THE ROLE OF CONTEXT REINSTATEMENT AND UPDATING IN RETRIEVAL PRACTICE

A Dissertation

by

ALAN HERNANDEZ CORTES

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Steven M. Smith
Committee Members,	Joseph M. Orr
	Louis G. Tassinary
	Darrell A. Worthy
Head of Department,	Heather C. Lench

August 2020

Major Subject: Psychology

ABSTRACT

There is a considerable amount of research showing retrieval practice consistently outperforms restudying as a learning method. This superior performance, also termed the *Testing Effect* (TE), is posited as a result of the encoding and retrieval of contextual elements by the *Episodic Context Account* (ECA). The ECA relies on four assumptions to explain the beneficial effects of retrieval practice. Because part of one assumption, the ECA states the difficulty involved in mentally, or self-reinstating, the context of a target memory is related to the enhancement it receives during its retrieval. The ECA also assumes the encoding and updating of contextual elements contributes to the benefit of retrieval practice. In six experiments, (pictorial) context self-reinstatement difficulty and updating were examined to test their roles within retrieval practice. Experiment 1 – 3 opted for a single practice/study block design before a final memory test. This set of experiments showed a lack of a TE, and even its reversal, when practice conditions enabled self-reinstatement of context and updating. Experiments 4 – 6 increased the number of practice/study blocks to three to test those same factors as part of an extended learning schedule. These latter experiments did yield TEs, but practice conditions, of varying degrees of self-reinstatement and updating, did not differ from one another. In general, it was the use of context cues, either through self-reinstatement or when provided during practices, that appeared to confer benefits to retention.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

The present work was conducted under the supervision of Professor Steven Smith, Professor Joseph Orr, and Professor Darrel Worthy of the Department of Psychological and Brain Sciences, and Professor Louis Tassinary of the Department of Visualization.

All work for this dissertation was conducted by the student independently.

Funding Source

The work in this dissertation required no outside funding.

TABLE OF CONTENTS

Page

ABSTRACT ii

CONTRIBUTORS AND FUNDING SOURCES iii

TABLE OF CONTENTS iv

LIST OF FIGURES vii

1. INTRODUCTION 1

2. RETRIEVAL PRACTICE 6

2.1. Retrieval Practice Frameworks 11

3. EPISODIC CONTEXT ACCOUNT 19

4. PRESENT EXPERIMENTS 30

4.1. Desirable Difficulties 31
4.2. Context Updating 35
4.3. Context Dependence 38
4.4. Practice Blocks 41
4.5. Context Specificity 42
4.6. Present Questions 46

5. GENERAL METHOD 54

6. EXPERIMENT 1 56

6.1. Method 57
6.1.1. Participants 57
6.1.2. Design and Materials 57
6.1.3. Procedures 58
6.2. Results 59
6.3. Discussion 60

7. EXPERIMENT 2 63

7.1. Method 65
7.1.1. Participants 65
7.1.2. Design and Materials 65
7.1.3. Procedures 65
7.2. Results 66
7.3. Discussion 68

8. EXPERIMENT 3 72

8.1. Method 73
8.1.1. Participants 73
8.1.2. Design and Materials 73
8.1.3. Procedure 74
8.2. Results 74
8.3. Discussion 76

9. EXPERIMENTS 4-6 79

10. EXPERIMENT 4 81

10.1. Method 82
10.1.1. Participants 82
10.1.2. Design and Materials 82
10.1.3. Procedures 82
10.2. Results 83
10.3. Discussion 86

11. EXPERIMENT 5 88

11.1. Method 88
11.1.1. Participants 88
11.1.2. Design and Materials 89
11.1.3. Procedures 89
11.2. Results 89
11.3. Discussion 93

12. EXPERIMENT 6 95

12.1. Method 96
12.1.1. Participants 96
12.1.2. Design and Materials 96
12.1.3. Procedures 96
12.2. Results 97
12.3. Discussion 99

13. SUMMARY AND CONCLUSIONS 102

REFERENCES 114

LIST OF FIGURES

Page

Figure 1 Basic Retrieval Practice Design 3

Figure 2 Context Retrieval and Updating 27

Figure 3 Phases of Experiment 1 55

Figure 4 Experiment 1 Results 60

Figure 5 Phases of Experiment 2 67

Figure 6 Experiment 2 Final Test Results 69

Figure 7 Experiment 2 Retrieval Practice Results 69

Figure 8 Phases of Experiment 3 75

Figure 9 Experiment 3 Results 76

Figure 10 Phases of Experiment 4 83

Figure 11 Experiment 4 Results 85

Figure 12 Phases of Experiment 5 90

Figure 13 Experiment 5 Results 92

Figure 14 Phases of Experiment 6 97

Figure 15 Experiment 6 Results 99

1. INTRODUCTION

It is an intuitive habit to mentally rehearse a piece of information we wish to commit to memory. We repeat phone numbers, names, or locations to ourselves as many times as possible as soon as it is determined such information needs to be remembered. Be it internally or through mutterings, we assume such repetitions in rapid succession will let us encode and retain them better. A similar method is applied during our formal education, that is, we restudy or re-read lectures notes, books, and other materials to prepare for an exam. Even if not explicitly acknowledged, we might assume that the more times we read through our notes the better they will be learned and proceed to re-read them multiple times. To best prepare for an exam we increase the amount of times we re-read our materials until we deem them well learned, correct? While repeated exposure to (to-be-learned) material can help improve retention, this might not be the best way to study. Specifically, simply re-reading or restudying might not confer the best results for long-term retention. A possibly more efficient method has been proposed for some time (Roediger & Karpicke, 2006a, 2006b), that of retrieval practice. This method has been shown to consistently outperform restudying (Rowland, 2014) through its use of "pre-" testing before a final test. There have been several attempts to explain the exact mechanisms responsible for the effectiveness of retrieval practice (e.g., Carpenter & DeLosh, 2006; Karpicke, Lehman, & Aue, 2014). One in particular seems to fare a bit better than the rest, as it appears more comprehensive and adequately supported by other studies. The Episodic Context Account (ECA, Karpicke, Lehman, & Aue, 2014) is such attempt and it does so by positing four assumptions. Despite the support it has received not all of its assumptions have been fully tested. To this end several experiments were conducted, and are herein presented, to help determine whether two of

the ECA's assumptions have an active role in retrieval practice, namely, the assumptions of context retrieval and updating during retrieval practice are tested.

The better performance by retrieval practice over restudy, on a final memory measure or test, is termed the Testing Effect (TE). In its simplest form a retrieval practice paradigm consists of a retrieval practice and a restudy condition. This paradigm would include an initial encoding block, followed either by retrieval practice block(s) which is sometimes accompanied by corrective feedback, or by a restudy block(s) in which materials are typically re-read, and then a final memory test (Figure 1). The number of practice or restudy blocks, the presence of feedback, and the type of final test are all variables which can have an impact on the strength of the TE (Rowland, 2014; Karpicke, Lehman, & Aue, 2014), albeit it not necessarily to the same extent. Their mention here is to underscore that an account aiming to explain how retrieval practice works will have to contend with detailing their impact on the TE. An account as expansive as that does not exist, but the ECA does appear to be the most well-rounded as it suggests mechanisms that are at play during retrieval practice. In stating what portions of a learning event are encoded, retrieved, and enhanced (or strengthened) it can give a clearer idea of what creates the TE. As a brief example, the better performance afforded by restudy, over retrieval practice, after a short retention interval (i.e., the amount of time between the last restudy block and a final test) reverses if the interval is expanded (by hours or days). The benefit of retrieval practice, for long-term retention, is handled by the ECA as the reliance on the retrieval of contextual elements. At short retention intervals it is easier to think back on the targets themselves or on associated contextual elements used as indirect retrieval routes to the targets. If attempting a direct retrieval fails, then if the test shares a similar temporal context to the restudy block (e.g., they are close in time), the contextual test elements can be used as cues to retrieve the context

elements associated with the targets. At longer intervals, temporal context changes (e.g., more time passes by) making context retrieval, or the use of an indirect route, more difficult. Retrieval practice is not hampered by this change in temporal context as it can enable the creation of more contextual associations, thus increasing the number of retrieval routes that can be used to retrieve an associated target memory. In contrast, this creation of extra retrieval routes is absent, or stunted, in restudy (see Episodic Context Account section below). This shows that the ECA assigns a key role to the encoding and retrieval of contextual associations and can explain some of the more peculiar results of the TE, such as the retention interval interaction. The ECA is not the only framework using context to explain this effect, but this serves to introduce how it incorporates several existing ideas/concepts into a cogent account. The integration of these concepts is part of the reason why it can explain several aspects of the TE. Moreover, it further specifies how mechanisms are to work within retrieval practice, which can be something that might be absent in the standalone version of such ideas. It still remains to be tested just how well exactly the assumptions of the ECA go together. The ECA's, seemingly facile, assimilation of the several features that make up its assumptions is what warrants their testing.

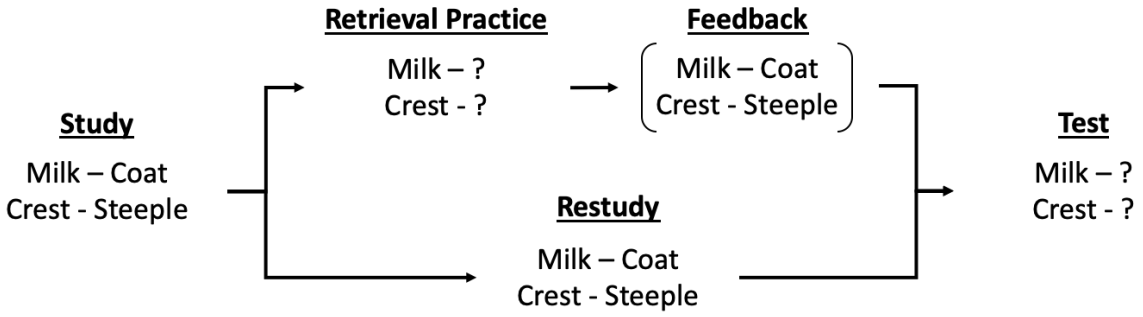

Figure 1. Basic Retrieval Practice Design. After an initial study of cue-target word pairs, a practice test can be given (i.e., retrieval practice) (top path) with corrective feedback or pairs can be restudied (bottom path), then followed by a final cued-recall test.

