AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

2024, Vol. 50, No. 11, 1722–1739
https://doi.org/10.1037/xlm0001391

When Confidence Reveals More Than Recognition Performance Does: The Case of Context Load

Oliwia Zaborowska[1], Beatrice G. Kuhlmann[2], Katarzyna Zawadzka[3], and Maciej Hanczakowski[3]
[1] Faculty of Psychology in Warsaw, SWPS University
[2] Department of Psychology, School of Social Sciences, University of Mannheim
[3] Faculty of Psychology and Cognitive Science, Adam Mickiewicz University

Context in which events are embedded is often hypothesized to serve as an independent cue for retrieval. This means that any effects of context need to obey two basic principles of cue-dependent memory: Memory retrieval should be augmented when, first, encoding context is reinstated and, second, this context uniquely specifies individual items stored in memory. Both of these regularities are well supported for recall tests, but they remain contentious in recognition tests. Here, in three experiments, we assess whether unique and nonunique contexts affect memory processes when reinstated during recognition. However, rather than focusing on measures of recognition performance, we looked at confidence judgments collected during recognition that should be particularly sensitive to recollective effects resulting from context cuing. Experiments 1 and 2, using old/new and forced-choice recognition tests, respectively, documented positive effects of context reinstatement on confidence in correct recognition identifications, but only for contexts uniquely associated with individual items. These effects emerged even when there were no reliable context effects in recognition performance measures. Experiment 3 showed the same effect of context reinstatement, moderated by context load, when spontaneous recognition of a previous study episode occurred during restudy. These results demonstrate the role of context as an independent retrieval cue both in deliberate and spontaneous recognition.

Keywords: context, recognition, confidence, reinstatement effect, fan effect

Episodic memory is defined by access to contextual details of a remembered event—the where and when of what one remembers (Tulving, 1972). It is thus unavoidable that much effort in studies on episodic memory has been devoted to context effects. One of the main questions asked by memory researchers is not only whether context details accompany retrieval of any information from episodic memory, but also how context can be used to elicit retrieval of details associated earlier with this context (see Smith & Vela, 2001). In other words, what are the conditions under which trying to remember an event is easier in the presence of the encoding context for this event?

Retrieval benefits of context reinstatement have indisputably been observed in a variety of recall tests, both when participants are given no specific cues for a free-recall test (e.g., Godden & Baddeley, 1975; T. K. Isarida et al., 2017; Smith & Manzano, 2010) and when such cues are provided in case of a cued-recall test (Smith et al., 2014). The issue of context dependence becomes, however, more contentious when it comes to recognition memory, with a number of studies failing to observe any benefits of reinstating context at the time of a recognition test (e.g., Godden & Baddeley, 1980; Hockley, 2008; Hockley et al., 2012; Murnane & Phelps, 1993, 1994, 1995; Smith et al., 1978). In relation to these

This article was published Online First October 7, 2024.
Jonathan Tullis served as action editor.
Oliwia Zaborowska https://orcid.org/0000-0001-9272-5081
Beatrice G. Kuhlmann https://orcid.org/0000-0002-3235-5717
Katarzyna Zawadzka https://orcid.org/0000-0003-0214-1184
Maciej Hanczakowski https://orcid.org/0000-0001-8980-4918
All data and materials are available on the Open Science Framework at https://osf.io/rbdg2/. This work was supported by the joint National Science Centre/Deutsche Forschungsgemeinschaft grant awarded to Maciej Hanczakowski (Grant 2018/31/G/HS6/01839) and Beatrice G. Kuhlmann (Grant KU3329/3-1, Project Number 426600523) and by Grant PPN/PPO/2018/1/00103 from the National Agency for Academic Exchange awarded to Katarzyna Zawadzka. This research was funded in whole, or in part, by the National Science Centre, Poland [Grant 2018/31/G/HS6/01839]. For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Oliwia Zaborowska played a lead role in formal analysis, investigation, and resources, a supporting role in conceptualization, and an equal role in methodology, software, and writing–original draft. Beatrice G. Kuhlmann played a supporting role in conceptualization, methodology, and project administration and an equal role in funding acquisition and writing–review and editing. Katarzyna Zawadzka played a supporting role in funding acquisition and an equal role in project administration, supervision, conceptualization, methodology, and writing–review and editing. Maciej Hanczakowski played a lead role in conceptualization and an equal role in funding acquisition, methodology, software, project administration, supervision, and writing–original draft.
Correspondence concerning this article should be addressed to Maciej Hanczakowski, Faculty of Psychology and Cognitive Science, Adam Mickiewicz University, Szamarzewskiego 89/AB, 60-568 Poznań, Poland. Email: maciej.hanczakowski@gmail.com

null effects, it has been argued that context may play a negligible role in recognition where it is outshone by information embedded in the recognition probe itself (Smith, 1988). Still, there are also studies in which the benefits for recognition performance have been observed (e.g., Ensor et al., 2023; Gruppuso et al., 2007; Hanczakowski et al., 2014, 2015; Murnane et al., 1999; Russo et al., 1999). This variability in documented empirical patterns necessarily raises a question about the possible moderators of this effect, which in turn implies a theoretical question concerning the exact mechanism of the somewhat elusive effects of context reinstatement in recognition.

On the theoretical side, the cuing-with-context account (Macken, 2002; Rutherford, 2004; see also Koen et al., 2013) proposes that reinstated context serves as an independent retrieval cue, beside the recognition probe itself, which allows for recollecting an item–context association established at study. As with any instance of recollection, such retrieval may lead to an upshot of correct recognition compared to a condition in which context is not reinstated and recollection often fails. Regarding the moderators of the context reinstatement effect, the cuing-with-context account points to one of the major principles of memory—the cue-overload principle (Watkins & Watkins, 1975; see also Goh & Lu, 2012; Nairne, 2002; Poirier et al., 2012). If context is used as an independent cue for retrieving item–context associations, then it should be effective to the extent to which it excludes competing associations. In other words, contexts associated with fewer study items—what is referred to as low context load—should be more effective in supporting recognition performance than contexts associated with multiple study items, with increasing loads gradually reducing the effectiveness of cuing.

An alternative account of context effects has been proposed in the Item, Associated Context, and Ensemble (ICE) model developed by Murnane et al. (1999). This is an instance of a global matching model of recognition memory which assumes that recognition decisions result from a unitary memory signal, often referred to as familiarity, without any contribution of a recall-like process of recollection. In the ICE model, context present at encoding may become integrated with a to-be-remembered item, creating a novel memory representation, called an item–context *ensemble*. In the words of Murnane et al. (1999), "although constructed from item and context information, ensemble information is thought to be a unique type of information that is different from either item or context information considered alone" (p. 404). At retrieval, the re-presentation of a recognition probe together with its context results in a match to the stored ensemble, augmenting the familiarity signal and increasing the chances that this probe will be correctly recognized as old.

The uniqueness of ensembles proposed in the ICE model to account for context reinstatement effects means that "ensemble information only produces a match for targets tested in the learning context" (Murnane et al., 1999, p. 404)—that is, ensembles produce a specific surge in hits only for items tested with reinstated contexts that results in augmented recognition discriminability as compared to any condition in which context is changed across encoding and retrieval. To consider the possible modulating role of context load in the ICE model, one needs to decide whether familiarity elicited by ensembles incorporating unique context features is any different from familiarity elicited by ensembles incorporating shared context features. This issue has not been specifically considered by Murnane et al., but in the most basic formulation of the ICE model, ensembles are unified memory representations, independent from their constituent parts, and thus should not differ in their ability to elicit familiarity signal based on the distinctiveness of their parts. In other words, the ICE model, through its strong delineation between item, context, and ensemble representations, predicts benefits of context reinstatement to recognition performance for any value of context load, at least as long as item–context ensembles are successfully created at encoding (T. Isarida et al., 2018; Rutherford, 2004). This will be the prediction tested here. It is worth noting, however, that the ICE model could be in principle modified to muddle the distinction between item, context, and ensemble information, in which case distinctiveness of ensembles would itself depend on the distinctiveness of its constituent context features, accommodating the potential modulating role of context load—an issue we revisit in the General Discussion.

Given that the role of context load has long been deemed crucial for elucidating the theoretical underpinnings of the context reinstatement effect in recognition (see Ensor et al., 2023; T. Isarida et al., 2018; Rutherford, 2004), various studies have been conducted to test the competing predictions of the cuing-with-context and the ICE accounts. However, their interpretation is arguably not settled. First, some studies failed to observe the effects of context reinstatement on recognition performance even under conditions of low load (Hockley, 2008; Murnane & Phelps, 1993), thus rendering moot the issue of whether this effect would be reduced by increased load, and testifying only to the elusive nature of these effects in general. Second, some studies that did find support for the moderating role of context load—and thus argued for a crucial role of recollection in the reinstatement paradigms—used methods that, while increasing the chances of observing context reinstatement effects in recognition, used procedures that make it questionable whether the studies were really about context memory. For example, in a series of studies, Reder and her colleagues (Diana et al., 2004; Reder et al., 2002, 2013) found greater context reinstatement effects in recognition with lower context loads but achieved this by directly instructing participants to encode context information at study. This method has been criticized previously as changing the role of context in memory from peripheral to focal and thus changing the assessed research question (Hockley, 2008). Third, other studies which found support for the role of context load and used incidentally encoded context relied on either a cross-experimental comparison (T. Isarida et al., 2018) or an interpretation of nonsignificant effects (Rutherford, 2004). Fourth, and perhaps most importantly for the present study, the most recent investigation of this issue by Ensor et al. (2023) found a reversed effect of context load, with greater context reinstatement effect on recognition performance for highly loaded contexts than for low-load contexts—a result these authors interpreted as supporting the ICE theory over the cuing-with-context account. We reserve the discussion of this last study for later, and now we just highlight the issue of continuing controversy of whether and how context load moderates context effects in recognition.

As underscored throughout the present article, one of the main challenges of testing the competing theories of context reinstatement effects in recognition is that this effect is less than robust even under conditions of low context load, when both the cuing-with-context and the ICE accounts do predict the effect.[1] The challenge is thus to find a reliable index of the memory effect that context reinstatement is thought to yield. Here, we propose that to achieve this goal, research could move away from its focus on recognition performance measures and assess the impact of the context reinstatement manipulation on metamemory experience. Specifically, we propose to use confidence judgments to index the context reinstatement effect. In one of our previous investigations (Hanczakowski et al., 2014, Experiment 1), we examined simultaneously the effects of (low load) context reinstatement on forced-choice recognition of faces and retrospective confidence judgments accompanying those recognition decisions. We found a clear effect of the context manipulation on confidence, while there was no such effect on recognition performance. This suggests that metamemory may be more sensitive to context effects than the widely examined recognition performance measures, opening an avenue for pursuing a more detailed issue of the moderating role of context load. Before, however, we present our experiments, we need to tackle the theoretical problem of how it is even possible to reveal memory effects in metamemory that are not present in memory measures—an issue that for some reasons remains largely unrecognized in the memory literature.

