 389–393.
- *Rhodes, M. G., & Castel, A. D. (2008). Metacognition and part-set cuing: Can interference be predicted at retrieval?. *Memory & Cognition*, 36(8), 1429–1438.
- Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., & Mortenson, E. (2020). Racial inequality in psychological research: Trends of the past and recommendations for the future. *Perspectives on Psychological Science*, 15(6), 1295–1309.
- Robin, J., & Moscovitch, M. (2017). Familiar real-world spatial cues provide memory benefits in older and younger adults. *Psychology and Aging*, 32(3), 210–219.
- Roediger III, H. L. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior*, 12(6), 644–657.
- Roediger, H. L., & Geraci, L. (2003). *Conducting implicit memory research: A practical guide. A guide to implementing cognitive methods with clinical populations*. American Psychological Association.
- Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology*, 8 (pp. 63–131). Elsevier.
- Roediger III, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803.
- *Roediger III, H. L., & Schmidt, S. R. (1980). Output interference in the recall of categorized and paired-associate lists. *Journal of Experimental Psychology: Human Learning and Memory*, 6(1), 91–105.
- *Roediger III, H. L., Stellon, C. C., & Tulving, E. (1977). Inhibition from part-list cues and rate of recall. *Journal of Experimental Psychology: Human Learning and Memory*, 3(2), 174–188.
- Roediger III, H. L., & Tulving, E. (1974). Part-list cuing and directed forgetting. In *meeting of the Midwestern Psychological Association, Chicago*.
- Roenker, D., Thompson, C., & Brown, S. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, 76(1), 45–48.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641.
- Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior*, 12(1), 43–50.
- Schooler, J. (2011). Unpublished results hide the decline effect. *Nature*, 470(7335), 437–437.
- Serra, M., & Nairne, J. S. (2000). Part-set cuing of order information: Implications for associative theories of serial order memory. *Memory & Cognition*, 28(5), 847–855.
- *Serra, M., & Oswald, K. M. (2006). Part-list cuing of associative chains: Tests of strategy disruption. *The Journal of General Psychology*, 133(3), 301–317.
- *Slamecka, N. J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology*, 76(4, Pt.1), 504–513.
- *Sloman, S. A. (1991). Part-set cuing inhibition in category-instance and reason generation. *Bulletin of the Psychonomic Society*, 29(2), 136–138.
- *Sloman, S. A., Bower, G. H., & Rohrer, D. (1991). Congruency effects in part-list cuing inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5), 974–982.
- Todres, A. K., & Watkins, M. J. (1981). A part-set cuing effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, 7(2), 91–99.
- Tulving, E. (1974). Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean that the

- memory trace has been lost; it may only be inaccessible. *American Scientist*, 62(1), 74–82.
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, 50(3), 289–335.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48: 10.18637/jss.v036.i03.
- *Wallner, L., & Bäuml, K. H. T. (2020) Self-paced part-list cuing. *Psychonomic Bulletin & Review* <https://doi.org/10.3758/s13423-021-01910-3>
- Watkins, M. J. (1975). Inhibition in recall with extralist “cues”. *Journal of Verbal Learning and Verbal Behavior*, 14(3), 294–303.
- Watkins, M. J., & Allender, L. E. (1987). Inhibiting word generation with word presentations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(4), 564–568.
- Watkins, M. J., Schwartz, D. R., & Lane, D. M. (1984). Does part-set cuing test for memory organization? Evidence from reconstructions of chess positions. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 38(3), 498–503.
- Zellner, M., & Bäuml, K. H. (2005). Intact retrieval inhibition in children’s episodic recall. *Memory & Cognition*, 33(3), 396–404.

Open Practices Statement

None of the data for the analyses reported here is available on a public repository but will be made available upon request to the corresponding author. The protocol for this report was not preregistered.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.