The importance behind further examining the TE is not only to clarify and explain how retrieval practice works, but also to further refine the method if possible. In identifying the most important factors, those conveying the greatest benefit to learning, and trimming or removing those that do not contribute meaningfully to the effect, retrieval practice could potentially be honed to become an even more effective strategy. Because retrieval practice has been shown to aid the learning of more than just lab friendly stimuli (e.g., word lists), such as prose passages and other conceptually complex stimuli (Karpicke & Aue, 2015; Roediger & Karpicke, 2006b), its coupling with other methods may further enhance its effects (e.g., concept mapping, see Karpicke & Blunt, 2011). Admittedly speculative, the exploration of other potential combinations could help refine retrieval practice to better suit the material being learned. Yet another possible line of research is the refining of the schedule of practices. As it is, an equally spaced interval between practices yields a similar level of performance as expanded practice, or the incremental increase of interval length between practices, (Balota, Duchek, & Logan, 2007). The exact reason for this result has not received a lot of attention and its research would help in determining what could constitute the optimal interval between practices. The similar performance between schedules could actually be explained in terms of the usage of temporal context elements, an idea shared by the ECA. But it would also be a sign that the idea of retrieval aid through contextual elements needs to be qualified as a strict application of it would predict expanded practice to outdo spaced practice. This suggests the manner in which context is used during retrieval practice is limited, by what and at which point(s) remains to be found and elucidated. Furthermore, the retrieval practice literature (at least the subset emphasizing context use) has been vague as to what aspects of temporal context are actually used during practice. Because the blanket term temporal context can include many types (e.g., physical, mental,

experimental) it can be difficult to exactly pin down which type is primarily being used in a given experiment. More research is needed to fully discover how specific types of context is used in retrieval practice. Further experiments using specific types of context could identify which might actually have a role in the TE. Also, context use during practice has mostly been proposed as a positive, but the prospect of developing context dependence (e.g., with the repeated use of context cues as exclusive means for target retrieval) has not been adequately addressed. That is, in terms of context use neither specificity nor its consequences have been thoroughly researched in regard to retrieval practice. Overall, there is still much to clarify and expand to better understand the TE.

The proceeding sections take the structure of an overview of retrieval practice, the ECA, which is followed by the presentation of several experiments conducted to test the assumptions of the ECA, and finally, the supporting and conflicting results for the assumptions are discussed along with future research topics. To briefly expand on this, the first of these general sections will breakdown several subtopics relating to retrieval practice, the basic findings associated with obtaining a TE, the further elaboration on the ECA and its strengths and weaknesses, and neglected areas of retrieval practice that still need to be addressed. The subsequent sections will begin with a general description for the present six experiments and fully explain their research agenda. This will include the assessment of the two aforementioned assumptions and other issues that will also be indirectly tested, as well. The last section will be devoted to accurately place the obtained results within the literature of retrieval practice. That is, given the mixed evidence found for the assumptions of the ECA, how is it interpreted in the context of previous work?

2. RETRIEVAL PRACTICE

It may not be common to think of retrieval as having a significant role in learning aside from that of assessment. The beginnings of the idea can be traced to findings in the early 1900s in studies relating to education (for a short historical account see Roediger and Karpicke, 2006a). Early work recognized the need for research closely matching classroom learning conditions. For example, Spitzer (1939) used grade school students as participants in his study researching the effect of testing on retention. Participants read and were later tested (at different time points) over articles intended to match the type of material learned in a classroom. Spitzer concluded early testing benefitted long-term retention. In addition, he suggested ideas that are still relevant to testing, such as the availability of feedback and the use of testing as learning aid and not just an assessment tool. Later, further work showed the positive impact testing even after failures. Izawa (1970) administered series of study-test blocks, in which participants would study first then take one or multiple tests before restudying again (which was followed by the same number of tests as before). He found greater amounts of testing improved retention and a greater number of tests in between study sections reduced errors on the tests after restudying. Such examples highlight the resistance to forgetting and long-term benefits testing provides retention as also found by other studies before the relatively recent "resurgence" of testing studies (for an extended historical account of early testing research see Richardson, 1985).

Eventually, the idea of fully delving into testing (or retrieval) as possibly having an important role in the learning process became what gave rise to the pair of articles (Roediger & Karpicke, 2006a, 2006b) that arguably started the more recent trend of retrieval practice studies. But, there was yet other work that held the process of retrieval to have a function beyond checking the contents of memory (e.g., final memory test). Retrieval need to be confined to the

role of assessing memory accessibility, that is, it is not an act that merely brings memories to mind, but one that also modifies them. This modification was suggested to be influenced by the difficulty of the retrieval process, as it might produce a more stable memory and allow the creation of several retrieval routes, assuming retrieval is not easy or direct (Bjork, 1975). If retrieval is a memory modifier it suggests it is possible to enhance learning by engaging in multiple retrieval attempts (or testing) before a final test. This means that practice test(s) can serve as a learning event(s) if, for example, recalling a memory could strengthen it. Accordingly, retrieval practice would confer a benefit to learning due to this notion.

Among the top reasons for recommending retrieval practice (over restudy) as a learning method is the help it gives to memory performance after long intervals. This is well exemplified by Roediger and Karpicke's (2006b) results on a final recall test after a given retention interval length. In their study participants read prose two passages and were asked to restudy one and practice retrieval on the other (after its initial reading/studying). A final recall test, over the contents of the passages, was given after 5 minutes, 2 days, or one week. Their results showed a clear benefit to practicing retrieval, rather than restudying, if tested after long intervals. Performance after 2 days and one week had a TE, whereas a reversal of the effect was found if tested after 5 minutes. This pattern was replicated in their second experiment, which was similar to their first but required the learning of only one passage and increased the amount of times participants would restudy or practice retrieval (e.g., four blocks of studying, three blocks of studying and one block of practicing retrieval, or one of block of study and three of retrieval practice). Participants showed a TE on a recall test after one week, but, once again, not after 5 minutes. Interesting enough, final test performance increased (with a long retention interval) as more testing was given during the acquisition blocks. This result was also reversed if tested

shortly after acquisition, that is, the less testing present during these blocks the higher the performance. This result of the TE is noteworthy for a few reasons. First, it shows that even practice testing only once (i.e., three study blocks and one practice) still obtained a TE, and practicing multiple times yielded the greatest benefit, attesting the effectiveness of retrieval practice to enhance long-term retention. Second, the results illustrate how performance should not be mistaken for learning (e.g., see Soderstrom & Bjork, 2015). If the experiments had only tested after a short interval the results would show a benefit to restudying rather than practice, but performance after 5 minutes does not actually show how well the passages were learned as seen from the better recall scores after one week. If the point of learning is to retain information long after its immediate exposure (or study) then restudying did not attain that goal as effectively as repeated testing. The authors reported forgetting percentages by condition for their second experiment and the only-study conditions showed the biggest percentage (i.e., just over fifty percent) in forgetting relative to the other conditions. The conditions with only one practice test almost halved that percentage (i.e., a bit less than thirty percent), and the multiple test conditions had the lowest forgetting percentage (i.e., fourteen percent). The better performance after restudying, and a short retention interval, would misleadingly represent how well that information would actually be retained in the long-term. On the other hand, the lesser forgetting (after one week) that occurred with practice testing shows that even assessing learning after 5 minutes still gives a better idea of how well that information will be retained because it is a closer estimate of what will be remembered in the long run, this is especially the case with the repeated testing condition. In short, practicing retrieval helped learn the presented material much better. Positive results such as these make retrieval practice a prime candidate as a learning strategy as it is more conducive to long-term retention.

Yet another issue to note is that, in general, Roediger and Karpicke's experiments used multiple blocks, be it studying or practice, during the acquisition phases. It is not surprising such an experimental choice was taken given that re/studying and retrieval practice are being looked at as learning strategies, so that conditions within their experiment should match, as closely as possible, the conditions under which people might use them in real-life. That is, while learning some material we might repeatedly go over it (be it by restudying or testing), so it makes sense to try to replicate such conditions. But this also brings up another issue, namely, that of interval length between learning blocks. Finding the optimal schedule, or spacing, of learning blocks has mostly shown that spaced presentations, rather than massed, are more conducive to learning (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Does that settle the type of schedule should retrieval practice follow? Will spaced *retrieval practice* receive benefit in the same manner as spaced *presentations* (or *exposures*)? The question has been addressed several times and, overall, the answer converges on the result that spacing does give better performance. Furthermore, research into possible benefits from the modification of spacing intervals has shown equal spacing between practice blocks to give a comparable result as expanded practice (Balota, Duchek, & Logan, 2007), meaning that increasing the length of each interval as blocks progress does not appear to give any benefit. Although, there is evidence expanded practice does result in better performance compared to equal spacing (Landauer & Bjork, 1978), such results have not been consistently found. One possibility for this discrepancy in the spacing of (retrieval) practice literature could be due to the content of the intervals. Specifically, if information that could potentially interfere with that of the to-be-learned material is present during an interval then expanding retrieval might be a better choice than equal spacing. Such a result was reported by Storm, Bjork, and Storm (2010) for experiments in which the intervals between practices were

filled with the reading of passages similar (i.e., geographic information) to that of the main, or to-be-tested, passage (i.e., containing information about Antarctica). Adding to the generalizability of the results, the benefit was found using free recall and cued-recall tests, as well as fill-in-the-blank questions. Plus, this benefit was found using a long retention interval of one week, which contrasts with previous work which found the same benefit but only with a short retention interval, and equally spaced practice actually resulted in better performance after a long interval (Karpicke & Roediger, 2007). The authors concluded that material with higher susceptibility to forgetting would benefit from being learned with an expanded schedule. Given the mixed set of results further clarification is still needed to better understand the conditions under which one schedule might be better than the other. Nonetheless, practice with spacing, be it equal or expanded, does appear to reliably help retention more than massed practice (Balota, Duchek, & Logan, 2007).

It is in the interest of developing a comprehensive account of retrieval practice, and the TE, that the superiority of spaced over massed practice should be thoroughly investigated. An oft repeated account for this effect, and the TE in general, is that of encoding variability (Melton, 1970; Cepeda, et al., 2006; Balota, Duchek, & Logan, 2007). Encoding variability (Martin, 1968) states that a stimulus seen multiple times can create different associations on each presentation, which should facilitate stimulus recall as it will have several retrieval routes. As it pertains to retrieval practice, these associations are posited to be contextual. That is, if a stimulus is practiced at two different times it is often assumed the contexts are quite dissimilar, especially the further they are spaced apart. This would imply that a stimulus would then become associated with two contexts or more retrieval routes, which can then be used to search for the stimulus, thus increasing the chances of successful retrieval. This version of encoding variability, which is

contextual variability, is at least the one proposed by the ECA. Evidence for a broader version of encoding variability is considered lacking (for a short account see Karpicke, Lehman, & Aue, 2014) and is part of the reason why the ECA explicitly states the type of variability that contributes to obtaining a TE. The exact manner in which the ECA proposes contextual variability is used will be detailed in a later section. Currently, it is enough to state that (at least one form of) encoding variability is suspected of being partially responsible for the TE. While there may be other explanations for the TE, encoding/context variability will be further discussed later on as it figures prominently in the ECA.