To assess the effect of context reinstatement under conditions of equated memory for the context itself, a comparison is made between the reinstated context condition, with studied items presented in their associated encoding contexts, and a re-paired context condition, where contexts at test are taken from other previously studied items (Macken, 2002). The recognition performance measure in this case is simply the hit rate—with the strength of targets and contexts being equated, any difference in the hit rate across the reinstated and re-paired context conditions can only stem from some form of recognition that the studied item and its context go together, caused by a preserved relationship across items and their contexts in the reinstated context condition.[2]

Both the cuing-with-context and the ICE theories place the locus of the context reinstatement effect on hit rates, with the former arguing for recollection of the item–context association and the latter arguing for increased memory strength when both components are simultaneously present as a recognition probe. But equally, both accounts underscore that hit rates also depend on another factor—the familiarity (or memory strength) of the studied item itself. If an item is recognized as old based on item information alone, then additional recollection/strength coming from its relationship with context does not change the hit rate in any way. This means that only when the studied item is not familiar enough to be recognized may context information increase the chances of this item being correctly endorsed as studied. One thus needs to consider the likelihood that the memory of a particular item itself is weak, not supporting its correct recognition, while memory for either its association with the encoding context or the item–context ensemble is strong enough to help recognition. Due to a variety of factors—including fluctuations of attention at study that necessarily result in correlated signals for item and context information (deBettencourt et al., 2018) and faster forgetting rates for nonfocal and hence sparsely encoded context information (Benjamin, 2010)—this likelihood may be very low. In a majority of cases in which item memory fails, context memory is also likely to fail to support correct recognition. This contingency of context influence on item memory failure could be the prime reason for the persistent difficulties to observe context reinstatement effects in the memory measure of recognition performance.

What, then, about metamemory? The crucial point here is that the influence of context on confidence judgments is not contingent on not recognizing the item itself. Even when one has sufficient evidence that an item has been studied to warrant an "old" recognition decision, realizing that this item was presented in a particular encoding context can further increase one's confidence that the "old" recognition decision is indeed correct. In other words, metamemory judgments can provide a graded assessment of the strength or volume of memory information retrieved for any given correct recognition identification. This makes confidence a more sensitive measure of context effects, which can be potentially expressed in this metamemory measure across all recognition trials in the reinstated context condition. Of course, just as in the case of recognition hits, it is definitely not the case that confidence is solely determined by context memory. Multiple factors are likely to determine confidence, but clearly, the memory effect that results from context reinstatement is one of the bases of confidence, as previous research on context effects on confidence has revealed (Hanczakowski et al., 2014). Participants become more confident in their decisions in a memory task if they remember additional contextual information, which remains consistent with a general observation that metamemory processes are "parasitic" on memory processes (Koriat, 1993)—people know that they know only because they make inferences based on the characteristics and products of memory retrieval itself. And again, just as in the case of recognition hits and due to the way in which the context reinstatement paradigm is constructed, any difference in confidence across reinstated and re-paired context conditions can only possibly arise from a memory effect—some form of a realization that an item and its context go together—that translates into increased confidence that one correctly recognizes an item. In this formulation, there is nothing unique about metamemory judgments apart from them providing a fine-grained expression of changes in memory processes.

In the present study, we conducted three experiments assessing the role of context load on context reinstatement effects in recognition. We directly manipulated context load by either presenting study items with individual contexts, shown only once at study, or presenting multiple study items with a single context. Subsequently, these contexts were either reinstated for their study items or re-paired in such a way that old items were presented with old contexts which previously accompanied other study items. The main novelty of the present investigation was that our focus was not so much on recognition performance measures but on confidence judgments provided by participants. For the individual contexts, we expected robust effects of context reinstatement revealed in increased confidence accompanying correct recognition identifications, that is, decisions for which better memory for context

[1] One should note that there are likely to be variables other than context load that moderate the context reinstatement effect in recognition. For example, T. Isarida et al. (2020) obtained reasonably consistent effects in recognition using video rather than photograph backgrounds as contexts. A study by Shin et al. (2020) showed that context effects (assessed in recall) are larger for items which were deemed relevant to the virtual environment in which they were presented. The nature of the context is thus also important for the context effects (see also Bayen et al., 2000; T. Isarida et al., 2018).

[2] Note also that false alarm rates are uninformative in this paradigm. For new items, contexts cannot be considered reinstated or re-paired, as their status is realized only when they are paired with study items and preserve or break the item–context pairing created at study, respectively.

information—elicited by reinstated contexts—could contribute. The question was whether these effects on confidence would be reduced with an increased context load, as predicted by the cuing-with-context account, or would remain intact, as predicted by the ICE model. In Experiment 1, we used an old/new recognition test coupled with retrospective confidence judgments to address this issue. In Experiment 2, we switched to a forced-choice recognition test. In Experiment 3, we focused on spontaneous recognition during restudy, changing our metacognitive measure to judgments of learning (JOLs)—prospective confidence judgments (Zawadzka & Higham, 2015, 2016) concerning future memory performance. To foreshadow, all experiments revealed reduced context reinstatement effects in metacognitive measures when context load was increased, documenting the use of context as an independent cue in recognition and thus supporting the cuing-with-context account.

Experiment 1

The present experiment examined whether context load moderates the effects of context reinstatement in recognition with the use of a metacognitive measure of retrospective confidence judgments. Participants studied lists of single words that were superimposed over unrelated contextual background photographs. In subsequent old/new recognition tests, old items were presented against contexts taken from the encoding phase: either reinstated contexts that accompanied them at encoding, or re-paired contexts previously presented with other studied words. New words were always presented against old contexts. Context load was manipulated at study either by pairing words with individual, unique contexts for the *low context load condition,* or by pairing 24 different words with the same context background for the *high context load condition*. The main focus of the experiment was on the pattern of confidence judgments following correct identifications of old items. The context-cuing account predicts smaller differences in confidence between reinstated and re-paired contexts in the high than in the low context load condition, while the ICE account predicts similar differences in confidence in both low and high context load conditions.

One more manipulation was included in the design of Experiment 1. We suspected that the effectiveness of encoding context information—either as item–context associations or as part of item–context ensembles—may depend on the frequency of changes of context during study (see T. Isarida et al., 2020). In the high context load condition, numerous words are associated with a single context, which means that contexts may often remain unchanged for consecutive study items. This continuance of context presentation may lead to habituation of context—less effective encoding of context that would undermine its role during recognition as compared to individual contexts that are not repeated during study. If that were the case, it would skew the results toward the pattern predicted by the cuing-with-context account, while stemming from encoding dynamics rather than the effectiveness of cuing during retrieval. To assess whether such context habituation takes place, we directly manipulated the frequency of context changes during encoding by using a between-list manipulation of context load for one group of participants and a within-list manipulation for the other group. With the between-list manipulation, context changes in the high context load condition are less frequent than with the within-list manipulation, where the presentation of multiply associated contexts is often interspersed with presentations of singly associated contexts. Thus, if context habituation during encoding undermines encoding of contexts and their subsequent use during recognition, we would expect a reduction in the context reinstatement effect for the high context load condition to be more readily observed with the between-list manipulation of context load. To preview, no such effect was observed here, suggesting that context habituation is not crucial in determining the dynamics of context reinstatement effects in recognition.

Method

Participants

Ninety-six English-speaking undergraduate students (age: M = 20.48, range: 18–45, 82 females and 14 males) participated in exchange for course credit. This sample size was established based on the previous research on context effects on confidence (Hanczakowski et al., 2014, Experiment 1), where the context effect of interest resulted in a difference of $d_z = 0.49$. A power analysis showed that to obtain power of .90 to get the effect of this magnitude, 47 participants would need to be tested. Since we had two groups in the experimental design, we tested 48 participants in each group. The experiment was conducted in a laboratory setting.

Materials

A set of 384 English nouns of medium frequency consisting of four to eight letters was collated from the Medical Research Council database. These were divided into two subsets, one consisting of 288 words to be used as study words and the other of 96 words to be used as foils in recognition tests. Both subsets were then divided into six lists of 48 and 16 words, respectively. Participants in the between-list design studied six lists of 48 words separately, with each list followed by a recognition test for 32 of the words mixed with 16 foils. Only 32 of 48 studied words in each list were included in the recognition test because the remaining 16 words were used solely as the source of context photographs to be paired with foils in the recognition test for the low context load condition. For participants in the within-list design, two lists comprising the low and high context load conditions were merged, resulting in three study lists of 96 words, each followed by a recognition test for 64 of the studied words mixed with 32 foils. Again, only a subset of studied words was presented as targets in recognition tests, with the remaining words serving only as the source of context photographs paired with foils in the low context load condition.

A set of 150 photographs of landscapes and animals was collated from various Internet sources and the authors' holiday photograph album. All photographs were converted to black and white, with a resolution of 170 × 192. The photographs were randomly divided into three lists of 48 photographs and three lists of two photographs. The lists of 48 photographs were then used for the low context load condition, in which each photograph was paired randomly with one study word, while the lists of two photographs were used for the high context load condition, in which each photograph was paired randomly with 24 different study words. It is perhaps worth noting that thus the load conditions remain unbalanced with respect to context familiarity, as individual contexts are presented much more often in the high context load condition. This confound is an inevitable characteristic of studies on context load (see Ensor et al., 2023, for a discussion), although it is worth bearing in mind that the context reinstatement effect refers to a comparison of reinstated and re-paired conditions *within* each context load condition, thus keeping context familiarity constant.

For the recognition tests, in the between-list design, 32 words were randomly chosen from each study list to serve as old recognition probes, mixed with 16 foils, while in the within-list design, 64 words were randomly chosen to serve as old recognition probes, mixed with 32 foils. Half of the old items were presented with the same context that accompanied them at study for the reinstated context condition. The other half of old items were presented with a re-paired context, which for the low context load condition meant a switched context previously paired with another old item serving in the re-paired context condition, while for the high context load condition it meant the other context used in this condition for a particular list. The assignment of targets to four different context conditions, created by crossing the reinstatement and the load manipulations, was counterbalanced across participants. In the low context load condition, each new recognition probe was paired with one of the contexts previously paired with one of the 16 studied items that were not tested, while in the high context load condition, each new probe was paired with one of the two contexts used in this condition for a particular list.

Design

The study conformed to a 2 (recognition context: reinstated vs. re-paired) × 2 (context load: low vs. high) × 2 (context load group: within-list vs. between-list) mixed design, with the first two factors manipulated within participants and the last factor manipulated between participants.

Procedure

Participants were randomly assigned to the within-list or between-list group. They were instructed that they would study multiple lists of words, with each word superimposed over a picture background and each list followed by a recognition test. Participants were explicitly told to focus on their encoding of words rather than pictures. During study, each word was presented individually in capital letters and red font, superimposed over a picture background for 3 s, with a 500-ms interstimulus interval. A recognition test instruction immediately followed each study list. In recognition tests, participants were presented again with individual words superimposed over picture backgrounds and asked to decide whether these words were old or new, pressing “o” or “n,” respectively. Immediately after the recognition decision, they were asked to provide their judgment of confidence in the correctness of this decision on a 1 (*guessing*) to 6 (*sure*) scale. The time for making recognition and confidence decisions was not limited.

Results

All data and materials for Experiments 1–3 can be found at https://osf.io/rbdg2/.

Recognition

Means for hit and false alarm rates are presented in Table 1. Recognition performance was analyzed separately for hits and false alarms. With the use of the re-paired context condition, the effects of context reinstatement on recognition performance would be evident whenever hit rates for the reinstated context condition were higher than for the re-paired context condition. For completeness, we later present an analysis of recognition discrimination in the form of the area under the curve (AUC) index, derived from confidence judgments.