2.1. Retrieval Practice Frameworks

But first, to present a brief survey of the literature and for it to serve as a contrast to the ECA, a second candidate vying to explain the TE should be discussed, the elaborative retrieval hypothesis (Carpenter, 2009). Basically, this account relies on the activation of semantic information during practice tests. For example, in trying to retrieve, at practice, the target from the word pair lamp – leaf (upon the presentation of the cue word, lamp) we may elaborate and begin associating semantically related words. That is, we may think of a lamp placed on a desk next to a plant which would have leaves. This means the elaboration or linking of lamp and leaf would contain other concepts such as desk and plant. These could then be used at retrieval to activate the target word. Whereas restudy is not held to elicit such elaboration which would significantly reduce the number of retrieval routes. This account also incorporates the idea of retrieval difficulty as potentially enhancing retention as greater difficulty would require greater amounts of elaboration. While the elaborative retrieval hypothesis has shown some support (Carpenter & DeLosh, 2006) it is worth noting that explicitly instructing elaboration during retrieval has not shown consistent results. Karpicke and Zaromb's (2010) experiments included a

comparison between a generation, retrieval, and a read condition. The first was considered as essentially an elaboration task as its instructions called for the completion of word fragment with the first word that came to mind. The second was taken as comparable to a standard retrieval practice condition and the third to a restudy condition. Overall, their results from multiple experiments found a pattern consistent with performance, on recall and recognition tests, being higher for the retrieval condition rather than the generation condition. In particular, one study showed generation doing about the same as reading. The authors explained this as the retrieval condition's use of intentional retrieval, whereas the generate condition relied on incidental retrieval. This difference in retrieval mode was concluded to be such a significant influence that thinking back to an original study event has been proposed, including by the ECA, as important to the benefits of retrieval practice (Karpicke, Lehman, & Aue, 2014). The elaborative retrieval hypothesis has also failed to receive supportive results as concept mapping (yet another task requiring elaboration) has also not outperformed "standard" retrieval practice nor restudying (Karpicke & Blunt, 2011). Interestingly, if both retrieval practice and concept mapping are combined the results are quite positive. In the second of Blunt and Karpicke's (2014) experiments they had participants read two short texts and practice retrieving them as either of two activities, that is, in paragraph form or by concept mapping. Participants would read one text for a few minutes then recall as much as they could in a paragraph or create a concept map. They would then re-read the text and do the same retrieval task. Next, they would repeat this same process with the second text. These two conditions (mapping or writing a paragraph) was crossed with the presence or absence of the text during that activity. A short answer test was given one week afterwards which showed both conditions had a comparable performance. In general, their results showed that both concept mapping and writing a paragraph in the absence of the text

yielded better results on the final test. This performance was relative to those same conditions when the text was available during the activities. This suggests attempting to retrieve the text conferred a benefit, whereas being able to consult the text while doing the activities (i.e., in the absence of retrieval) undercut their potential. As whole, evidence supporting the elaborative retrieval hypothesis is not strong (for other objections see Karpicke, Lehman, & Aue, 2014), although given the foregoing results there can still be a place for a form of elaboration to contribute to the TE. Nonetheless, evidence points to a key aspect of the superior performance of retrieval practice being the active, or intentional, retrieval of a study event.

Indeed, the intentional retrieval of a previous event has been previously suggested as partially responsible for helping enhance memory. A proponent of this idea is the *reminding effect* (Benjamin & Tullis, 2010). In short, the effect states that when the presentation of a stimulus acts as a reminder to a previous event the memory for the latter is enhanced. For example, if the cue-target pair fiber – edge is studied and the subsequent presentation of the cue reminds of the target, then the target's memory is enhanced. The effect can technically apply to both repeated practice and restudy as long as the event is retrieved. But, as applied to retrieval practice it helps explain how it may create the TE. In such a case, a practice trial serves as a potential reminder for the original study event, which if recalled, can be strengthened. The magnitude of the enhancement is posited to be inversely proportional to the likelihood of a reminding. For instance, the more likely a cue is to remind of the target, such as if they were strongly semantically related, then the weaker the enhancement the target would receive. The less likely a reminding occurs, then the greater the enhancement. A low likelihood of reminding can be due to the stimuli (e.g., cue-target pair) being semantic unrelated or if a considerable amount of time has passed between the original and the reminding event. The latter example is

particularly relevant to retrieval practice given the better results of spaced over massed practice. It is also suggested the *reminding effect*, just as the ECA, may also benefit from the intentional retrieval of a past event (Tullis, Benjamin, & Ross, 2014). But, aside from the shared dependence on a form of desirable difficulties (Bjork & Bjork, 2011), i.e., enhancement from reminding likelihood, and intentional retrieval, the *reminding effect* does not have the same scope as the ECA to explain the TE. Granted, the reminding effect was not developed for the purpose of detailing the workings of retrieval practice, rather it was proposed as a way to explain distributed practice in general, which can include a variety of types of stimuli presentations. Because there is bit of a lack of studies looking specifically at the reminding effect within retrieval practice it does not really make it an attractive alternative to explain the TE. In addition, reminding has shown the interesting effect of mainly strengthening the reminded memory and not the one doing the reminding (e.g., a cue, Tullis, Benjamin, & Ross, 2014) which has not quite been seen within a retrieval practice study, although it would be interesting to research such a topic. It would show a negative "side effect" of practicing retrieval that has not really been considered before. Unlike the ECA or the elaborative retrieval hypothesis, the reminding effect does not fully develop a mechanism to explain distributed practice. Instead, it is referenced to emphasize the roles of explicit retrieval and retrieval difficulty as recurring topics for explaining how retrieval practice works. The ECA has arguably made the best use and specification of both, which can make it at least one of, if not the most, comprehensive takes on retrieval practice. It is noteworthy such ideas are recycled in different, or sometimes the same, forms as it might signal further development and research of the ideas is warranted (because a satisfactory version has not been obtained if others keep being proposed) or that a new approach might be needed.

Both the reminding effect or elaborative retrieval hypothesis have received mixed support for their claims, and should also be held accountable for their negative, albeit interesting, "side effects" that can impact learning. As briefly mentioned above, the former had the effect of strengthening the *reminded memory* and not the one doing the reminding. If the TE is the result of such reminding, then it might be a bit difficult to reconcile its positive results with such a potential flaw built into retrieval practice. If it is possible for retrieval practice, as a whole (i.e., after multiple blocks), to strengthen the presentations (i.e., original and practice), then that would make it difficult for an account such as the ECA to explain the TE. The occurrence of this strengthening at the expense of what would typically be a second, third, or later practice and should ultimately be a hindrance at some point to learning. Given the ECA relies on the accumulation of contextual associations created across practice trials, if reminding of the original even happens without any enhancement to subsequent presentations, then the building and influence of contextual variability would be limited. The implication of "asymmetrical" strengthening is that the original event is enhanced enough to provide, at least a considerable part, of the TE. That is, multiple practices, or in this case remindings, will strengthen the original study event. How likely that is remains unclear, especially in other "types" of practice conditions. Although one supporting example comes from the context dependence literature in which some studies show cue-target pairs (e.g., a face and a name) superimposed over contextual images, or videos, and later practice target retrieval upon the presentation of the cue and image. Smith and Handy (2014) did just that across five practices for two conditions, either (1) the same context video (along with the cue) was shown across all practices or (2) a different video was shown each practice. Performance, on a cued-recall test two days later, was best under the varied context condition. This result is partially consistent with the reminding effect. Based on previous

findings (Tullis, Benjamin, & Ross, 2014) it is posited memories of the presentations during practice should remain unchanged. Whether practices are under the same or different context the cue-target pair should be strengthened by the same amount, which was not the case given the final cued-recall test. One explanation for this result is that the difference between conditions was due to the reminding difficulty during practice. That is, if the same context is presented at every practice then the reminding of the original event is more likely and the recipient of a lesser enhancement. Whereas a different context is less likely to trigger the reminding of the original event, which would give a greater enhancement. If the difference between conditions is mainly the outcome of this difficulty, then the *reminding effect* accounts for this effect just as other frameworks that rely on retrieval difficulty. This limits its usefulness as it does not provide explanations distinguishable from other frameworks. The elaborative retrieval hypothesis has also shown effects that might not make it completely suitable to explain a boost in memory performance (i.e., TE). Because it hinges on the creation of semantic associates that later act as retrieval routes, it would be a significant shortcoming if for that exact reason performance suffers. Specifically, elaboration has been shown to be susceptible to the output of intrusions whereas retrieval practice does not show such an effect as found by Lehman, Smith, and Karpicke (2014). The reason for this interference was explained by the authors as an increase in the search set, or candidate memories searched to find a target, whereas retrieval practice enables its restriction and protects from the incorrect selection of non-targets. Furthermore, retrieval practice has shown similar protection against other forms of interference (e.g., retrieval-induced forgetting and output interference; Kliegl & Bäuml, 2016). This smaller search set has been explained as due to the retrieval of contextual information. That is, retrieval practice allows for the association of contextual elements and targets. Because each practice is (temporally) set apart

it is accompanied with a distinct set of contextual elements which can become associated with the studied material (i.e., targets). During final retrieval, such as a final recall test, probing for a given memory can be restricted to the set which are associated with its context. This implies that searching through a greater number of candidates would lead to longer search times as a retrieval attempt has to cycle through more memories. Similarly, search times can be reduced if the search set is smaller. Indeed, reduced search times have been found with retrieval practice relative to elaboration (Lehman, Smith, & Karpicke, 2014). But, why would elaboration show lower performance and higher latencies if it also depends on a form of context? In an elaboration task semantic associates may be activated which, in essence, would serve as contextual information. Why does such context not help with performance? The answer could be in the act of retrieval, or rather its practice. As discussed above, if elaboration (i.e., concept mapping) is paired with retrieval practice then performance is comparable as when only engaging in retrieval practice (Blunt & Karpicke, 2014). This suggests that despite any contextual elements that may be encoded during elaboration, without an explicit (retrieval) practice activity, they might not be of much help. This means only creating and associating contextual information may not be as effective a learning strategy if not compounded with retrieval. This latter activity was lacking in the conditions mentioned above where elaboration was compared to retrieval practice. That is, they were comparisons between activities using either elaboration or practice. Naturally, such exclusivity between them was needed to draw a clear distinction and determine which account is better at explaining the TE. That elaboration, when coupled with retrieval, is just as effective emphasizes the key role retrieval plays during practice. The question now becomes, what it is about retrieval that makes it so conducive to learning?

While the singular act of retrieval may not fully explain the TE (despite its possible role as a memory modifier) it is proposed it plays a complementary role along with other factors. Namely, the retrieval of context has received considerable notice (e.g., Bäuml, 2019; Polyn, Norman, & Kahana, 2009; Lohnas, Polyn, & Kahana, 2011; Mensik & Raaijmakers, 1989) as it has been consistent with performance on various memory assessments (e.g., free recall, cued-recall). As already mentioned, context retrieval may help by restricting the number of memories that are searched, but it may yet have a more expansive role. The ECA could be considered among the foremost frameworks that rely on context and provides an appealing explanation for the TE. This is due to the amount of evidence it has accrued in its favor as well as its use of existing concepts that it is able to incorporate and give a coherent account of how they create the TE.

3. EPISODIC CONTEXT ACCOUNT

So far, the repetition, and up to a point insistence, that a single account fully explain the TE may seem unreasonable as there are potentially too many variations to retrieval practice. Each variation (e.g., the inclusion of corrective feedback, length between practice intervals or retention intervals, and number of practice trials) can alter the magnitude, and even the presence, of the TE. A framework making such an ambitious attempt must be put through rigorous testing for an adequate evaluation. Presently, this is the case for the ECA. Its assumptions cover factors ranging from the backdrop of retrieval practice, or the conditions in which it occurs, to the selection of a memory for output. For all its breadth it has received a considerable amount of support, but nonetheless there are still a few blind spots that need addressing. That is, a couple of its assumptions could use further testing to make sure they are justifiably considered to play a role in retrieval practice. Overall, the four assumptions of the ECA are rather simple and can be quite convincing, but if not fully stated they would not entirely reveal their own assumptions and research gaps. Each of its assumptions merits a discussion in turn.

The first assumption states an event's temporal context slowly progresses and its elements can become available for encoding (Estes, 1955; Howard & Kahana, 2002). In general, context is assumed to fluctuate at a slow pace so that contextual elements at two given moments will have greater overlap the closer they are to each other. The further apart any two points are then they will have less shared elements. This means that stimuli encoded at the beginning of an event may become associated with a set of contextual elements that may differ if the same stimuli are presented a second time at a later point in the event. This first assumption begins to introduce the concept of contextual variability, which will only take on a more important role and will be discussed in due time. For now, further clarification is needed to better understand

what is meant by context and how it may become associated with a given stimulus. The mention of temporal context as part of the assumption does not make it very clear which type of context is meant. There can be quite a lot of different types of context. For instance, temporal context can include the physical surrounding of an event (e.g., furniture, decorations, location), the semantic context of a tasks (e.g., being in a history class), mental or internal context (e.g., moods and thoughts), and peripheral information within a task (e.g., the font and color of words presented within a lab experiment, (Smith, 2013). Thus, the temporal context of an event could have multiple types of context. Which is/are the context(s) that become associated with a stimulus? The vagueness of the assumption in regard to context could be considered both a strength and a weakness. That is, by not specifying a context the ECA can generalize to better accommodate the different circumstances in which retrieval practice may happen. As noted, there can be a significant amount of variations between retrieval practice studies, and among them, context could be considerably different. Indeed, as we will later discuss, multiple studies have given their practices contexts that affect the TE. On the other hand, the lack of specificity can be a shortcoming as contexts may differ not only in type, but also in their saliency, encoding, retrieval, and overall influence. Aside from the broad categorization listed above, context could be further divided into a taxonomy which adds to the complexities already noted. Bjork and Richardson-Klavehn (1989) created a taxonomy of context which includes type, relationships, and processing as its three main dimensions. Context type refers to how intrinsic it is relative to a stimulus. For example, the font in which a word list is presented would be intraitem context, whereas the furniture in which the list is learned would be extraitem context. The relationships between a stimulus and context can be incidental, integral, or influential. Most relevant to the current discussion is that of incidental and integral context. An example of the latter can be an

extraitem integral context in which a context element, such as an object, is intentionally associated with a stimulus. An incidental context would be an element not explicitly encoded or associated along with a stimulus. Lastly, the third dimension is that of data driven or conceptually driven processing at test. Because each level of a dimension can be paired with each of the levels in all dimensions explaining each combination will lead us a bit off topic (but for examples of each see Bjork & Richardson-Klavehn, 1989). The reference to the taxonomy is made to (1) show how complex the role of context can actually be, (2) suggest more research on the context of retrieval practice is warranted as it covers a sizeable area, and (3) show the ECA does address key issues about context, as we will shortly see. Given the large variety of contexts, it appears as if the generalizability of the ECA could be considered a shortcoming. But, despite the great number of contexts the ECA does narrow which will actually have an impact on performance. Namely, it considers intentionally encoded context as the most influential and beneficial for retrieval practice. In contrast, incidental context is not expected to help as much. The explicit encoding of context better assures the creation of a more direct association, which can then be used as a retrieval route. But, to further complicate the issue, this advantage of intentional encoding may not always be present or of the same magnitude. If context is salient enough, such as if it is video or picture presented along a to-be-learned word, then incidental encoding could be enough to help learning (Smith & Handy, 2014, 2016). Overall, the topic of context is one much too broad and dynamic for a single framework to encompass. The approach the ECA appears to take is a limited one, but useful. By emphasizing the intention of encoding it narrows the scope of what eventually will become associated.

More often than not, temporal context refers to the links made between items or interitem associations. For example, in a word list memorization task, each word presented singly would

(hopefully) receive the full attention of the learner. As words switch from one to the next there begins to form a temporal context. That is, the temporal context of a given word is made up of the adjacent words on the list (Howard & Kahana, 2002; Lohnas, Polyn, & Kahana, 2011; Howard, Jing, Rao, Provyn, & Datey, 2009). The suggestion of interitem associations as context is accurate and for the purposes of retrieval practice it is adequate. This is especially the case when one remembers retrieval practice is used as a learning tool applied to material, such as study or class notes, that can be used in a similar manner. Although, given that the ECA does explicitly state its assumption relies on temporal context, without any further specification, it is not unreasonable to test the generalizability of this term. The testing would not merely be academic, rather in finding the boundaries of the ECA and retrieval practice both may be modified and refined to better suit their respective goals. As a preview of the experiments presented below, a specific type of context will be used for all present experiments, namely pictorial context.

The second assumption involves the use and self-reinstatement of context during retrieval as way to facilitate the remembrance of a target. This strategy is attempted if the context cues available during retrieval are not useful enough. That is, the contextual elements available at retrieval can be examined to determine whether they could be used as effective cues. If they are then memory is probed for stored contextual memories related to the target. If contextual elements are not helpful, then the process of contextual self-reinstatement is attempted to recall the context in which a target was encoded. This reinstatement strategy basically serves as an indirect target retrieval route. As a brief note, henceforth the term self-reinstatement will refer to this process of recalling a context, whereas the more general term reinstatement will be used when referring to any form in which an event's context is brought to attention. For example, it

will to refer to either self-reinstatement or the *experimental* re-presentation of context, as when an earlier seen or encoded context is fully shown a second time as part of an experiment. Self-reinstatement also affords the benefit triggering other memories that share the same context. As such, context can be an effective and economical cue, in so far as it may be associated with various targets. As mentioned, context is posited to fluctuate slowly and, as a possible consequence, may overarch the presentation of various stimuli which would share the same context. Hypothetically, using a context cue, be it provided externally or internally, should make it more likely that stimuli learned in close proximity should be retrieved successively as they would be activated by the same cue. Evidence for this kind of retrieval pattern is found when measuring output clustering and the probability of recalling adjacently learned stimuli. While both basically refer to the same concept (i.e., sequential output of stimuli sharing a characteristic, such as temporal context) their separate mention is meant to convey the different ways clustering is calculated. Output clustering has been observed for stimuli based on their temporal (Polyn, Norman, & Kahana, 2009), semantic, and even spatial context (Miller, Lazarus, Polyn, & Kahana, 2013). In the case of semantic clustering, the effect has been observed when learned stimuli are output based on categorical membership. But, this output pattern emerged only when task instructions called for the semantic assessment of the stimuli (i.e., pleasantness ratings, Whiffen & Karpicke, 2017). Similarly, the same study found greater temporal clustering when task instructions was a list discrimination task. These results are noteworthy for their dependence on task instructions, which is reminiscent of Bjork and Richardson-Klavehn's (1989) taxonomy emphasizing the role of process (albeit in testing). Without explicitly focusing on a contextual element while still in, what may be termed, an acquisition phase clustering did not occur. A lack of low grouped output was observed on restudy conditions as opposed to list discrimination and

pleasantness ratings. The ECA's specification of intentional encoding of context is applicable in this case, that is, without paying attention to an aspect of context (whatever the kind) it might not be used at retrieval. Nonetheless, at the risk of nitpicking, in a possible attempt to generalize it risks glossing over relevant factors by mostly focusing on intentionality. The results from studies showing clustering oddly support the ECA's assumption while possibly pointing out a potential flaw, namely the lack mention about process. Still, such findings are mostly a win for the ECA as they support its prioritization of intentionality. It should be noted that a form of clustering, the probability of retrieving items presented close in time to each other, has been found with tasks that do not explicitly require the encoding of contextual elements. Studies in which word lists are learned, by presenting each word singly, show that once a word is retrieved at testing the probability of outputting neighboring words decreases the further away or the greater the lag between them. This conditional retrieval has also come up when the encoding task is simply to study the words without another task requiring the encoding of context (e.g., Kahana, Howard, Zaromb, & Wingfield, 2002). This use of interitem context shows may show that other, more conventionally considered contextual elements, need not be used to facilitate recall. Further support for this benefit from context is from transitive associations. That is, if two "overlapping" cue-target word pairs, presented at different temporal contexts or time points, (e.g., shredder – common and common – ribbon) are encoded it is possible to retrieve words across pairs due to their shared context by way of their overlapping word. For example, in using the cue shredder to recall common this target may trigger the retrieval of ribbon from the second pair (Howard et al., 2009). In sum, for our purposes the use context, as specified by the ECA's second assumption, in this manner is reasonably justified. In a simple retrieval practice design the encoding phase would consist of the initial exposure or study of, say, a word list. This first phase would provide

a temporal context that could be retrieved on a later practice to help recall, for example. Afterwards, the same could be done on a final memory test in which the original, and practice, context may be used. If the context, either at practice or test, is similar to the original (or practice) context then it may be used as a cue, if not, then self-reinstatement could help retrieval. In instances when temporal context is used a retrieval guide, it would be expected to see a clustering of answer (on a final test) based on the context of a phase or block. Lehman, Smith, and Karpicke's (2014) study had participants learn multiple lists and their retrieval practice condition displayed more same list transitions during a final test. This clustering was taken as a sign the temporal context of a given list was mentally reinstated by participants and used to search for "nearby" or same list items. Each list was different with its own temporal context in this example, were the same list practiced multiple times (i.e., in multiple contexts) then the idea of context updating, or the next assumption, comes into view.

The third assumption deals with contextual associations made across retrieval attempts. Upon the retrieval of a stimulus its original context becomes associated or updated with the context in which retrieval occurs (Figure 2). This updating has the effect of creating a composite of the temporal context elements belonging to each of the contexts. Such an account of contextual updating has been proposed by some memory models (e.g., Siegel & Kahana, 2014; Lohnas, Polyn, & Kahana, 2015). Context updating can amount to the increasing of contextual variability, which can be considered a limited variant of encoding variability. Just as encoding variability would predict that its increasing should lead to greater number of retrieval routes, and hence higher chances of successful retrieval, so does contextual variability is said to boost retrieval. Increasing contextual variability can make it more likely that some context elements available at retrieval will match those from a target's context and can serve as cues. Variability

can increase if events are spaced apart or temporally distant, letting context fluctuate, so they can acquire more unique elements. For example, the advantage of spaced practice over massed practice can be interpreted as due to contextual variability. The multiple spaced practices give the opportunity to update and bolster context variability to help retrieval at test. The simplicity of this account on updating is both an asset, as it is supported by the benefit of spaced practice, and yet another instance in which the ECA does not fill in details. Again, it is possible it does not attempt to do so as it might intend to position itself as explaining how retrieval practice works in a general way. Nonetheless, to not specify how this updating takes place with, say, different types of context, or how the composite is formed, can be construed as a shortcoming. Because the ECA basically relies on the existing literature (memory models) that address context updating it seems to avoid trying to detail that information. But, for the most part those accounts have also not provided much information on how it may apply to other types of contexts or the interplay between contexts elements as they form a composite. Perhaps it would be a bit unfair to lay the burden of specification solely on the ECA, but to include the assumption of context updating is to invite its questioning. As it is, aside from the updating of temporal context there is not much work that looks into either of the two issues raised above. Some of the present experiments described later are a start to investigating, at least, the first of these issues. The selection of pictorial context is suited for the experimental manipulation of (contextual) variability as the number, type, and duration of contexts can be predetermined enabling enough control to assess the effect of variability.