Hit rates were subjected to a 2 (recognition context: reinstated vs. re-paired) × 2 (context load: low vs. high) × 2 (context load group: within-list vs. between-list) mixed analysis of variance (ANOVA), which yielded a significant main effect of context load, $F(1, 94) = 10.97$, *mean square error (MSE)* = .005, $p = .001$, $\eta_p^2 = .11$, reflecting overall higher hit rates in the low context load condition, $M = 0.75$, $SD = 0.15$, than in the high context load condition, $M = 0.73$, $SD = 0.15$. More importantly, the main effect of recognition context was also significant, $F(1, 94) = 13.90$, $MSE = .004$, $p < .001$, $\eta_p^2 = .13$, reflecting overall higher hit rates in the reinstated context condition, $M = 0.75$, $SD = 0.15$, than in the re-paired context condition, $M = 0.73$, $SD = 0.14$. This confirms that a small but reliable effect of context reinstatement on recognition performance was found—only some form of recognition that items and reinstated contexts go together could produce higher hit rates when both the strength of targets and the strength of contexts were equated across the reinstated and re-paired context conditions. The main effect of group was also significant, $F(1, 94) = 6.83$, $MSE = .076$, $p = .01$, $\eta_p^2 = .07$, reflecting overall higher hit rates in the between-list design, $M = 0.77$, $SD = 0.14$, than the within-list design, $M = 0.70$, $SD = 0.14$. This effect likely arose due to differences in list length,

Table 1

Proportions of Old Responses as a Function of Context Load (High vs. Low), Type of Item (Old Item in a Reinstated Context, Old Item in a Re-Paired Context, New Item), and Context Load Group (Between-List vs. Within-List) in Experiments 1 and 2

Experiment and context load group	High context load			Low context load		
	Old reinstated	Old re-paired	New	Old reinstated	Old re-paired	New
Experiment 1						
Between-list	.77 (.16)	.74 (.15)	.17 (.15)	.81 (.14)	.77 (.14)	.20 (.17)
Within-list	.69 (.16)	.69 (.13)	.21 (.13)	.72 (.15)	.70 (.15)	.24 (.14)
Experiment 2	.80 (.13)	.80 (.12)		.81 (.11)	.80 (.13)	

Note. In Experiment 2, only the within-list manipulation of context load was implemented, with the use of a forced-choice recognition test meaning no separate false alarm rates. Standard deviations are given in parentheses.

with recognition being easier with shorter lists used in the between-list group.

The only significant interaction was the interaction between recognition context and group, $F(1, 94) = 4.69$, $MSE = .004$, $p = .033$, $\eta_p^2 = .05$, which arose because context reinstatement affected recognition performance—as measured by hit rates—in the between-list group, $t(47) = 4.63$, $p < .001$, $d = 0.67$, with higher performance in the reinstated, $M = 0.79$, $SD = 0.14$, than in the re-paired context condition, $M = 0.76$, $SD = 0.14$, but not in the within-list group, $t(47) = 1.04$, $p = .304$, $d = 0.15$, with comparable performance in the reinstated, $M = 0.71$, $SD = 0.15$, and the re-paired, $M = 0.70$, $SD = 0.13$, context conditions. Such an interaction was not expected, and thus, it is not discussed further, also because it was not obtained in the confidence or discriminability results presented later. Of more importance, the triple interaction was not significant, $F(1, 94) = .02$, $MSE = .003$, $p = .900$, $\eta_p^2 < .001$, $BF_{incl} = 0.21$. The interaction of recognition context and context load was also not significant, $F(1, 94) = 2.50$, $MSE = .003$, $p = .117$, $\eta_p^2 = .03$, but given our specific predictions concerning the differential effects of context reinstatement in the high and low context load condition, we nevertheless analyzed these effects separately, collapsing across the group variable. This revealed that context reinstatement reliably affected performance in the low context load condition, $t(95) = 4.51$, $p < .001$, $d = 0.46$, with better performance for the reinstated, $M = 0.76$, $SD = 0.16$, than for the re-paired context condition, $M = 0.73$, $SD = 0.15$, while the same effect was not reliable in the high context load condition, $t(95) = 1.43$, $p = .156$, $d = 0.15$, with comparable performance in the reinstated, $M = 0.73$, $SD = 0.16$, and re-paired, $M = 0.72$, $SD = 0.14$, context conditions. We also computed Bayes factors for these specific comparisons, finding extreme evidence for the context reinstatement effect for the low-load context condition, $BF_{10} = 897.26$, and moderate evidence against the effect for the high-load context condition, $BF_{10} = 0.30$. While these effects are in line with the predictions of the cuing-with-context account, they obviously need to be treated with caution given that the interaction of recognition context and context load was not significant.

False alarm rates were analyzed with a 2 (context load: low vs. high) × 2 (context load group: within-list vs. between-list) mixed ANOVA. Note once again that the analysis of false alarm rates does not include the variable of recognition context because context cannot be reinstated for novel items—they were always tested with old contexts that could be presented earlier either with a single studied item or with multiple studied items. The ANOVA yielded a significant main effect of load, $F(1, 94) = 6.90$, $MSE = .005$, $p = .010$, $\eta_p^2 = .07$, which arose because false alarms were more common in the low context load condition, $M = 0.22$, $SD = 0.16$, than in the high context load condition, $M = 0.19$, $SD = 0.14$. Neither the main effect of group, $F(1, 94) = 2.27$, $MSE = .038$, $p = .135$, $\eta_p^2 = .021$, nor the interaction, $F(1, 94) = 0.06$, $MSE = .005$, $p = .805$, $\eta_p^2 < .001$, was significant.

Recognition discriminability was analyzed by computing AUC indices. To derive those, confidence judgments were recoded so that rather than reflecting confidence in responses on a scale of 1–6, they would refer to confidence in the status of an item on a scale from 1 (*sure new*) to 12 (*sure old*). So, for example, if a given item was classified as old with confidence 6, this was translated into the judgment of 12; if it was classified as old with confidence 5, this was translated into the judgment of 11, and so on. These values of so-called Type-1 confidence were used to construct receiver operating characteristic (ROC) curves (which are presented in the Appendix) and to derive AUC scores across all experimental conditions.[3] The AUC scores, which are presented in Table 2, were analyzed with a 2 (recognition context: reinstated vs. re-paired) × 2 (context load: low vs. high) × 2 (context load group: within-list vs. between-list) mixed ANOVA. This yielded a significant main effect of recognition context, $F(1, 94) = 18.00$, $MSE = .001$, $p < .001$, $\eta_p^2 = .16$, because recognition performance was better in the reinstated ($M = 0.82$, $SD = 0.12$) than in the re-paired context condition ($M = 0.81$, $SD = 0.12$), and a significant main effect of context load group, $F(1, 94) = 5.72$, $MSE = .05$, $p = .019$, $\eta_p^2 = .06$, because recognition performance was better in the between-list design ($M = 0.84$, $SD = 0.11$)—with shorter study lists—than in the within-list design ($M = 0.79$, $SD = 0.11$). No other effect was significant, including the crucial recognition context by context load interaction, $F(1, 94) = 1.93$, $MSE = .001$, $p = .168$, $\eta_p^2 = .02$. Again, just as in the case of hit rates analyzed earlier, while this interaction was not significant, direct comparisons, collapsed across the group variable, showed that while the difference indicative of the context reinstatement effect was not significant in the high context load condition, $t(95) = 1.97$, $p = .052$, $d = 0.20$, with inconclusive evidence in either direction, $BF_{10} = 0.71$, it was significant in the low context load condition, $t(95) = 4.36$, $p < .001$, $d = 0.45$, with extreme evidence for the context reinstatement effect, $BF_{10} = 528.99$. Altogether, these results are consistent with the analysis of hit rates and show that context reinstatement did affect recognition performance, but also that it seemed to play a larger role in the low rather than high context load condition.

Confidence

All analyses presented here were performed only on confidence judgments made on trials with correct recognition decisions. This was done for two reasons. First, the bases of correct and incorrect recognition decisions are necessarily different, which means that the bases of confidence judgments related to those decisions are also different, precluding a sensible collapsed analysis. Second, and more importantly for the purpose of the present study, a realization that an item and a context go together cannot possibly occur when the item itself is identified as new, and for this reason, context reinstatement can only potentially affect confidence—via its effect on memory—in cases of correct identifications of targets. While it would be in principle possible to analyze confidence in incorrect identifications separately, such false alarms are relatively rare in recognition tests, leading to noisy confidence data, which we thus decided not to analyze.

The means for confidence judgments for correct recognition decisions (both hits and correct rejections) are presented in Table 3. A 2 (recognition context: reinstated vs. re-paired) × 2 (context load: low vs. high) × 2 (context load group: within-list vs. between-list) mixed ANOVA on confidence in hits yielded a main significant effect of recognition context, $F(1, 94) = 14.67$, $MSE = .039$, $p < .001$, $\eta_p^2 = .14$, with higher confidence for correct responses in the reinstated,

[3] Although we collected Type-2 confidence judgments in our study—that is, confidence related to one's decision—after recoding, we obtained Type-1 confidence, which is confidence related to the status of a memory probe, and thus, the constructed ROC curves are Type-1 ROC curves. The resulting indices need to be treated with caution as it is possible that they would be different if the approach of collecting Type-1 confidence judgments was adopted because Type-1 and Type-2 judgments may depend on different cues (but see Higham et al., 2009).

Table 2
Values of the Area Under the Curve Index as a Function of Context Load (High vs. Low), Recognition Context (Reinstated vs. Re-Paired), and Context Load Group (Within-List vs. Between-List) in Experiment 1

Context load group	High context load		Low context load	
	Reinstated	Re-paired	Reinstated	Re-paired
Within-list	.79 (.11)	.79 (.11)	.79 (.12)	.77 (.12)
Between-list	.84 (.12)	.83 (.12)	.86 (.12)	.83 (.12)

Note. Standard deviations are given in parentheses.

$M = 5.11$, $SD = 0.64$, than in the re-paired context condition, $M = 5.03$, $SD = 0.69$. This again reflects the context reinstatement effect also found in the recognition measures. The only other significant effect was an interaction of recognition context and context load, $F(1, 94) = 4.60$, $MSE = .04$, $p = .035$, $\eta_p^2 = .05$. This arose because confidence in hits in the low context load condition was higher for reinstated contexts, $M = 5.16$, $SD = 0.64$, than for re-paired contexts, $M = 5.04$, $SD = 0.72$, $t(95) = 4.21$, $p < .001$, $d = 0.43$, while the same difference was not reliable in the high context load condition, $t(95) = 1.17$, $p = .247$, $d = 0.12$, with comparable confidence for reinstated, $M = 5.06$, $SD = 0.70$, and re-paired contexts, $M = 5.03$, $SD = 0.72$. These results were further confirmed by Bayesian analyses, which yielded extreme evidence in favor of the context reinstatement effect in the low context load condition, $BF_{10} = 304.95$, but also evidence against such an effect in the high context load condition, $BF_{10} = 0.22$. These results are in line with predictions of the cuing-with-context account, and—due to a reliable interaction—they are less ambiguous than the accuracy data in revealing a reduction of the context reinstatement with increased context load. No other effect was significant, $F(1, 94) = 2.77$, $MSE = .10$, $p = .099$, $\eta_p^2 = .03$, for the main effect of context load; $F(1, 94) = 2.52$, $MSE = 1.72$, $p = .116$, $\eta_p^2 = .03$, for the main effect of group; $F(1, 94) = 3.40$, $MSE = 0.10$, $p = .068$, $\eta_p^2 = .04$, for the context load by group interaction; $F(1, 94) = 1.66$, $MSE = 0.04$, $p = .201$, $\eta_p^2 = .02$, for the recognition context by group interaction; and $F(1, 94) = 0.02$, $MSE = 0.04$, $p = .883$, $\eta_p^2 < .001$, $BF_{incl} = 0.27$, for the triple interaction.

To provide further insight into the nature of confidence judgments, Figure 1 presents the proportions of hits made at each level of confidence across recognition context and context load conditions, but collapsed across the context load group variable that did not affect confidence in previous analyses. From these, it can be seen that while for the high context load condition there is little difference across recognition context conditions at any level of confidence, in the low context load condition, there seems to be a shift toward hits made with the highest confidence in the reinstated context compared to the re-paired context condition. This pattern was confirmed with paired comparisons. They showed that while in the high context load condition the proportion of hits made with the highest confidence did not differ across the reinstated and re-paired context conditions, $t < 1$, in the low context load condition, there were indeed more hits made with the highest confidence in the reinstated than in the re-paired context condition, $t(95) = 3.28$, $p = .001$, $d = 0.33$. Overall, this suggests that the effects of context reinstatement emerged in confidence judgments made in the low context load condition because participants were more likely to make highest confidence identifications of old items when they encountered them in the presence of their original rather than changed contexts.

For completeness, we also analyzed the effects of retrieval context on confidence in correct rejections. Once again, this analysis concerns new items so the reinstatement manipulation is absent from here. A 2 (context load: low vs. high) × 2 (context load group: within-list vs. between-list) mixed ANOVA yielded only a significant interaction, $F(1, 94) = 4.28$, $MSE = 0.07$, $p = .041$, $\eta_p^2 = .04$, although the difference between contexts from the high and low context conditions was not reliable in either the within-list group, $t(47) = 1.80$, $p = .078$, $d = 0.26$, or the between-list group, $t(47) = 1.01$, $p = .315$, $d = 0.15$. Neither the main effect of context load, $F(1, 94) = 1.22$, $MSE = 0.07$, $p = .272$, $\eta_p^2 = .01$, nor the main effect of context load group, $F(1, 94) = 1.20$, $MSE = 1.92$, $p = .275$, $\eta_p^2 = .01$, was significant.

Discussion

The results of the present experiment are in line with the predictions of the cuing-with-context account of context reinstatement effects in recognition. They demonstrate that while context reinstatement does affect recognition processes—an effect that is often difficult to obtain when only recognition performance is considered (Hockley, 2008; Murnane & Phelps, 1993, 1994, 1995; Smith et al., 1978)—this effect is larger when contexts are individually paired with single study items than when contexts are shared across multiple study items. Such a role of context load is consistent with context being used as an independent cue in recognition tests (Macken, 2002; Rutherford, 2004), revealing the role of the cue-overload principle (Watkins & Watkins, 1975):

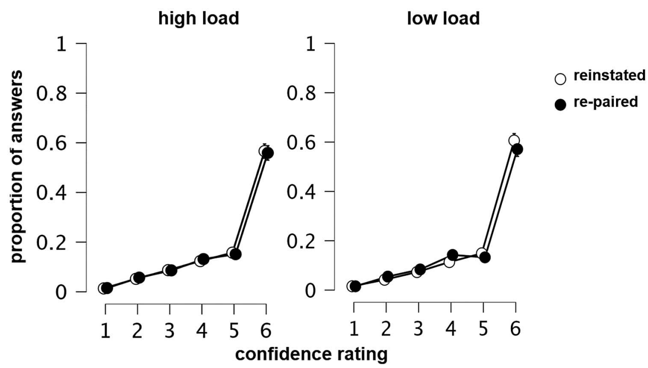
Table 3
Confidence in Correct Responses as a Function of Context Load (High vs. Low), Type of Item (Old Item in a Reinstated Context, Old Item in a Re-Paired Context, New Item), and Context Load Group (Between-List vs. Within-List) in Experiments 1 and 2

Experiment and Context load group	High context load			Low context load		
	Old reinstated	Old re-paired	New	Old reinstated	Old re-paired	New
Experiment 1						
Between-list	5.12 (0.69)	5.11 (0.73)	4.02 (1.08)	5.28 (0.63)	5.19 (0.69)	4.14 (0.94)
Within-list	4.99 (0.70)	4.94 (0.70)	3.88 (0.98)	5.03 (0.64)	4.88 (0.73)	3.85 (0.99)
Experiment 2	4.48 (0.93)	4.46 (0.95)		4.66 (0.081)	4.51 (0.93)	

Note. In Experiment 2, only the within-list manipulation of context load was implemented, with the use of forced-choice recognition test meaning no separate false alarm rates. Standard deviations are given in parentheses.

Figure 1

Proportions of Answers at Each Confidence Rating Level Across Recognition Context and Context Load Conditions in Experiment 1, Collapsed Across Groups



Note. Error bars denote the standard errors.

Contexts associated with fewer study items, more uniquely pointing to information that needs to be retrieved from memory, are more effective as retrieval cues. By the same token, the present results contradict the predictions of the ICE model (Murnane et al., 1999), which argues that context reinstatement effects arise due to compound cuing with the use of context features in conjunction with item features, with the resulting compound cue matching item–context ensembles stored in memory. The ICE theory, with its denial of the role independent context cues, has no role for context load and thus would not predict the current pattern of results.

It is important to stress that the support for the role of context load observed in the present experiment comes primarily from the confidence rather than recognition performance data. Although an overall effect of context reinstatement was observed on hit rates, and it seemed to derive from the effect observed in the low context load condition, the interaction of recognition context and context load was not reliable when hit rates were analyzed. This is in line with a common pattern of results in the context reinstatement literature, where reinstatement effects are not always picked up by the recognition measures employed. By contrast, this interaction was reliable in the confidence results. We do not wish to make strong conclusions solely on the basis of the lack of a significant interaction in one measure and a significant interaction in another—especially as the conclusions from both the frequentist and Bayesian t test results were similar regardless of whether we looked at recognition or confidence data. What we want to argue here, however, is that using a measure of confidence in addition to a measure of recognition performance can prevent accidentally overlooking the influence of context reinstatement on participants' recognition processes.

Additional analyses suggest that context reinstatement in the low context load condition specifically boosted old identifications assigned the highest confidence ratings. These results are also consistent with two previous studies concerning context reinstatement in recognition. Macken (2002) showed that context reinstatement affects mainly recognition decisions accompanied by remember judgments, which can also be conceptualized as high confidence endorsements (Donaldson, 1996), and Koen et al. (2013) demonstrated that context reinstatement affects the recollection component of the dual-process model, which is derived based on high confidence endorsements (see Yonelinas, 1994). Overall, these results chime with our initial intuition that confidence, which is not contingent on failures to recognize items embedded in retrieval contexts, may be a more sensitive measure of context effects on recognition processes than recognition performance measures (Hanczakowski et al., 2014). However, with both accuracy and confidence measures largely pointing in the same direction in this experiment, such a suggestion still remains tentative.

A clear dissociation of accuracy and confidence in recognition was observed previously with a forced-choice testing format (Hanczakowski et al., 2014) rather than the old/new format used in Experiment 1. Recently, it has been argued that forced-choice testing should be the preferred way of testing various effects in recognition memory due to a straightforward way of interpreting accuracy results, without the need to consider how to disentangle discriminability—the memory component of recognition—from bias, which is the decisional component involved in translating memory evidence into explicit responses in a test (Brady et al., 2023). Context reinstatement effects are not subjected to the conundrum of separating discriminability and bias as long as the re-paired context condition is included in the experimental design—a point underscored here by the consistency of Experiment 1 results, where the analyses of hit rates produced exactly the same results as the analyses of the AUC measure of discriminability derived from confidence judgments. As argued earlier, in this design, only recognizing the actual pairing of an item and its encoding context could feed into the decisional process, meaning that any effect observed in recognition hits must necessarily be memory based. Still, using the forced-choice format should also reveal the effects of context reinstatement on recognition memory. If the forced-choice format is the answer to the problems with interpreting recognition data, then we would expect context effects to be clearly discerned in this type of test. However, if previous investigations of context effects in recognition are taken into account, it is possible that this format of testing will reveal a dissociation between measures of accuracy and confidence.

Experiment 2

The present experiment sought to confirm the conclusions derived from Experiment 1. We again assessed the role of context load in moderating context reinstatement effects in recognition, testing whether these effects are reduced when context load is increased, as the cuing-with-context account would predict, or remain unchanged by increased context load, as predicted by the ICE theory. The change introduced here compared to Experiment 1 was that we used a forced-choice testing format. From the perspective of theories of recognition, this should make little difference as there is no reason to hypothesize that recognition processes are altered by the format of testing (Jang et al., 2009). Thus, while it could be argued that both the cuing-with-context and the ICE theories predict the context reinstatement effect to be revealed in hit rates when context load is low, we argue that such an effect may be difficult to obtain due to item memory obscuring the effect of context reinstatement. By contrast, from the metacognitive perspective advocated here, forced-choice testing could further serve to demonstrate a dissociation across recognition accuracy and confidence results in recognition, with robust effects of context reinstatement on confidence in correct recognition identifications. Based on a previous study using this format of testing (Hanczakowski et al., 2014), we would expect any effects of context reinstatement to be clearly visible in the measure of confidence.

Method

Participants

Seventy-one English-speaking undergraduate students (age: M = 19.44, range: 18–46, 59 females and 12 males) participated in exchange for course credit. The sample size was determined by the availability of participants in the subject pool in the semester in which the study was conducted. The power to detect the interaction of interest (for confidence judgments) of a magnitude observed in Experiment 1 ($\eta_p^2 = .05$) was .91. The experiment was conducted in a laboratory setting.

Materials, Procedure, and Design

The materials were the same as in Experiment 1, bar the addition of a new set of 192 words that were used as foils in the recognition test: While in Experiment 1 there were more targets than foils at test, and also not all targets were tested because a number of contexts from the low context load condition had to be reserved for use with foils, here the use of a forced-choice test necessitated the use of the same number of targets and foils and allowed for testing all targets. The procedure was based on that from Experiment 1, although this time context load was only manipulated within lists. The main change was the type of a recognition test. Here, two words were displayed superimposed over a picture background, one above the other. The tests required participants to identify old words by pressing either "z" for the word at the top of the picture or "m" for the word at the bottom. Old words were displayed equally often at the top and bottom of the pictures. After making the recognition decision, participants had to rate their confidence in it on a scale from 1 (*guess*) to 6 (*sure*). The time for the recognition judgment and for the following confidence judgment was again not limited. The design of the experiment was 2 (recognition context: reinstated vs. re-paired) × 2 (context load: high vs. low), with both factors manipulated within participants. The assignment of target words to four experimental conditions was counterbalanced across participants.

Results

Recognition

Given the use of a forced-choice recognition test, only hit rates were now available for analysis. These are presented in Table 1. A 2 (recognition context: reinstated vs. re-paired) × 2 (context load: high vs. low) within-participants ANOVA yielded no significant effects, $F(1, 70) = 0.25$, $MSE = .01$, $p = .619$, $\eta_p^2 = .004$, for the main effect of context load; $F(1, 70) = 1.69$, $MSE = .002$, $p = .198$, $\eta_p^2 = .02$, for the main effect of recognition context; and $F(1, 70) = 0.70$, $MSE = .002$, $p = .407$, $\eta_p^2 = .01$, for the interaction. When we assessed evidence for the context reinstatement effect separately for the high- and low-load context conditions, in both cases, evidence favored no effect, $BF_{10} = 0.14$ and $BF_{10} = 0.40$, respectively, although evidence in the latter case was only anecdotal.