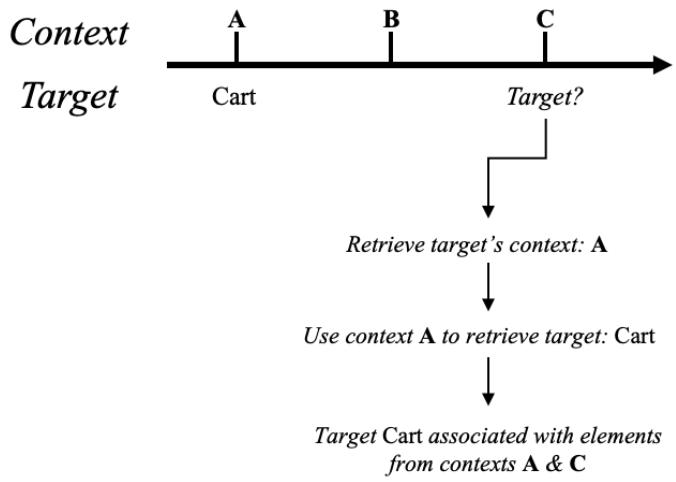

Figure 2. Context Retrieval and Updating. To assist target retrieval, its original context may be retrieved and, if successful, may become associated with elements from the current context. As displayed in this image, retrieving the word Cart is aided by the retrieval of context A, which becomes associated with elements present at context C (i.e., an association between the original and retrieval contexts).

As the fourth assumption, the ECA states that retrieval benefits from a restricted search set. This set allows a more efficient probing of memory due to the contextual updating described in the previous assumption. For instance, after the initial study and practice(s) of a given target it may become associated with multiple contexts (i.e., that of study and practice). On a final test, target retrieval is limited to only memories associated with a specific combination of contexts. The updating of context in effect acts as a filter that can restrict which memories are searched. In limiting the search pool, retrieval can become more efficient as there are fewer candidates amongst which to select for output. The use of a restricted search set, and the efficient retrieval that should follow, can manifest in reduced response times. Lehman et al., (2014) showed a quicker arrival at asymptote for the cumulative recall in a retrieval practice condition relative to

control and an elaboration condition. Similar results were reported by Roediger and Karpicke (2006a) showing more (practice) testing reduced the time needed to asymptote.

As important as this fourth assumption may be for the ECA it arguably, out of the latter three assumptions, contributes to the TE the least. Where the assumptions for context retrieval and updating (assumptions two and three, respectively) address the core of which and how stimuli is encoded, the fourth relates how retrieval is eased as a consequence of the former two. The first assumption sets the conditions in which retrieval practice happens, that is, it posits the fluctuation and availability of context which enables the rest of the assumptions. Because the first and last assumptions basically let the other two "carry the load" of learning it would not be unreasonable to focus on them as the main sources of the TE. Indeed, the present experiments will take the middle assumptions and test the retrieval and updating of context. Again, not to dismiss the first and fourth assumptions as they do add cohesion to the ECA, but assumptions on context retrieval and updating are the two that could be the most susceptible to changes in a retrieval practice paradigm. That is, context is basically assumed to fluctuate and even though the pace might vary from event to event its basic idea of change is the same. Likewise, a restricted search set could be assumed to mostly work in the same way regardless of alterations to how retrieval practice is done. But, if we change the stimuli, the type of context, the number of practices, or the availability of updating, which are variables not uncommonly manipulated in studies, then the middle assumptions take on a greater role. Because these two assumptions might be the most likely to experience different demands by the conditions of retrieval practice, the spotlight is mainly placed on them to explain the results from study to study. The ECA may rest most of its ability to generalize to the wide variety of retrieval practice studies on the assumptions of context retrieval and updating. While the other two assumptions might be the

most constant (across studies) and also in need of further investigation, to properly evaluate them would require dedicated research that would soon reveal they are more general concepts applicable to many more scenarios than just retrieval practice. Whereas the middle assumptions, while still applicable to other situations, are a bit more tailored for the conditions in retrieval practice. As such, if the purpose is to explain, as best as possible, how retrieval practice creates the TE then a starting spot for the endeavor would be to test the two specialized assumptions.

Interestingly, context retrieval and updating have been investigated as part of other (similar) literatures (e.g., Finn, 2017; Smith & Handy, 2016; Smith & Vela, 2001). Their work can be quite informative for the purposes of assessing the ECA and the TE. For instance, research on context dependence and memory updating could well guide and bring up issues potentially ignored up to now in retrieval practice. The insistence on the use of context from study to practice(s) to test does legitimately raise the question of whether, in some cases, context can be overly used and negatively impact learning. Also, the process of retrieval may not always have a positive effect on learning or updating of memories (Finn & Roediger, 2013; Davis & Chan, 2015). These issues will be discussed in the next section but are introduced to note that the present experiment can address the context retrieval and updating assumptions of the ECA as well as other, more general, questions about learning.

4. PRESENT EXPERIMENTS

There are two main questions being asked across six experiments. The first being, is contextual reinstatement is a desirable difficulty? The second, does context updating actually occur during retrieval practice? These questions are to test the second and third assumption of the ECA, respectively. The reason for this testing is two-fold, to asses retrieval practice on both practical and theoretical grounds. The benefit of the TE is consistently found and is, naturally, the basis for its recommendation as a learning tool/strategy (e.g., Larsen, Butler, & Roediger, 2008). Even the consistency of this benefit does not necessarily indicate retrieval practice should not be modified for the better. By identifying the key aspects (i.e., what contributes the most to the TE) and the redundant or unnecessary ones, then an optimal design (if possible) may be constructed. And as a brief speculation, the finding of this optimal design might depend on the material being learned, that is, different material might require slight modification to the design. But, retrieval practice has produced a TE with different types of stimuli (e.g., Johnson & Mayer, 2009). The studies with more ecological validity tend to be those that try to recreate the learning done in a school setting or aimed at a similar application in mind (Butler & Roediger, 2007). This means the TE has often been found with materials friendly to educational/school purposes. This suggests the generalizability of retrieval practice across material expands mostly to this category. Far from being a limitation, this further indicates its recommendation to students is warranted (Putnam, Sungkhasettee, Roediger, 2016). Regardless of design adjustment based on material, changes due to other issues can still take place. For instance, as already mentioned prior work did not seem to find any difference or benefit of expanded practice over equally spaced practice, unless practice intervals contained information that may cause interference. These results suggest a change of design based on how learning is planned, that is, if a student must

study for more than one class then he/she should plan their retrieval practice activities based on whether notes might interfere with one another (i.e., if the cover similar information). Aside from the main two questions posted above, the present experiments could help find out whether the type of context used therein also adds to the TE. Knowing whether it is actually beneficial, not advantageous, or has no influence could help change how practice is done. Lastly, the inclusion of a given aspect shown to benefit learning is a rather obvious decision to take as is the exclusion of one that hinders learning. But, factors which might not show any influence on learning might deceptively signal that their presence can be tolerated and not require any action as to their removal. The reason this could transform into a problem for learning is that studying/retrieval practice is not always conducted in ideal settings (e.g., a lab experiment). In a real-world scenario, the presence of multiple "neutral" factors might actually turn out to be a distraction. In reducing retrieval practice to its most basic components it can be constructed without much concern for including extra factors that may go on to negatively impact performance. On the one hand, the TE is already a robust finding with the existing practice designs, so there is not much need to change it. On the other hand, modifying the design could only help streamline it and this would not come at the cost of time-consuming exploratory experiments, rather it would come about naturally from theory driven studies.

4.1. Desirable Difficulties

From a theoretical perspective there are a few issues that could use further clarification as to their workings within retrieval practice. There are several factors, all held to occur within assumptions two and three, that are in need of testing, namely, the postulation context retrieval can be a desirable difficulty, the updating of context at retrieval, the use of a specific type of context (i.e., pictorial), and context dependence. The concept of desirable difficulties has been

criticized for not specifying a mechanism (Karpicke, Lehman, & Aue, 2014) for the enhancement memories are posited to receive due to difficulty. For all the supporting evidence ascribed to desirable difficulties (e.g., increased performance with spaced over massed practice) it is of limited utility without identifying what constitutes a difficulty. The ECA posits the degree to which context is reinstated, or rather self-reinstated, to be the desirable difficulty. Greater amounts of reinstatement are considered as a more difficult task which means under such circumstances it should confer a greater enhancement to memory. The less self-reinstatement needed, then the smaller the enhancement. For example, if a context were reinstated experimentally, that is, fully re-presented a second (or later) time then there would not be a need to mentally reinstate it. Interestingly, it is possible for such conditions to not just diminish the enhancement, but also be a hindrance to learning. A full reinstatement (i.e., no self-reinstatement needed) can "deceptively" show higher memory performance during practice(s) if the final test does not reinstate the same context. This could be termed deceptive because the learner might mistake his/her practice performance as accurately reflecting how well information is learned regardless of the presence or absence of any specific context. The context dependence literature has shown changing contexts between practice and final test can lead to worse performance on the latter as context is not able to be used a retrieval cue anymore. This is consistent with the proposal from the ECA, if there is no need to mentally reinstate context at practice, then memory is not enhanced as much as it could be and, in some cases (as in the aforementioned), may not help in strengthening target memories to the point of decontextualization. Granted, if a final test is administered under the same context as practice(s) then the lack of self-reinstatement might not actually hurt but improve performance. Still, it is often assumed that a final assessment (i.e., a test) will be conducted in a context different from which practice occurred and a goal in

learning is to be free of dependencies for the retrieval of knowledge. The notion of greater self-reinstatement as a desirable difficulty does appear consistent with the results of context dependence studies. But, a stricter interpretation of these findings could be made that, at least, on the lower end of self-reinstatement the results are consistent. That is, when there is no need to retrieve context memories are not strengthened, but what about when context is difficult to retrieve? The ECA receives some backing on this front from studies showing that learning, or practicing, under different contexts can help on a final test. These studies (e.g., Smith & Handy, 2014) show final test performance to be best when practices changed (pictorial) context from block to block. When retrieval of context was difficult or required greater self-reinstatement, then better retention on a final test emerged. Aside from a similar finding, another one of these studies showed other striking results in regard to the changing of context during practices. Schwoebel, Depperman, and Scott (2018) crossed context reinstatement (same *v*. different) and spacing (massed *v*. spaced) of practice using pictorial contexts. The authors reported main effects for both context reinstatement and spacing, with different context and spaced practice as doing better than their respective counterparts. In both conditions in which practice is considered to be more difficult there was a greater boost to performance on a final test given after one week. It appears that desirable difficulties based on reinstatement difficulty does show promise as way to explain better performance after engaging in retrieval practice. Although, oddly enough the authors did not explicitly report any group comparisons. For example, the different context massed practice group did just as well on the test as the same context spaced practice group, and the different context spaced practice scored higher than both of those groups. At least, this is according to a graph included in the article. Perhaps the authors were simply not interested in what could be gleaned from such comparisons, especially given the topic of retrieval difficulty

were not part of the study's stated purpose. Or, they might have thought the graph was enough to convey those differences. Whatever the reason, as it is the article might have missed out on being able to offer a more conclusive answer as to retrieval difficulty. Specifically, the possibility of being able to compare a condition that compounds difficulties such as the difference in pictorial context and temporal context (i.e., spaced practice) against a condition of only one difficulty (same context spaced practice). An informal, and admittedly not satisfactory, comparison by way of inspecting their graph shows the former did score noticeably higher than the latter, suggesting an interesting idea. By combining the difficulty needed in the degree of self-reinstatement plus the difficulty in having to think back further in time (to a different temporal context) there seemed to be a greater enhancement as seen on the final test. Further research is needed to better assess which and how contexts are compounded while still maintaining a TE. The compounding of difficulties is not really a subject that has been discussed before, but one that certainly fits in well with the ECA. Overall, there is support for self-reinstatement as constituting a desirable difficulty, but even the study described above could be expanded upon to assess this issue in a slightly different way. As another preview of the present experiments, the effects of context retrieval difficulty at practice will be checked at an earlier stage in learning, while also trying to minimize the potential for the building of context dependency. In the study above, there was an extended acquisition phase (i.e., four retrieval practice blocks) in which trials were always shown with the same contexts. This could have had the effect of building memory strength before there was any need to engage in difficult retrieval, which would come in a later practice block. Plus, by the time this block began context dependency might have already been established. The present experiments will try to avoid these issues as will be explained later.