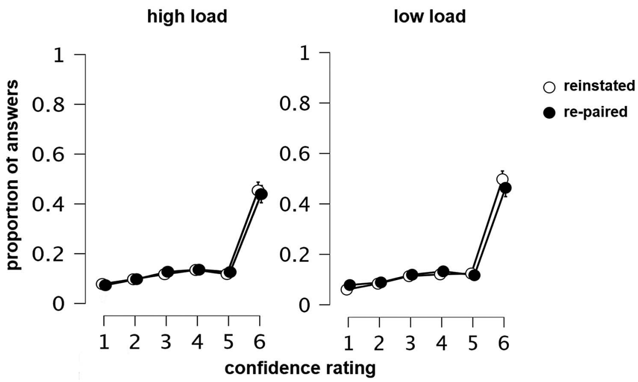
Confidence

The means for confidence judgments for correct recognition decisions are presented in Table 3. A 2 (recognition context: reinstated vs. re-paired) × 2 (context load: high vs. low) within-participants ANOVA yielded a significant main effect of context load, $F(1, 70) = 8.87$, $MSE = 0.10$, $p = .004$, $\eta_p^2 = .11$, which arose because participants were generally more confident in their correct responses in the low context load condition, $M = 4.58$, $SD = 0.85$, than in the high context load condition, $M = 4.47$, $SD = 0.93$. The main effect of recognition context was also significant, $F(1, 70) = 7.96$, $MSE = 0.06$, $p = .006$, $\eta_p^2 = .10$, and this reflected higher confidence in correct responses made in the presence of reinstated contexts, $M = 4.57$, $SD = 0.85$, than in the presence of re-paired contexts, $M = 4.49$, $SD = 0.93$. Thus, in contrast to the recognition performance results, confidence results revealed a clear context reinstatement effect. Finally, and most importantly, the interaction was significant, $F(1, 70) = 7.86$, $MSE = 0.04$, $p = .007$, $\eta_p^2 = .10$, which reflected the fact that while the context reinstatement effect for confidence was robust in the low context load condition, $t(70) = 3.48$, $p < .001$, $d = 0.41$, the same difference was not significant in the high context load condition, $t(70) = 0.57$, $p = .572$, $d = 0.07$. The Bayesian analyses confirmed strong evidence for the context reinstatement effect when load was low, $BF_{10} = 28.82$, but also strong evidence against such an effect when load was high, $BF_{10} = 0.15$.

To gain further insight into the nature of confidence judgments, in Figure 2, we present the proportions of hits made at each level of confidence across recognition context and context load conditions. These plots suggest that, as in Experiment 1, in the high context load condition there was little difference across recognition context conditions at any level of confidence, including the highest level, for which the difference across reinstated and re-paired context conditions was not significant, $t(70) = 1.65$, $p = .102$, $d = 0.20$. At the same time, in the low context load condition, there was a shift toward hits made with the highest confidence in the reinstated compared to the re-paired context condition, confirmed by a significant difference in the proportion of the ratings of 6 when

Figure 2

Proportions of Answers at Each Confidence Rating Level Across Recognition Context and Context Load Conditions in Experiment 2



Note. Error bars denote the standard errors.

contexts were reinstated rather than re-paired, $t(70) = 3.21$, $p = .002$, $d = 0.38$. Overall, these results again suggest that the effects of context reinstatement emerged in confidence judgments when context load was low because participants were more likely to recognize old items with the highest confidence when they encountered them in the presence of their original contexts.

Discussion

The results of Experiment 2 confirm the conclusions derived from Experiment 1. First, we found again that context reinstatement affected recognition processes. Second, this effect was modulated by context load, as the effect of context reinstatement was present when contexts were individually paired with study items, but absent when contexts were associated with multiple study items. Third, the context reinstatement effect for the low context load condition was revealed in the measure of confidence in correct responses and caused in particular by high confidence endorsements of targets, while at the same time it was absent from the measure of recognition performance. The first important theoretical conclusion, then, is that the present results unambiguously demonstrate the role of context as an independent retrieval cue in recognition. Despite the change of test format, context was again subjected to the cue-overload principle by which overloaded contexts, pointing to multiple study items, are less effective as retrieval cues, in line with the cuing-with-context account of context reinstatement effects in recognition.

The second important point demonstrated by the present results is again that not all changes to recognition processes are reflected in measures of recognition performance. Although the performance–confidence dissociation was not particularly convincing in the old/new test used in Experiment 1, where the robust context reinstatement patterns found in confidence were also reflected—albeit far more tentatively—in the hit rate and AUC scores, here this dissociation was clear. Context reinstatement failed to affect the recognition performance measure of hit rates in any detectable way. Why, then, could context constitute such an ineffectual cue for supporting recognition performance in the present experiment, while small effects on recognition performance were observed in Experiment 1? One option is that the perceptual match between encoding and retrieval was not as perfect here as in the previous experiment because single words were superimposed over context pictures at study and two words (the target and a lure) were presented at test.[4] With the overall test display not perfectly matching encoding even in the reinstated context condition, the memory effects of context reinstatement could have been reduced, leaving accuracy unaffected. But such recognition effects were clearly not eliminated, as testified to by the confidence results.

Several previous investigations of context reinstatement in recognition obtained null results in an accuracy measure (e.g., Hockley, 2008; T. Isarida et al., 2018; Russo et al., 1999), just as the present experiment. Often such null results were used to conclude that context reinstatement does not affect recognition under some specific experimental conditions. However, the present results indicate that null results in terms of performance can well be accompanied by robust effects on confidence. These confidence patterns indicate that context reinstatement did affect recognition processes—participants retrieved more information from their memory when item-specific contexts were reinstated—even if this information failed to contribute to the accuracy of their overt recognition responses. We hold this as one of the major conclusions of the present work—null results in accuracy measures should not be interpreted as indicating that memory processes were unaffected by a particular experimental manipulation. If the purpose of researchers is to investigate psychological processes rather than performance measures, then multiple measures of these processes should be utilized, and metacognitive measures could be employed to describe the workings of memory.

The discussed dissociation across measures of recognition performance and metacognitive appraisal of this performance opens another potential avenue for investigating context effects in recognition. Recognition is often elicited not by an explicit requirement of a particular memory test but by an incidental encounter with a familiar stimulus which is spontaneously recognized as previously experienced (Hintzman, 2011). Such spontaneous recognition cannot be investigated with explicit memory tests, as asking for explicit judgments of recognition would inevitably change the spontaneous nature of the process, but it can be pursued with the use of metacognitive judgments (e.g., Tullis et al., 2014). In the following experiment, we thus extended our investigation of the role of context load to a case of spontaneous recognition during study, as assessed by metacognitive appraisals of the effectiveness of encoding operations.

Experiment 3

The primary aim of this third experiment was to once again examine a possible modulating role of context load for the effects of context reinstatement on metamemory. While Experiments 1 and 2 provided a demonstration of such a role in explicit recognition tests, here we focused on spontaneous changes in memory processes that may occur while studying. Participants were presented with pairs of words to study in two separate study phases. These pairs were presented either with individual contexts or with contexts paired with multiple study pairs. The restudy phase served a similar role as tests in Experiments 1 and 2, so that at restudy contexts were either reinstated for a particular pair from the first study phase or re-paired. Immediately after completing a restudy trial, participants were asked

[4] We thank William Hockley for this suggestion.

to provide a JOL for this particular pair—a judgment of prospective confidence regarding remembering the second word from this particular pair when later cued with the first word at test. The impact of the context reinstatement manipulation on the magnitude of JOLs was examined.

What might happen when context is reinstated at restudy? Two previous studies using this methodology and employing only a low context load condition (Saenz & Smith, 2018; Zawadzka et al., 2018) have shown that context reinstatement at restudy affects JOLs, which are higher when contexts are reinstated rather than re-paired. This indicates that some memory process unfolds differently under context reinstatement and this change in memory feeds into metacognitive assessments, which become shaped by this mnemonic cue (Koriat, 1997). But the crucial question concerns the nature of this memory process modified by context reinstatement. Tullis et al. (2014), who observed the effects of spontaneous reminding on JOLs, noted that this effect could take two forms: either that of conscious recollection or a more implicit form of recognition characterized by increased fluency of item processing. These two forms can be thought of as mapping onto the two different mechanisms of context reinstatement effects pursued here, with conscious recollection of item–context associations stemming from context reinstatement according to the cuing-with-context account, and increased fluency being synonymous with familiarity elicited by ensemble matching.

If it is assumed that both of these memory effects translate into increased JOLs, then both the cuing-with-context and the ICE theory can account for previous observations that context reinstatement at restudy shapes JOLs. However, this also means that these accounts formulate different predictions for the role of context load. The cuing-with-context account predicts item–context recollection to be dampened—and hence the JOL effect reduced—under conditions of increased context load, and the ICE theory, assigning no role to context load, predicts similar effects of context reinstatement for JOLs in both context load conditions.

Method

Participants

Fifty-four Polish-speaking undergraduate students (age: M = 26.91, range: 18–45, 48 females and six males) participated in Experiment 3 in exchange for course credit. Given that the current procedure differed from the procedure of Experiments 1 and 2, we consulted the effect size of the context reinstatement effect for JOLs from the study by Zawadzka et al. (2018, Experiment 1). The power to detect such an effect of $d = 0.56$ in the present experiment was .98. The experiment was conducted online.

Materials

A set of 240 Polish nouns of medium-to-high frequency consisting of four to eight letters was chosen from the Subtlex-pl database (Mandera et al., 2015). These words were randomly paired to create 120 cue–target pairs and then divided into two lists of 60 pairs each. These lists were assigned to either the low context load condition or the high context load condition, with the assignment of list to conditions counterbalanced across participants. Thus, participants completed two experimental blocks, each consisting of two study phases and a cued-recall test, with one block serving in the low context load condition and the other serving in the high context load condition. The order of the blocks was random.

The 62 pictures used for this study were taken from the set used in Experiments 1 and 2. Of these, 60 were used for the list of word pairs assigned to the low context load condition, and the remaining two were used for the list assigned to the high context load condition. For the low context load condition, each context was thus paired with an individual study pair, while for the high context load condition, each context was paired with 30 different study pairs. In the second study phases of both blocks, contexts from the first study phase were reinstated for half of the study pairs. For the remaining half, contexts were re-paired, which for the low context load condition meant taking context from a yoked study pair, while for the high context load condition meant using the other context used in this condition. The assignment of pairs to experimental conditions was counterbalanced across participants.

Design

The design of the experiment was 2 (restudy context: reinstated vs. re-paired) × 2 (context load: low vs. high), with both variables manipulated within participants. Recognition context was manipulated within study lists, while context load was manipulated between lists used in two separate experimental blocks.

Procedure

Participants were first presented with 60 word pairs for study, superimposed over context backgrounds. They were instructed to concentrate on memorizing these pairs, which would later be tested. Each pair was presented for 4 s. Immediately after the first study phase, a restudy phase followed, where the same pairs were presented in a new random order, superimposed over either the same contexts as before or over contexts taken from other studied items. Here, each word pair was presented for 2.5 s, and this was followed by a JOL prompt, asking participants to rate on a 1–5 scale their confidence in recalling the target word at test when cued with the first word from the pair. The time for providing JOLs was not limited. After the completion of the second study phase, participants were given a cued-recall test in which first words from the studied pairs were used as cues for eliciting their respective second, target words. The time to provide responses was 10 s, and no contexts were shown in this phase of the experiment. The cued-recall test for the first block was immediately followed by the second experimental block, after which the procedure terminated.