4.2. Context Updating

Furthermore, there is another issue still in need of more stable theoretical backing, that of context updating. The stance of the ECA is that this phenomenon happens at practice and makes a composite of contextual elements, old and new. It is not among the goals of the present experiments to rule out or support the updating contextual memories in all circumstances, let alone memories in general. Rather, these experiments attempt to test whether such updating occurs within the design of retrieval practice. In a typical practice trial, a cue might be presented for a limited time to retrieve its associated target. Immediately afterwards, although not always the case, corrective feedback will show both cue and target for a few seconds. In such a trial there would be the retrieval of a target (and context) and further encoding of cue-target pair at feedback. If in addition to this the encoding of new context is tacked on, then that might make it a task too many. The proximity in which these processes deploy might not seem like much, but they might end up interfering with each other so that learning might actually be impaired. This has been described as the impairment of learning by testing. This would make it unlikely for new context elements to be encoded and updated with an existing context memory, which runs counter to the third assumption of the ECA. It also counters the basic idea of retrieval practice conferring a benefit to learning due to testing, which makes this impairment all the more interesting. An explanation for this impairment is the *borrowed time hypothesis*. Davis and Chan (2015) formulated this hypothesis which states learners focus on relearning information seen as more important for an ongoing task/test at the expense of learning new information. Unsurprisingly, this shift in priority negatively impacts the latter's encoding quality. Furthermore, if a target were to not be retrieved, say, at practice, then the learner might expectedly devote most of the time to encoding the target during feedback (assuming it is given).

The same authors reported, within the same article, results supporting said hypothesis. Their basic design was to first encode cue-target pairs in the form of a face and a name, respectively. This was followed by a filler task, which was in turn followed by a practice test (i.e., name retrieval cued by a face) with feedback or restudying the face-name pairs. Critically, after each trial on both conditions, a profession was shown alongside the pair. This presentation was to be the associating of new information. After another filler task the final tests were given. The first test called for the retrieval of the profession and the second test for the name, both upon the presentation of a face cue. In two out three experiments that followed this design, they found an impairment for learning new information in the testing conditions. In regard to name retrieval, their first experiment (described above) found a TE, but for profession retrieval the restudy condition did better. Their second experiment, identical to their first except professions were encoded prior each testing or restudying trials, also found impaired learning for professions in the testing condition. Interestingly, there was no TE for name retrieval. When the authors reverted the design to the one in the first experiment with the key addition of more name feedback time that attenuated impaired new learning, although they found a non-significant trend in that direction (and a name retrieval TE, by the way). It was not until their fourth experiment in which testing and restudying blocks were separated from new learning that the impairment went away. That is, after either testing or restudying only the face-name pairs, another block was given in which the professions were to be learned. They found better performance for profession cued-recall in the testing condition relative to restudy, although there was a revered TE for names. To unpack these results, the authors explained them in terms of their hypothesis by stating that, in experiment one and two, participants in the testing condition were devoting their time to ensuring they learned the correct names. This means time was borrowed from learning

the professions to better learn the names, either when both name and profession was presented at the same time (Experiment 1) or when a name was presented at feedback which preceded the profession for the next trial (Experiment 2, i.e., borrow time from the next trial to rehearse the previous). When they finally gave participants more time for name learning by extending feedback time "borrowing" from new learning began to attenuate (Experiment 3). And, when the need to switch between learning different stimuli is gone, the impairment is reversed (Experiment 4). These results pose problems for the ECA as updating (or new learning) of context could be unlikely or difficult according to the *borrowed time hypothesis*. For example, on a given practice trial, and granting that a target was successfully retrieved by way of self-reinstating context, learners might continue to encode that target and not pay attention to the current context. Likewise, during feedback, they would spend time further encoding the target rather than associating new information or context. If attention is mainly focused on target stimuli and not context, then how is it supposed to be updated? This might be doubly difficult if context is not explicitly instructed to be encoded. That is, the incidental encoding of context might make it even less likely to be attended in favor of target stimuli. But, the learning context might not be as unlikely if specific types are available. For instance, the presence of pictorial context could make it easier to incidentally, or even intentionally, encode given it might be more salient. The present experiments below are poised to help assess such an issue, albeit not directly. Their use of pictorial context and explicit encoding instructions could let us know whether the context is foregone by learners to better learn the targets, as held by the *borrowed time hypothesis*. Or, whether the richness of the contexts might still allow some elements to be encoded, despite the focus of attention being on the targets, and updated. Aside from the issue regarding context type, retrieval practice may have a slightly more fundamental problem if time

borrowing does impinge on learning. Most retrieval practice design are strikingly similar to Davis and Chan's first experiment, with the exception of adding a new piece of information on each trial. This suggests a possible limitation to the benefits of retrieval practice in the case new information is attempted to be learned incrementally. The alternative would be to set apart, in a different block, the learning of the new content. But perhaps that might not be ideal, considering Davis and Chan found a reversed TE for names when this change to practice was made. Overall, this limitation on retrieval practice appears relevant the assumption of context updating as that technically qualifies as new learning. This is another instance where the present experiments can serve as starting point for further research into a peculiar aspect of retrieval practice.

4.3. Context Dependence

Yet a third issue, mentioned before, that is in need of closer inspection is that of context dependence. Even though the ECA does not directly mention it, its assumptions' strong reliance on context leaves the possibility of developing this "crutch" that may not be conducive to long-term retention. In the event context is repeatedly used as a cue, say, at practice, the learner become dependent on it as a retrieval route which may not be a good strategy if there are no other routes available. For example, learning the word *tiger* in context A and practicing (and using A to retrieve the target) in context B will, according to the ECA, become associated with a combination of context A and B. This means context A has received the most strengthening, from initial encoding and its use in practice, and context B has received only one encoding attempt. If context B was not encoded well, and it is a possibility given the *borrowed time hypothesis*, then it might not serve as an effective cue on test. This implies context A is depended on as the main retrieval route and if unavailable target retrieval through context self-reinstatement might not be successful. Context dependence studies, in which a single context is

overly used as a cue, consistently show its detrimental effect at testing if unavailable (Smith & Handy, 2014, 2016; Smith, Handy, Angello, & Manzano, 2014). The ECA does state that contextual *variability* is what contributes to the TE, so having more than one practice (as described above) would supposedly help encode other contexts (e.g., B, C, D). According to the ECA this should ease the target retrieval because of the larger number of possible retrieval routes, but does that necessarily evade the possibility of context dependence? Would not context dependence apply to the composite context memory? Smith and Handy (2016) provide helpful results in addressing this issue. In their experiment, they showed participants cue-target word pairs superimposed over a picture background during an initial study phase. Then they practiced retrieval over the targets for 1, 2, 3, 4, or 5 blocks. Plus, there were four other conditions to which participants were assigned, each differing as to the context displayed on each practice trial. One type of condition always reinstated the original study context across all practice blocks and a second type changed the context picture on every block. In total, there were six exposures to the same or different contexts. Another key manipulation was context relation to the cue-target pair. There were two types of relations, incidental or supportive. The former indicated a context that had no relation to the cue-target pair and the latter showed a related image. For example, the Tagalog - English pair tagaugit – pilot would be displayed with the context of a beach (incidental) or airport (supportive). Because context relation and reinstatement (i.e., same or different) were crossed in their design they had four context conditions which they assessed on a cued-recall test given with a blank background. The results indicated that after 5 practices, there was better performance on the test for those in the varied incidental context condition than in the other three groups. This is directly relevant to the ECA, as it supports the building of contextual variability as a key contributor to retention and ultimately, given enough practices as in this case,

allow for decontextualization of targets. Interestingly enough, that same condition performed among the lowest conditions, if not actually the lowest, when participants had 1-3 practice blocks. It was not until four, or more, practices were given that the condition started to perform better relative to the others. Up until that point, a "small" number of practices with constant (both varied and incidental for 2-3 practices) and varied supportive contexts (for 1-3 practices) seemed to be of greater help to final test performance. For up to three practices the repeated use of the same context, or at least one thematically related, worked best. This runs a bit counter the context dependence expectation as in two conditions practicing under the same context was ultimately better than switching contexts across practice blocks. Because contextual variability is a top factor in the ECA these results deliver a mixed record to its expectations. This is because two conditions with the least amount of context actually had better cued-recall than the condition with the greatest variability. Still, the supportive varied group contained a moderate amount of variability and performed just as well as the aforementioned conditions after two or three practices and did the best after only one practice. Given the manipulations of practice and context, the experiment is particularly placed to, somewhat, help address the issues of building context variability and context dependence. In regard to the former, the results do not give a clear answer as context variability appeared to sometimes give an advantage and at times comparable results to conditions with lesser amounts of variability. These two outcomes were dependent on the number of practice blocks. As it pertains to context dependence and retrieval practice, the results are also not entirely conclusive as the presence of constant context during practice paid similar or greater performance dividends relative to conditions where context varied. But, number of practice blocks was again a decisive factor. With four or five practices the advantage of constant context began to dissipate. This is quite consistent with the ECA's position on

context variability helping retrieval, but results showed this to be the case mostly after a considerable number of practices (i.e., 4-5). In terms of context dependence, a greater number of practices may have allowed the decontextualization of targets (Smith & Handy, 2014), which ultimately helped recall on the final test. Thus, a retrieval practice design, with constant and varied contexts, partially supports the superiority of context variability as a performance aid but mainly when using a protracted design. Schwoebel et al. (2018) showed converging results with a similar design as discussed above. This is in keeping with the ECA's stance that increasing context variability can eventually reduce dependence on context by decontextualizing memories.

4.4. Practice Blocks

The importance of number of practices on the effective building of context variability should not be ignored. With a few exceptions, context variability only showed itself as beneficial with a "large" amount of practices. Having practices in the low range actually benefited more with lower amounts of variability. The inclusion of more practices actually began to make context variability less of a factor on performance. Final test scores for all conditions (with low or high variability) was still relatively high when four or five practices were used. For both of those practice conditions performance was approximately in the range of 80-90%. The greatest difference (~10%) appeared between conditions after five practices, and between the varied incidental context group (i.e., greatest amount of variability) and the other three conditions that performed very similarly. While a ten percent increase is still a sizeable difference, this suggests even conditions with low contextual variability can still substantially benefit from retrieval practice, provided more than a few practices are given. This result is not entirely inconsistent with the ECA, because in the end more variability did confer a better advantage at test, but it does qualify the importance of contextual variability, which may have a lesser role as practices

increase. But it should also be emphasized the experiment did not report pairwise comparisons between groups, so this discussion remains informal and merely suggestive, which is reason to further investigate this effect of variability and practice as do the present experiments.

4.5. Context Specificity

Lastly, the issue of context type is both of practical and theoretical interest and still much in need of further research. The enriching of learning material with the use of pictures is not uncommon. For example, in learning a foreign language a given word can be presented with an image of the object or concept it represents. This image would now form, if not the primary, at least part of the foreign word's context. According to the ECA, presenting the French word for car, *voiture*, with images of different cars on every practice block should allow the images to act as retrieval routes. Likewise, it is not hard to image how this form of learning can easily be adapted to other types of materials/subjects that use images. Hence, retrieval practice research should closely look at how this type of context may act given the role assigned to it by ECA. As it is, the ECA mention of context is almost exclusively referred to as temporal context. As noted, the issue with this term is that it can encompass a large variety of contexts. The extent to which different types of context will be used in the same manner during retrieval practice remains unknown. But, at least in regard to pictorial context, the studies mentioned throughout this section help in addressing this issue. In particular, Smith and Handy (2016) showed that supportive contexts, in the long run, do not seem to provide as great a boost to performance as incidental (i.e., unrelated) contexts. As applied to our previous educational example, does this mean presenting different images of cars is not as beneficial as using unrelated images during practice? To briefly reiterate the conclusion above, the answer depends on the number of practices. Thus, at least for pictorial contexts, the manner in which it is used in retrieval practice

hinges on another variable and based on these results caution should be taken before trying to generalize to other context types. As such, different contexts types may exert different amounts of influence on the benefit of the TE. A practical implication is that teachers and learners should be aware of how, for example, images may impact vocabulary learning using retrieval practice. The present experiments will attempt to build on Smith and Handy's results by also using pictorial contexts, but this will do so in a more measured approach by using a slightly different way to control context variability (a fuller description is found below).

From a more theoretical standpoint, albeit still with practical import, the selection of a type of context would allow an experimenter to have greater control to, say, test a specific aspect of the ECA. In the present case, the selection of pictorial context enables quite a few conveniences. For example, to preview the materials used in the present experiments, they use images displaying scenes such as parking lots, buildings or streets. These scenes, while not enough to be distracting given their common locations, can be salient enough to attract attention and better ensure they were attended and encoded. Plus, as expanded on later, the scenes will be presented to participants with instructions to intentionally encode the contexts. Without the selection of pictorial context, and its intentional encoding, it would have been more difficult to try to identify which context elements were (possibly) attended by participants. Admittedly, this does not completely seal off the possibility that other types of context elements (e.g., room furniture, moods) are also encoded and used during practice, but it does help in reducing it as attention may be placed towards the task stimuli. This in turn will help assess the third assumption of the ECA, context updating. In narrowing down which kinds of elements are encoded, and thus available to update, it can be known if pictorial context can justifiably be included as part of the temporal context mentioned by the ECA. Furthermore, as different images

will delimit study and practice trials, that is, each cue-target stimulus is paired with one image, this also helps in eliminating the possibility that item-item associations (a type of context) are formed during the experiment. Because each trial will display an image unrelated to adjacent trials the change in context will be abrupt and provide enough information (i.e., a new image to encode) in need of attending to discourage forming stimuli associations across trials. Even though item-item associations can be technically considered context the aim of the present experiments is to test retrieval and updating as it pertains to context directly associated to a given stimulus (or stimuli pair such as a cue-target). The reason being that in the case of item-item associations, the retrieval of an item might trigger the memory of its adjacent neighbor(s), which combines both context and target retrieval. In pairing a single pictorial context with a cue-target stimulus the retrieval route of context is solely that of an image. In other words, the retrieval of context has its own dedicated route or association. In using this dedicated association the assumptions of context retrieval and updating can be tested more accurately. Likewise, by selecting pictorial context it could be argued that the results could extend to a different type of context, the one type images were meant as a stand-in, environmental context. To be clear, to conclude pictorial and environmental context have the same influence during retrieval practice would require further experimentation explicitly comparing both types. Arguably, if one type of context will be used to generalize to another, this might be a starting point, but one should proceed with caution as this might lead to an overestimation of the influence of environmental context on retrieval practice. At least in the context dependence literature, in which the effects of environmental context are researched, there is a rather weak effect of context (Smith & Vela, 2001). Granted this effect is in regard to the matching of context at study and test and does not necessarily include practice testing, so the effects might not be of the same magnitude in a

retrieval practice design. Nonetheless, this highlights the modest effect of environmental context on memory, which given the lack of similar studies within the retrieval practice literature it serves as helpful reference. It is worthwhile to consider the potential effects of environmental context considering retrieval practice is recommended as learning strategy/tool, one that can be implemented in many places. Students can study/practice retrieval in class, their dorm, school library, coffee shop, or home, so if the building of contextual variability does occurs as stated by the ECA, then it should be possible for it to do so in different environments. Further work is needed to investigate the influence of physical environment on retrieval practice and in the absence of such work the present experiments may help by using a substitute for such context. Granted, the adequacy to generalize from pictorial context to environmental context is debatable. A discussion on this subject, while relevant as it relates to the issue of context types, will ultimately lead too far off topic for the present purposes.

Overall, addressing these practical and theoretical issues would greatly benefit retrieval practice research by giving greater clarity as to its mechanism, such as by testing the assumptions of the ECA, and give more confidence over its efficacy as a learning strategy by incorporating only those aspects shown to contribute to the TE. The role of desirable difficulties is still not clear as a mechanism has not been proposed which limits its usefulness regardless of how convincing the idea might be due to how well it seemingly accounts for studies' results. Context updating across practices, or events, is yet another idea still in need of further testing. Whether such a process occurs during practice(s) *and* with different types of contexts, as encapsulated within the general term of temporal context, are worthwhile investigations because they inform several areas of research, such as memory updating, retrieval practice and its potential real-world applications. Likewise, context dependence and its potential for negatively impacting learning if

used as a crutch has not been fully examined within a retrieval practice design. Plus, the three aforementioned factors are associated to the amount of practice as their impact on learning may depend on it. A few studies discussed above are quite informative in this regard, but could use further research to expand their findings. Lastly, if the ECA is to be tested context specificity is needed as all the factors above may also depend on the encoding and retrieval of a given type. The present experiments are positioned to help address these issues albeit some more directly than others. Unfortunately, they cannot assess them all equally as that would naturally increase the scope of their two main goals, testing the context retrieval and updating assumptions of the ECA.

4.6. Present Questions

Because the context retrieval and updating assumptions detail factors that arguably drive most of the TE, they are herein taken as the most significant and, given the scarce support for the imputed role of desirable difficulties and context updating, the ones that are presently tested. As a quick recap, the second assumption of the ECA states context elements available at retrieval as used as cues to help search memory. These cues are used to recall a target's original context as a way to retrieve the target itself. If contextual elements, at retrieval, differ from those of the original context to the point they may not serve as effective cues, then self-reinstatement is said to take place. While the ECA does use this exact term, self-reinstatement is here used to refer to the mental reinstatement of context (the reason for this choice is explained above). This self-reinstatement is used in the same manner, that is, to guide the probing of memory. It is further posited by the ECA that the extent to which retrieval practice confers a benefit, or enhancement, to memory is based on the degree to which context is self-reinstated. To draw a finer distinction as to what is tested by the present experiment, the second assumption as whole is not the focus.

Rather, the present experiments basically assume that context is used a practice and what is being tested is whether that context retrieval constitutes a desirable difficulty. Namely, the greater the amount of context that needs to be self-reinstated (or retrieved) the greater the benefit to the memory of its associated target. As already introduced, the present experiments use pictorial contexts, specifically, images of common locations. This means that conditions that would require a greater amount of self-reinstatement would be those which show a different context or image than originally paired with a target. Whenever context is experimentally reinstated, or re-presented, the need to retrieve that same context is removed, which should confer a weak or no benefit. Testing this aspect of the context retrieval assumption with pictorial context in this manner is helpful as conditions varying reinstatement differ quite starkly. That is, the difference between conditions, in terms of required self-reinstatement, is that one will receive no cues to help that self-reinstatement whereas the other will not require any self-reinstatement of context. This will help compare two levels of self-reinstatement. It is expected that if it is a desirable difficulty, then conditions showing no context or a different one during practice should outperform the condition in which practice fully reinstates a target's original context. This would provide support for the ECA's second assumption and be quite a significant finding because such a result would show desirable difficulties to be operable with the mechanism proposed by the ECA. Of course, this would not mean desirable difficulties cannot yield the same results if other stimuli besides context are used. The ECA states that a difficulty by itself is not necessarily desirable, but that is the degree to which reinstatement is needed (Karpicke, Lehman, & Aue, 2014) that makes it so. For the present purposes, this is interpreted as the amount of reinstatement, such as the amount contextual information or elements that are retrieved. It could be worthwhile for future work to further dissociate difficulties in retrieval. For example, is

accessibility also a desirable difficulty? Suppressed or blocked memories are more difficult to retrieve, would retrieving them give a greater benefit too? The present experiments simplify this issue by assuming the accessibility of context is similar for all images and the difficulty arises from how much context is retrieved. Furthermore, it is important to note that a decision was made to not explicitly ask for the retrieval of context during practice or final test trials. While this may not seem conducive towards the goal of assessing context retrieval in those instances, including instructions for context retrieval, *at practice or test*, could have the unintended consequence of creating context dependence. If the instructions mention context could be used to retrieve targets, then when corrective feedback is given, displaying the target's context, participants may spend more time encoding the context and not the target (or its cue). This could make targets even more context dependent than they would be if presented under "neutral" context instructions (i.e., not emphasizing them as cues). This bias towards context dependence could result in bigger differences between self-reinstatement/context conditions. This would mainly impact the context (experimentally-)reinstated conditions, because according to the ECA it should have lower performance due to a lower demand to self-reinstate context. But, this lower performance would instead be the result of spending more time spent encoding and relying on context during practice, rather than a lack of enhancement due to the lack of context self-reinstatement. These results would suggest a stronger desirable difficulties effect when they would actually be due (in part) to context dependence. As such, no reference is made as to using context as a cue in the practice or test instructions. In sum, one of the primary goals for the present experiments is testing whether context retrieval during retrieval practice is a desirable difficulty. Again, if the ECA is correct, then it is expected that the greater the self-reinstatement needed the greater the benefit to memory on a final test. Results showing a lack of an effect

would suggest pictorial context retrieval may not be necessary to obtain the benefits of retrieval practice, that is, a TE. In the case of a reversal, this could suggest pictorial context might give a benefit not based on its self-reinstatement, but as a strengthened cue. That is, by presenting the context during practice(s) then participants may have more time to encode it further. Then it could be easier to remember during a final test, when it is not shown, and retrieve its associated target. While context self-reinstatement could still play a role, it might not be one consistent with desirable difficulties, rather one in which context needs to be learned well enough to act as a retrieval route. Technically, this result is still consistent with the second assumption of the ECA as context is being retrieved, at least on a final test, but desirable difficulties would not be part of the what makes retrieval practice effective.