Results

JOLs

Mean JOLs from the restudy phase are presented in Table 4. JOL magnitude was analyzed with a 2 (restudy context: reinstated vs. re-paired) × 2 (context load: high vs. low) within-participants ANOVA. This yielded a significant main effect of restudy context, $F(1, 53) = 4.26$, $MSE = 0.09$, $p = .044$, $\eta_p^2 = .07$, which reflected higher JOLs given when contexts at restudy were reinstated, M = 2.31, SD = 0.73, rather than re-paired, M = 2.23, SD = 0.68, documenting the context reinstatement effect in spontaneous recognition during restudy. The main effect of context load was

Table 4

Proportions of Correctly Recalled Items in the Final Cued-Recall Test and Mean Judgments of Learning (on a 1–5 Scale) at Restudy as a Function of Context Load (High vs. Low) and Type of Context (Reinstated vs. Re-Paired) Provided at Restudy

	High context load		Low context load	
Measure	Reinstated	Re-paired	Reinstated	Re-paired
Cued recall	0.30 (0.21)	0.29 (0.20)	0.27 (0.21)	0.27 (0.19)
JOLs	2.27 (0.79)	2.27 (0.75)	2.35 (0.81)	2.19 (0.71)

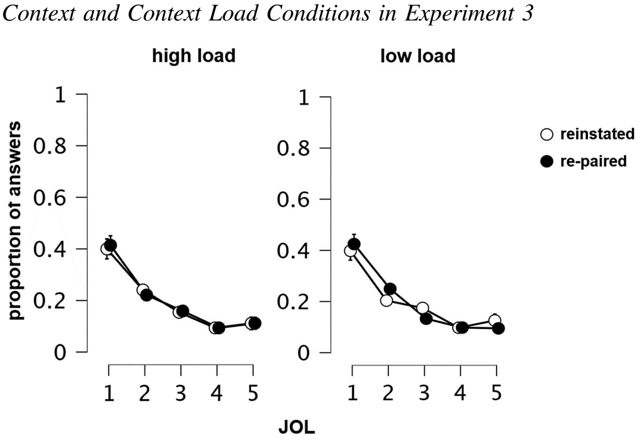
Note. Standard deviations are given in parentheses. JOLs = judgments of learning.

not significant, $F(1, 53) < .01$, $MSE = 0.31$, $p = .983$, $\eta_p^2 < .001$, but the Restudy Context × Context Load interaction was, $F(1, 53) = 8.72$, $MSE = 0.04$, $p = .005$, $\eta_p^2 = .14$. This interaction reflected the fact that while the context reinstatement effect emerged in JOLs provided in the low context load condition, $t(53) = 3.11$, $p = .003$, $d = 0.42$, it was absent from the high context load condition, $t(53) = 0.15$, $p = .884$, $d = 0.02$. Bayesian analyses revealed strong evidence for the context reinstatement effect when load was low, $BF_{10} = 10$, but also strong evidence against such an effect when load was high, $BF_{10} = 0.15$.

In Figure 3, we present the proportions of JOLs made at each level of confidence across restudy context and context load conditions. These plots suggest that for the high context load condition, there was very little difference across restudy context conditions at any level of confidence, including the highest level of 5, for which the difference across reinstated and re-paired context conditions was not significant, $t < 1$. At the same time, in the low context load condition, there seemed to be two separate context-dependent shifts. First, participants seemed to make more JOLs in the reinstated context compared to the re-paired context condition at the level of 3 (the exact midpoint of the confidence scale), as confirmed by a significant difference, $t(53) = 3.07$, $p = .003$, $d = 0.42$, as well as at the level of 5 (highest confidence), $t(53) = 2.37$, $p = .021$, $d = 0.32$. This latter shift is consistent with recognition results obtained in

Figure 3

Proportions of Answers at Each JOL Rating Level Across Study Context and Context Load Conditions in Experiment 3



Note. Error bars denote the standard errors. JOL = judgment of learning.

Experiments 1 and 2, and it indicates that reinstated contexts at restudy increase the likelihood of highest confidence responding, even when such responding concerns future rather than past performance. However, the shift toward the value in the middle of the confidence scale in the reinstated context condition was not observed for retrospective confidence.[5]

Cued Recall

For completeness, we also analyzed performance in the final cued-recall test as a function of context present at restudy. A 2 (restudy context) × 2 (context load) within-participants ANOVA yielded no significant effects, $F(1, 53) = 3.54$, $MSE = .002$, $p = .065$, $\eta_p^2 = .06$, for the main effect of context load; $F(1, 53) = 0.27$, $MSE = .007$, $p = .608$, $\eta_p^2 = .005$, for the main effect of restudy context; and $F(1, 53) = .01$, $MSE = .005$, $p = .910$, $\eta_p^2 < .001$, for the interaction.

Discussion

The present results confirm and extend the results of Experiments 1 and 2. Once again, the analysis of confidence judgments yielded a context reinstatement effect that was nevertheless limited to the low context load condition. In other words, reinstating encoding contexts at restudy affected recognition processes, but only when contexts were individually paired with the to-be-remembered items. Thus, context load once again moderated the effect of context reinstatement, providing evidence that context serves as an independent retrieval cue. These results are consistent with the cuing-with-context account of the context reinstatement effect: Reinstating unique contexts at restudy led participants to recollect item–context associations, and this conscious recollection fed into their predictions of future memory performance.

It is important to note that while Experiments 1 and 2 showed the role of context reinstatement in explicit recognition tests, here context was used to cue memory at restudy, when no explicit requirement to retrieve information from memory was imposed on participants. The influence of context reinstatement on recognition of previously studied materials is thus at least partially independent of a specific requirement to remember. At the same time, it is worth bearing in mind that participants in the present experiment faced a memory task, so it cannot be ruled out that participants did try to deliberately engage retrieval processes in support of effective learning. The goal of the present experiment also necessitated the inclusion of JOLs as our measure of spontaneous recognition, and it is now well-known that eliciting metacognitive judgments might change the very cognitive processes that these judgments are supposed to measure (e.g., Mitchum et al., 2016), making participants more focused on any type of information that may support

[5] It is worth noting that JOLs are different from retrospective confidence inasmuch as while recollection is clearly diagnostic of previous study, leading participants to give highest retrospective ratings when it occurs, it is not fully diagnostic of future memory performance. Just because someone recollects an item–context association does not mean that they will necessarily remember the item in the future. The two peaks in JOLs that are caused by increased recollection of item–context associations when context is reinstated could thus reflect two separate shifts in prospective confidence—from "I will not remember this item in the future" to "I may remember this item in the future" and from "I may remember this item in the future" to "I will remember this item in the future."

subsequent memory performance (Halamish & Undorf, 2023; Soderstrom et al., 2015). Thus, it remains possible that context serves as a cue mostly when participants remain in an explicit learning mode, having their attention directed toward contextual factors by a requirement to monitor their own learning.

Finally, the effect of context reinstatement was observed here in a measure of confidence different from retrospective confidence examined in Experiments 1 and 2. This time, the pattern of changes to recognition processes was revealed in confidence relating to future memory performance (Zawadzka & Higham, 2015, 2016), underscoring that a wide array of metacognitive measures are sensitive to context effects (see also Hanczakowski et al., 2017). This supports our contention that metacognitive judgments can be used more generally to probe memory processes, particularly when examining memory performance is either impossible or undesired due to methodological considerations such as the need to preserve the spontaneous nature of the examined memory process.

General Discussion

The present study assessed whether context features present at retrieval serve as an independent or an integrated cue in the process of recognition. Our test of choice for these accounts was whether retrieval context is subjected to the cue-overload principle (Watkins & Watkins, 1975), with potentially diminishing effects of contexts associated with multiple rather than single study items. While the use of context as an independent cue predicts such an effect of context load, its use as an integrated cue together with item features does not. Our chosen method was an investigation of context effects on metacognitive measures of confidence. This approach yielded two main conclusions. The first, a narrower one, is that context is subjected to the cue-overload principle, which suggests it is used as an independent cue for recognition. The second, and more general, is that confidence can be used to elucidate memory effects of context manipulations, sometimes more effectively than measures of recognition performance. We discuss these two issues in turn.

Regarding the cue-overload principle, our predictions were formulated on the basis of the cuing-with-context account (Macken, 2002; Rutherford, 2004) which argues that context is used as an independent cue in recognition, facilitating recollection of item–context associations created at study, as well as the ICE account which argues that context is only used as a cue in conjunction with features of a recognition probe, facilitating access to item–context ensembles created at study. While the former account predicts that context should behave as any other independent cue, with its cuing effects rendered less and less effective with an increasing number of associations shared by a given cue (Watkins & Watkins, 1975), the latter account in its pure formulation does not seem to predict such an effect, indicating that item–context ensembles are emergent entities, independent of other ensembles that may share the same context features. Three experiments conducted within the present study firmly support the prediction of the cuing-with-context account. In Experiments 1 and 2, it was found that reinstated contexts inflated retrospective confidence in correct recognition compared to re-paired contexts, not paired during study with recognition probes embedded in them, but this held true only for contexts paired with a single study item. For contexts paired with 24 different study items, no effects of context reinstatement on confidence in correct recognition decisions were observed. The same modulating role of context load was observed in Experiment 3 for prospective confidence judgments in the form of JOLs made during restudy. Restudying items in reinstated individual contexts led participants to expect better subsequent memory performance, but an analogous metacognitive pattern failed to emerge for contexts paired with 30 different study items. Together, these results indicate that context serves as an independent cue for recognition, both in explicit recognition tests and in more spontaneous instances of recognition that occur during restudy.

Throughout the article, we have assumed that the ICE model of context reinstatement (Murnane et al., 1999) does not provide any role for context load. This is based on an assumption inherent in this model, according to which ensembles are new memory representations, interactively linking item and context features, but in such a way that the emerging ensembles are equally distinctive whether they share their constituent features with other ensembles or not. Can the ICE model be modified to accommodate the modulating role of context load? If a strong assumption according to which ensembles are memory representations independent of item and context features is relaxed, and ensembles are allowed to differ in terms of distinctiveness—and thus their ability to elicit the familiarity signal—depending on an overlap in contextual features, then the ICE model can easily account for the modulating role of context load for context reinstatement effects.

Still, there is one issue that remains problematic for the ICE model when accounting for the current results, which gets to the heart of this model. The ICE model is a global matching model, which assumes only a single basis of recognition decisions—the familiarity signal elicited based on matching the probes to the contents of memory. With only a single recognition process, it is difficult to understand why a given manipulation such as context reinstatement would produce different results in measures of recognition performance and in metacognitive reflections on these processes, as most clearly visible in the results of Experiment 2. If low-load context simply increases familiarity of targets tested in those contexts, then why is this effect not seen in forced-choice recognition performance? As described later in the discussion, the dual-process model can handle such a result by assuming that familiarity sometimes occludes effects that are limited to recollective processes, but we see no such mechanism in a single-process model such as the ICE model of context reinstatement. We thus conclude that the overall patterns we report in the present article are unlikely to be accounted for by any modified version of the ICE model that would keep the fundamental assumption of a single recognition process.

Continuing the considerations of a single-process account of context reinstatement, it is worth noting that the present results, while consistent with some previous studies pointing to a role of context load in recognition (e.g., T. Isarida et al., 2018; Reder et al., 2013; Rutherford, 2004), contrast starkly with the conclusions of a recent study by Ensor et al. (2023). Ensor et al. compared the recognition effects of contexts that were either novel or familiar from outside the experimental setting, and thus presumably characterized by increased context load. They found the context reinstatement effect on recognition accuracy, but only for familiar contexts. We argue that such a reversed effect of context load is not predicted either by the cuing-with-context account, which predicts greater recognition benefits for low-load contexts, or by the ICE account, which does not consider the role of context load. This reversed effect of context load was nonetheless interpreted by Ensor et al. as supporting the ICE account, under the assumption that familiar contexts facilitated creation of item–context ensembles during study. Two issues regarding this pattern and related to the present study need to be noted. First, while an increased ease of creating ensembles for familiar contexts is one way to interpret the results of Ensor et al., an equally likely assumption could be that items are more easily associated with already familiar contexts and these stronger associations override any interference effects caused by increased context load at retrieval. Indeed, several studies have demonstrated such increased efficiency of creating associative bindings when stimuli are already familiar (e.g., Greve et al., 2017; Reder et al., 2016). Thus, the results of Ensor et al. may not uniquely favor the ICE account of context reinstatement in recognition over the cuing-with-context account.

Second, it could also be argued that the context-cuing and ICE perspectives do not need to be seen as opposite but rather can be relatively easily reconciled in a unified account of context effects. Dual-process theories of recognition, of which the cuing-with-context account is one implementation, focus on the distinction between familiarity and recollection, but they also acknowledge what can be described as an intermediate state of familiarity resulting from unitized representations. Unitization is a process by which two separate items become merged into a single entity (Graf & Schacter, 1989; Parks & Yonelinas, 2015), similar to what the ICE theory describes as ensembles when items and their contexts are involved. While in many studies unitization was promoted by specific encoding instructions (e.g., Diana et al., 2008; Robey & Riggins, 2018), it stands to reason that elements would be variably prone to being unitized depending both on their individual properties and reciprocal relationships. For example, T. Isarida et al. (2018) investigated context load effects in recognition and showed that while for semantically unrelated contexts the ability of highly loaded context to support recognition performance seems to be reduced, such a limit is absent when semantically related contexts are examined. This may well suggest that some types of contexts are easier to unitize with their respective items, giving rise to item–context ensembles and consequently to an augmented familiarity signal when later those items are presented with reinstated contexts during recognition. From this perspective, it is possible that the reversed context load effect observed by Ensor et al. does indeed reflect facilitated unitization—the process of creating item–context ensembles—for already familiar contexts. Ultimately, thus, the conclusion of the present study is emphatically not that item–context ensembles cannot be created and then utilized at test. What our study clearly demonstrates, however, is that the ICE theory cannot explain context effects in recognition on its own. Recollection of item–context associations due to context reinstatement is a real phenomenon, and one that can be prevalent under standard conditions of unrelated contexts incidentally accompanying memory encoding.

The evidence collected in the present study comes from metacognitive measures of confidence rather than measures of recognition accuracy. This is atypical. It may seem like a tautology that investigating any effect in recognition memory requires evidence coming from a measure of recognition accuracy. Indeed, much effort has gone in recent years into elucidating what this proper measure of recognition accuracy should be (see Brady et al., 2023, for a discussion). But, it seems overly simplistic to identify recognition memory with recognition accuracy. Accuracy is a single index that arises due to an aggregation of a multitude of memory processes that can give rise to the feeling of recognition (Wixted, 2007). Not all changes to one or more of these processes may be reflected in this index because of masking from other processes. Of course, in this respect, recognition accuracy is no different from metacognitive measures examined here, which also build on a variety of memory processes and thus may be variably sensitive to differences in some of these processes. An important difference, though, is that most metacognitive measures are not collected using a binary/forced-choice response format for distinguishing between old and new items but are more fine-grained.[6] They thus allow expressions of varying beliefs for any one type of a memory response, thus capturing some effects that are masked in measures of recognition accuracy. The general logical point that seems to be often ignored when interpreting empirical results, which pertains both to accuracy and metacognitive measures, is that observing a particular effect in a given measure may allow for inferring a change in a particular memory process, but it simply does not follow that a lack of such an effect allows for inferring that a particular memory process remains unchanged. The case of context effects investigated here serves as an example case of this conundrum.

The fact that context reinstatement—under conditions of low context load—affected confidence measures in the present study can only be explained by arguing that context affects recognition processes. Quite simply, reinstating context for recognition targets increased the chances of recollecting an association linking this target to its context, and this translated into participants' increased confidence that their "old" responses were correct in Experiment 1, that their identifications of targets were correct in Experiment 2, and that they would remember the cue–target pairs in the future in Experiment 3. In particular, this increased confidence was mainly driven by a shift toward responses made with highest confidence, in accord with a common assumption that recollection results in highest confidence responding (Koen et al., 2013). This memory effect on metacognition remains also fully consistent with a general observation that metacognitive judgments are based on cues that prominently feature mnemonic cues—the by-products of the memory process itself (Koriat, 1997). Indeed, it would be odd to even imagine that recollecting that a particular item was presented in a particular context would not make people more confident that their identifications of those targets are correct. This scenario is perhaps somewhat more plausible when it comes to JOLs—remembering an item–context association may or may not convince one that an item will be remembered in the future, as the same contextual features may or may not be present then. However, even in the case of JOLs, there is now ample evidence that these confidence judgments are sensitive to recollection of context information (Saenz & Smith, 2018; Zawadzka et al., 2018) and that people do infer future memorability from such recollections. To reiterate the main point, confidence, either in the form of retrospective of prospective judgments, reflects multiple cues—some of them having nothing in common with recognition (see Busey et al., 2000, for an example)—but those cues do include mnemonic cues such as contextual

[6] Note that metacognitive measures can also be assessed with a binary response format such as betting decisions (e.g., Hanczakowski et al., 2013). However, even betting decisions are more fine-grained than memory decisions because there are two responses—bet or no bet—that can be provided for each of the new and old decisions.

recollection, and thus, changes in confidence can be used as an index of recollective processes.

Perhaps a more contentious part of our argument is that there are situations in which there is a recognition memory effect, yet it finds no reflection in a recognition accuracy measure. This is particularly clear in the results of the current Experiment 2. One commentator on the present article suggested that a manipulation that enhances recollection of context associations may leave accuracy unaffected only if at the same time it impairs familiarity of targets, which seems highly unlikely in the case of context reinstatement, or any other manipulation for that matter. We would argue, however, that there is no need for such an assumption because it suffices to argue that familiarity of an item and the likelihood of recollection remain correlated. The power of recollection to improve recognition performance lies in its ability to increase the likelihood of responding "old" to a target in the case of old/new recognition or to identify a target from an array in the case of forced-choice recognition. This presupposes that recollection may change accuracy only when there is a reasonably large number of trials for which item memory fails (depending on one's theoretical perspective, it either fails entirely or familiarity falls below the old/new criterion), yet contextual recollection succeeds due to the experimental manipulation. The stronger the correlation between item and context memory, the fewer such trials should be present, undermining the sensitivity of any accuracy measure. And there is every reason to expect that item and context memory are correlated. For example, focusing attention on stimuli at encoding is likely to result in strong item and context memory, while withdrawing attention is likely to affect both. Thus, the majority of trials for which context reinstatement results in additional recollection of item–context associations could be trials for which item memory is already sufficiently strong to justify an "old" response anyway, with a net result that recollection, when it happens, has no additional effect on recognition performance.[7]

One final thought experiment may serve as an example of a situation in which a difference in memory processing determines metacognition yet does not affect a measure of accuracy. Imagine that encoding of items is so good as to result in a ceiling effect on hits—all targets are recognized correctly. But then, we apply a manipulation that enhances participants' ability to remember something particular about those targets—perhaps that on one of the study cycles, they were accompanied by a particular context. There is a memory effect there, but then, there is simply no way for it to be expressed in any measure of accuracy of item memory. One would specifically have to assess context memory to reveal it. But at the same time, a metacognitive measure concerning item memory may well be sensitive to this manipulation, with participants becoming more confident for all of their hits when accompanied by recollection. We have already described something along those lines in our previous investigation of JOLs (Zawadzka & Higham, 2015, Experiment 3). In this study, participants studied pairs of words in three study–test cycles, with JOLs collected on the last cycle. For pairs remembered correctly on one or two previous cycles, cued-recall performance on the third cycle was at ceiling—almost all of them were remembered correctly again. But JOLs did differentiate across those pairs, with higher JOLs for pairs recalled successfully on both preceding cycles rather than just on one. Clearly, those pairs were remembered differently depending on the number of previous successful retrievals, which was picked up by the metacognitive measure even as the performance measure was no longer able to reveal this difference. A ceiling effect is thus an extreme example of a situation in which performance measures become insensitive to differences still detectable with metacognitive measures.

Finally, the present analysis leads to a conclusion that a difference should be drawn between theories that deal with recognition processes and recognition accuracy. One could ask what affects both item and various types of contextual memory when previously presented items are re-presented. This is a question about memory processes, and as such, it may be answered by a variety of measures: measures of recognition accuracy on the one hand, but also measures of metacognition that do not require any modifications of the re-presentation paradigm—one still re-presents items, refrains from asking specifically about contextual details, and infers memory effects from changes in metacognitive judgments such as retrospective or prospective confidence. In our view, both cuing-with-context and ICE theories fall into this category of theories of recognition processes. Yes, they were proposed to account for recognition accuracy patterns, but they deal with memory effects rather than those patterns per se. But a different class of theories, dealing specifically with recognition performance patterns, is also possible. Indeed, a common framework of signal detection can be described as such a theory, a measurement rather than a process model. In terms of context reinstatement, we mentioned earlier an argument according to which this effect may be rarely observed in recognition performance measure due to outshining (Smith, 1988). Here, the idea was that "the environment can be suppressed at test, diminishing the likelihood that ambient environmental information will be used in the construction of memory probes" (Smith & Vela, 2001, p. 206). This is a formulation in terms of recognition processes, which our study refutes, demonstrating that environmental context is used as an independent cue in construction of memory probes. But the outshining hypothesis can be reformulated in terms of a theory of recognition accuracy, in which case context is outshone by item memory in its ability to influence performance measures, even if it does affect recollection of item–context associations.

[7] It is worth acknowledging the potentially crucial role of the strength of the effect of interest for these considerations. If reinstating the encoding context constituted a particularly powerful manipulation, its effects would likely be picked up by the measures of accuracy. However, both the modest effect sizes of the reinstatement effects observed in the present study and the numerous examples of failed attempts to document context reinstatement in recognition accuracy measures present in the published literature (e.g., Hockley, 2008; Murnane & Phelps, 1993) clearly indicate that context reinstatement is not a particularly strong manipulation for recognition processes. In other words, it needs to be emphasized that context reinstatement may lead to some additional recollection, but it is not likely to affect recognition for all targets included in a recognition test.

References

Bayen, U. J., Phelps, M. P., & Spaniol, J. (2000). Age-related differences in the use of contextual information in recognition memory: A global matching approach. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *55*(3), P131–P141. https://doi.org/10.1093/geronb/55.3.P131

Benjamin, A. S. (2010). Representational explanations of "process" dissociations in recognition: The DRYAD theory of aging and memory judgments. *Psychological Review*, *117*(4), 1055–1079. https://doi.org/10.1037/a0020810

Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2023). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. *Psychonomic Bulletin & Review*, *30*(2), 421–449. https://doi.org/10.3758/s13423-022-02179-w

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence–accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, *7*(1), 26–48. https://doi.org/10.3758/BF03210724

deBettencourt, M. T., Norman, K. A., & Turk-Browne, N. B. (2018). Forgetting from lapses of sustained attention. *Psychonomic Bulletin & Review*, *25*(2), 605–611. https://doi.org/10.3758/s13423-017-1309-5

Diana, R. A., Peterson, M. J., & Reder, L. M. (2004). The role of spurious feature familiarity in recognition memory. *Psychonomic Bulletin & Review*, *11*(1), 150–156. https://doi.org/10.3758/BF03206475

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2008). The effects of unitization on familiarity-based source memory: Testing a behavioral prediction derived from neuroimaging data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 730–740. https://doi.org/10.1037/0278-7393.34.4.730

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, *24*(4), 523–533. https://doi.org/10.3758/BF03200940

Ensor, T. M., Surprenant, A. M., Neath, I., & Hockley, W. E. (2023). Using preexperimental familiarity to compare the ICE and cue-overload accounts of context-dependent memory in item recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(4), 533–546. https://doi.org/10.1037/xlm0001169

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Godden, D. R., & Baddeley, A. D. (1980). When does context influence recognition memory? *British Journal of Psychology*, *71*(1), 99–104. https://doi.org/10.1111/j.2044-8295.1980.tb02735.x

Goh, W. D., & Lu, S. H. X. (2012). Testing the myth of the encoding-retrieval match. *Memory & Cognition*, *40*(1), 28–39. https://doi.org/10.3758/s13421-011-0133-9

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

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, *94*, 149–165. https://doi.org/10.1016/j.jml.2016.11.001

Gruppuso, V., Lindsay, D. S., & Masson, M. E. J. (2007). I'd know that face anywhere! *Psychonomic Bulletin & Review*, *14*(6), 1085–1089. https://doi.org/10.3758/BF03193095

Halamish, V., & Undorf, M. (2023). Why do judgments of learning modify memory? Evidence from identical pairs and relatedness judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(4), 547–556. https://doi.org/10.1037/xlm0001174

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

Hanczakowski, M., Zawadzka, K., & Coote, L. (2014). Context reinstatement in recognition: Memory and beyond. *Journal of Memory and Language*, *72*, 85–97. https://doi.org/10.1016/j.jml.2014.01.001

Hanczakowski, M., Zawadzka, K., & Macken, B. (2015). Continued effects of context reinstatement in recognition. *Memory & Cognition*, *43*(5), 788–797. https://doi.org/10.3758/s13421-014-0502-2

Hanczakowski, M., Zawadzka, K., Pasek, T., & Higham, P. A. (2013). Calibration of metacognitive judgments: Insights from the underconfidence-with-practice effect. *Journal of Memory and Language*, *69*(3), 429–444. https://doi.org/10.1016/j.jml.2013.05.003

Higham, P. A., Perfect, T. J., & Bruno, D. (2009). Investigating strength and frequency effects in recognition memory using type-2 signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 57–80. https://doi.org/10.1037/a0013865

Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the 'coordinates of truth'. *Perspectives on Psychological Science*, *6*(3), 253–271. https://doi.org/10.1177/1745691611406924

Hockley, W. E. (2008). The effects of environmental context on recognition memory and claims of remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1412–1429. https://doi.org/10.1037/a0013016

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

Isarida, T., Isarida, T. K., Kubota, T., Higuma, M., & Matsuda, Y. (2018). Influences of context load and sensibleness of background photographs on local environmental context-dependent recognition. *Journal of Memory and Language*, *101*, 114–123. https://doi.org/10.1016/j.jml.2018.04.006

Isarida, T., Isarida, T. K., Kubota, T., Nakajima, S., Yagi, K., Yamamoto, A., & Higuma, M. (2020). Video context-dependent effects in recognition memory. *Journal of Memory and Language*, *113*, Article 104113. https://doi.org/10.1016/j.jml.2020.104113

Isarida, T. K., Kubota, T., Nakajima, S., & Isarida, T. (2017). Reexamination of mood-mediation hypothesis of background-music-dependent effects in free recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *70*(3), 533–543. https://doi.org/10.1080/17470218.2016.1138975

Jang, Y., Wixted, J. T., & Huber, D. E. (2009). Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory. *Journal of Experimental Psychology: General*, *138*(2), 291–306. https://doi.org/10.1037/a0015525

Koen, J. D., Aly, M., Wang, W.-C., & Yonelinas, A. P. (2013). Examining the causes of memory strength variability: Recollection, attention failure, or encoding variability? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1726–1741. https://doi.org/10.1037/a0033671

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

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Macken, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 153–161. https://doi.org/10.1037/0278-7393.28.1.153

Mandera, P., Keuleers, E., Wodniecka, Z., & Brysbaert, M. (2015). Subtlex-pl: Subtitle-based word frequency estimates for Polish. *Behavior Research Methods*, *47*(2), 471–483. https://doi.org/10.3758/s13428-014-0489-4

Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, *145*(2), 200–219. https://doi.org/10.1037/a0039923

Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 882–894. https://doi.org/10.1037/0278-7393.19.4.882

Murnane, K., & Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory & Cognition*, *22*(5), 584–590. https://doi.org/10.3758/BF03198397

Murnane, K., & Phelps, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 158–172. https://doi.org/10.1037/0278-7393.21.1.158

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, *128*(4), 403–415. https://doi.org/10.1037/0096-3445.128.4.403

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*(5–6), 389–395. https://doi.org/10.1080/09658210244000216

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

Poirier, M., Nairne, J. S., Morin, C., Zimmermann, F. G. S., Koutmeridou, K., & Fowler, J. (2012). Memory as discrimination: A challenge to the encoding-retrieval match principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 16–29. https://doi.org/10.1037/a0024956

Reder, L. M., Donavos, D. K., & Erickson, M. A. (2002). Perceptual match effects in direct tests of memory: The role of contextual fan. *Memory & Cognition*, *30*(2), 312–323. https://doi.org/10.3758/BF03195292

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

Reder, L. M., Victoria, L. W., Manelis, A., Oates, J. M., Dutcher, J. M., Bates, J. T., Cook, S., Aizenstein, H. J., Quinlan, J., & Gyulai, F. (2013). Why it's easier to remember seeing a face we already know than one we don't: Preexisting memory representations facilitate memory formation. *Psychological Science*, *24*(3), 363–372. https://doi.org/10.1177/0956797612457396

Robey, A., & Riggins, T. (2018). Increasing relational memory in childhood with unitization strategies. *Memory & Cognition*, *46*(1), 100–111. https://doi.org/10.3758/s13421-017-0748-6

Russo, R., Ward, G., Geurts, H., & Scheres, A. (1999). When unfamiliarity matters: Changing environmental context between study and tests affects recognition memory for unfamiliar stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 488–499. https://doi.org/10.1037/0278-7393.25.2.488

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

Saenz, G. D., & Smith, S. M. (2018). Testing judgments of learning in new contexts to reduce confidence. *Journal of Applied Research in Memory and Cognition*, *7*(4), 540–551. https://doi.org/10.1016/j.jarmac.2018.07.003

Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davis & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13–33). Wiley.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*(4), 342–353. https://doi.org/10.3758/BF03197465

Smith, S. M., Handy, J. D., Angello, G., & Manzano, I. (2014). Effects of similarity on environmental context cueing. *Memory*, *22*(5), 493–508. https://doi.org/10.1080/09658211.2013.800553

Smith, S. M., & Manzano, I. (2010). Video context-dependent recall. *Behavior Research Methods*, *42*(1), 292–301. https://doi.org/10.3758/BRM.42.1.292

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

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 553–558. https://doi.org/10.1037/a0038388

Tullis, J. G., Benjamin, A. S., & Ross, B. H. (2014). The reminding effect: Presentation of associates enhances memory for related words in a list. *Journal of Experimental Psychology: General*, *143*(4), 1526–1540. https://doi.org/10.1037/a0036036

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4), 442–452. https://doi.org/10.1037/0278-7393.1.4.442

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

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

Zawadzka, K., & Higham, P. A. (2015). Judgments of learning index relative confidence, not subjective probability. *Memory & Cognition*, *43*(8), 1168–1179. https://doi.org/10.3758/s13421-015-0532-4

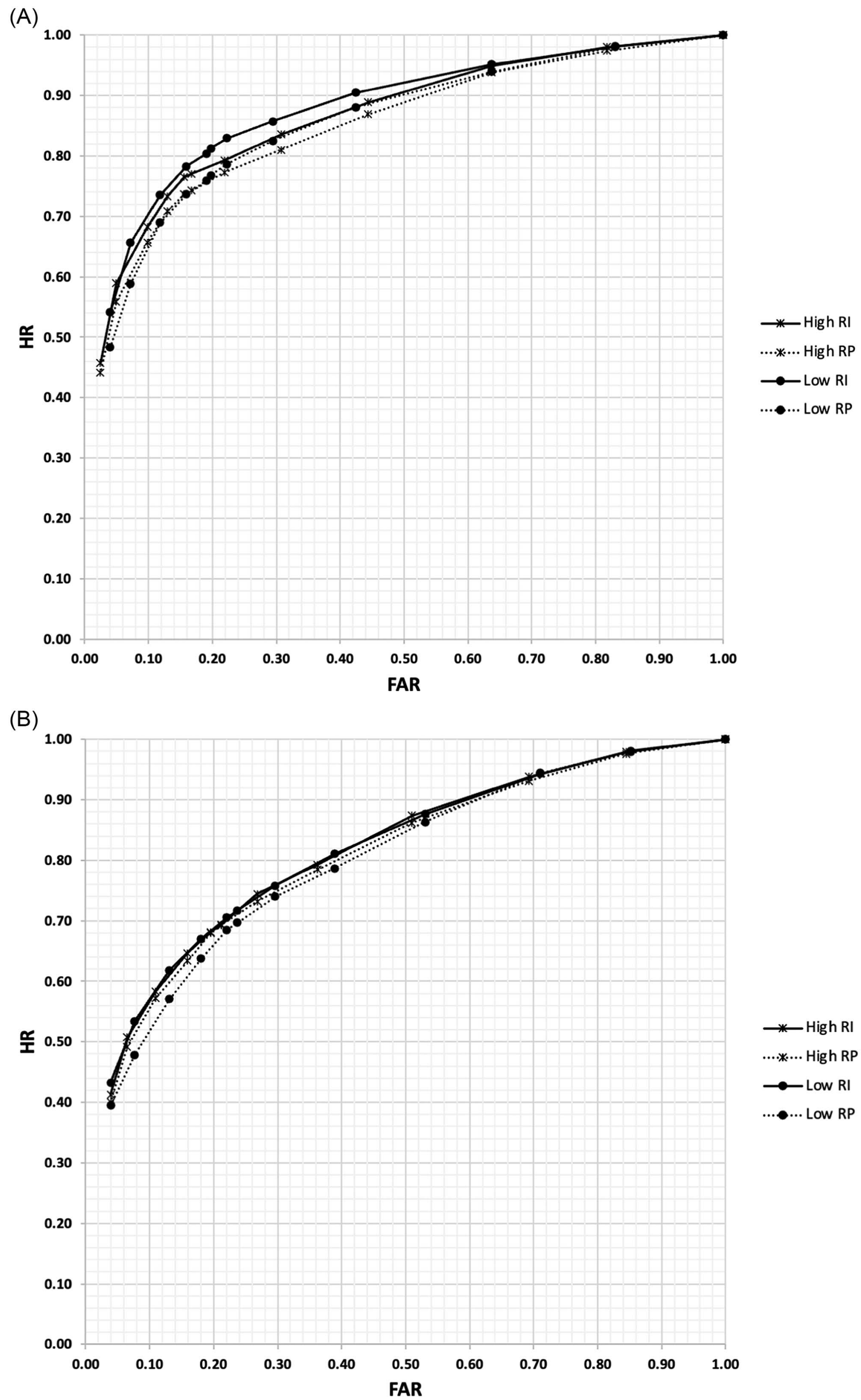
Zawadzka, K., & Higham, P. A. (2016). Recalibration effects in judgments of learning: A signal detection analysis. *Journal of Memory and Language*, *90*, 161–176. https://doi.org/10.1016/j.jml.2016.04.005

Zawadzka, K., Simkiss, N., & Hanczakowski, M. (2018). Remind me of the context: Memory and metacognition at restudy. *Journal of Memory and Language*, *101*, 1–17. https://doi.org/10.1016/j.jml.2018.03.001

(*Appendix follows*)

Appendix

ROC Curves as a Function of Context Load (High vs. Low) and Recognition Context (Reinstated vs. Re-Paired) in the Between-List (Panel A) and Within-List Conditions (Panel B) in Experiment 1



Note. HR = hit rate; FAR = false alarm rate; RI = reinstated context condition; RP = re-paired context condition.

Received October 4, 2023
Revision received June 4, 2024
Accepted June 13, 2024 ■