The second main question addressed by the present experiments, as a whole, is that of context updating. This corresponds to the third assumption of the ECA, which states that upon the retrieval of a target it can become associated with both its original context and the context at retrieval. As the original context memory is now associated, or updated, with the newer context they are said to form a composite memory with elements from both contexts. In the present experiments, this updating should occur with the contexts shown during early acquisition phases (e.g., study) and practice. Because the experiments use pictorial context, updating would be the forming of associations between images, for example, between study and practice(s) images. This means that for updating to occur there will have to be at least two different images or contexts presented with a target. Some of the present experiments do not look into this issue of context updating and do not vary a given target's pictorial context. This is done by always pairing it with one image and/or with a blank background. Technically, in either of these instances it could be said that a target's context is still different, namely, its temporal context. By

simply presenting a target multiple times, spaced throughout the experiment, its temporal context will be different regardless of any accompanying image. While this is correct, it is expected that pictorial context will be associated strongly enough with targets that it should serve as their primary form of context. Given the richness, or detail, of the images they should be the first recourse for context use, especially considering the time constraints of trials which can make participants rely more on what comes to mind more easily, such as intentionally encoded pictorial elements rather than other forms temporal context that might be more difficult to retrieve. This is the reason for not expecting the updating of context in experiments that do not provide multiple pictorial contexts for each target. For the experiments that do provide such change, the opportunity for context updating is present. If the ECA is correct in its assumption of updating, then it is expected that conditions allowing for the creation of this composite memory will perform better on final test. The updating will provide an increase in contextual variability as elements from two distinct pictorial contexts should be present as part of the composite. A reversal of this outcome, that is, conditions with no opportunity to update doing better on a final test than conditions enabling updating, could signal that encoding a target in different pictorial contexts is a hindrance to their retention. This could be due to the demands of encoding a new context. That is, less time is spent on learning the target as attention slightly shifts to encode the new context. As a result, the target is not learned as well as if it were paired with a single context. This short account of this alternative outcome is addressed later on as part of the appropriate experiments. To be clear, either of these outcomes does not rule out the possibility updating may occur with other types of context. Similarly, the results would be more directly applicable to retrieval practice designs with lower amounts of contextual variability. Because the present experiments only use two different pictorial contexts for each target it is still possible for

updating to take place under conditions that allow a greater building of contextual variability. Using two contexts, at the most, for these experiments is potentially a limitation, but the decision to restrict this number was due to the intention to test the idea of context updating under conditions in which are not difficult. For example, switching contexts across multiple practices will make it harder to make a strong enough association between them and their targets. Because each context is shown only for a few seconds on each practice and multiple contexts are paired with a target, there might be less of an incentive from participants to encode the context. While this might actually have the beneficial effect of decontextualizing targets, which is ultimately desired, this result would have been obtained not because context was retrieved and updated, but because it started to be ignored and an emphasis might be placed on encoding the targets. As targets, and their cues, are what is constant across practice/restudy blocks participants may think of them as being more important to the purpose of the experiment than the changing background contexts. This shift in perceived importance can make participants prioritize target learning which has been associated with improved performance when using retrieval practice (at least when target learning was explicitly prioritized as in Davis and Chan, 2015). This might be a positive result for learning, but it is not what is being investigated in these experiments, rather it is the issue of whether context is updated during retrieval practice. As such, it was decided to not use a greater number of contexts per target as seen in Smith and Handy (2016) and Schwoebel, Depperman, and Scott (2018). While their results do support the idea that contextual variability does help retention, the demands of their practice blocks were such that they may have encouraged a shift in prioritization from the variety of contexts to the fixedness of targets. In using a more modest amount of contextual variability the present experiments hope to supplement those studies.

In summary, the assumptions, or at least aspects, of context retrieval and updating are the two main issues that will be tested. The assumption involving context retrieval will not be tested in its entirety, that is, it will not be tested whether context is retrieved or not. Rather, what will be assessed is the possibility of context retrieval being a desirable difficulty, which contributes to the TE. When referring to this aspect of context retrieval it will often be done as the context retrieval assumption, merely for the sake of simplicity. On the other hand, the testing of the assumption of context updating will look at its occurrence during retrieval practice. As previously noted, these experiments will also help assess other issues, such as context dependence and the number or practices used, but their discussion will be temporarily postponed until the relevant experiment(s).

Overall, there are six experiments which can be grouped into two sets of three. This grouping is based on design and thematic similarities. Naturally, all experiments share many of these, but their grouping or separation is done on the characteristics of practice. Some experiments will have more retrieval practice blocks than others. To further specify, experiments 1 – 3 include only one retrieval practice, while 4 – 6 increase this number to three practices. There are other aspects, among their similarities, that differentiate them, but this is the main one. As a preview, the increase in practice marks the change from investigating the assumptions at an "early" stage to that of their development after multiple practices. Including more practices gives an opportunity to increase contextual variability, helps recreate the studying conditions in which retrieval practice might be used by learners, and can allow a difficulty to become desirable. The first of these refers to the more favorable conditions for encoding a new pictorial context, which should then be part of the context composite memory created by updating. By presenting a second pictorial context across multiple practices it can be learned better as participants will have

more than one block to study it. Previous studies (Smith & Handy, 2016; Schwoebel, Depperman, & Scott, 2018) have used a "one-shot" style of learning for new contexts which would not be as appropriate for the current purposes for the reasons described above (regarding the number of contexts employed by the present experiments). Plus, as retrieval practice is meant to be a learning tool it is worthwhile to use multiple practices as learners often use repeatedly go over their material, either within or across study sessions. Testing both assumptions under this condition will help determine whether they could be helpful in an applied scenario. And, more practices can help assess the question of desirable difficulties at two different points, when the difficulty is supposedly beginning, as it would be with only one practice, and after several chances for it to develop into a desirable one. That is, on the first practice context retrieval might still be a bit difficult to do as at this point it might not be encoded well enough. But, after the corrective feedback from the first practice, which would display the pictorial context, and its attempt at context retrieval as part of the trials, this can become less difficult in the next practice blocks. This decrease in difficulty can ease context retrieval but still pose challenge. As such, these three issues act as the rationale for the increase in practices across experiments, which in turn is the basis for their grouping.

5. GENERAL METHOD

Because most of the present experiments have a similar design, this preliminary section will go over their shared phases to avoid repeating them as much as possible. The following phases are included in all experiments with the exception of Experiment 5. The fifth experiment has a streamlined design which will be described in its corresponding section. For Experiments 1, 2, 3, 4, and 6 the first three phases were the same (see Figure 3). The first phase had participants learn the pictorial contexts they saw throughout the experiment and had them create a one/two-word label for each. After all contexts had been presented and labeled the second phase had participants recall the labels they gave to each of the contexts. This phase was intended to help further learn the contexts and familiarize participants with the retrieval of context. The third phase was the initial presentation of the pictorial contexts shown with their cue-target pairs. That is, each context was presented with a face and a name. Participants were instructed to study this triad, or ensemble, of stimuli (i.e., context, face, and name). Then, the remaining phases for each experiment differed according to their subject of inquiry. After those intermediate phases, *all* experiments (1-6) had the same final cued-recall test. This test presented faces, superimposed over blank backgrounds, as cues for name retrieval. The order of presentation of stimuli in each phase was randomized in all experiments.

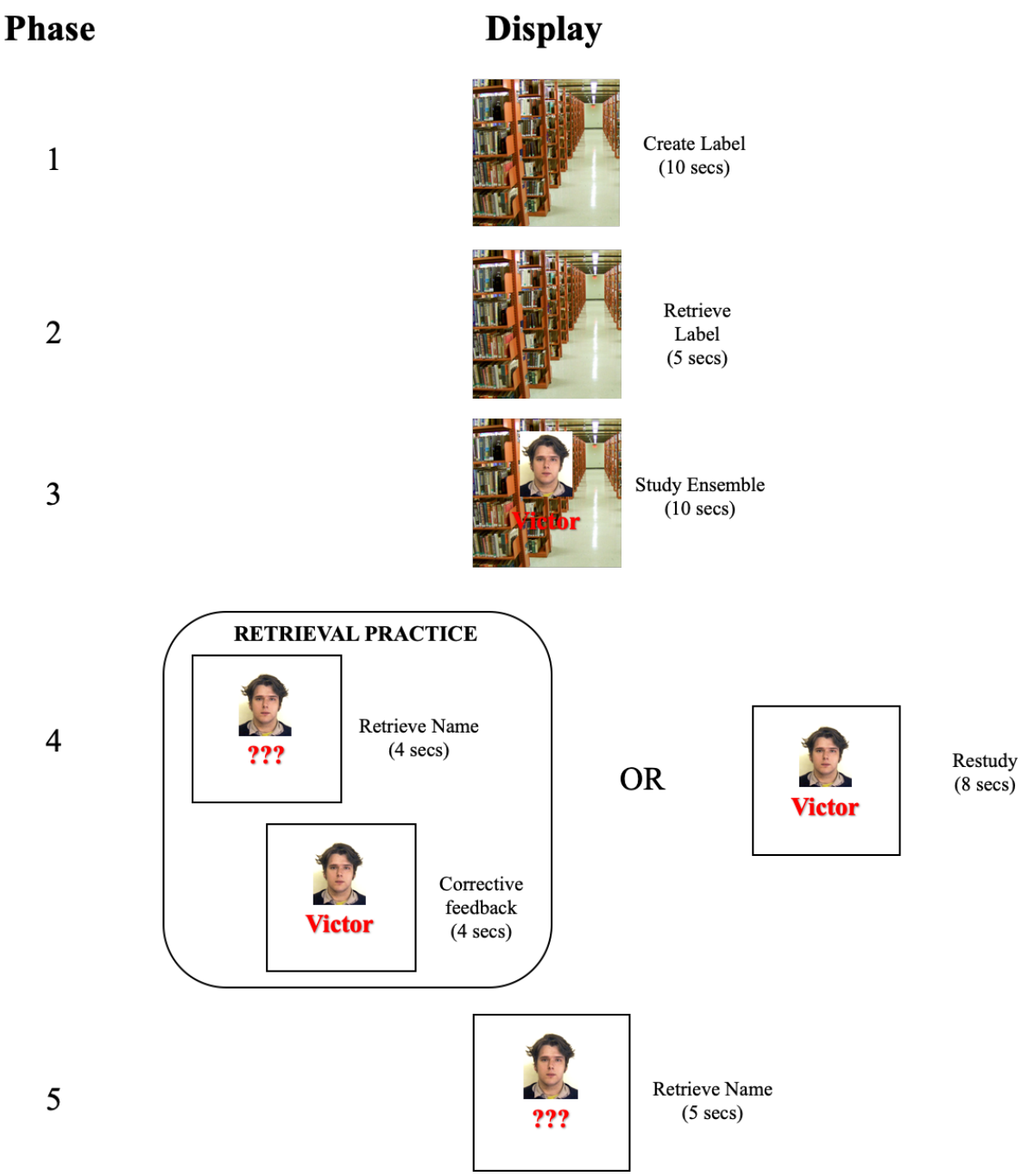

Figure 3. Phases of Experiment 1. Phases 1 and 2 consisted of creating and retrieving the label of images (i.e., contexts), respectively. Phase 3 was the study of all face-name-picture ensembles. Phase 4 was the restudy or retrieval practice of only the face-name pairs. Practice trials involved the retrieval of a name for a given face cue, followed by corrective feedback. Phase 5 was the final cued-recall test. Context images and faces selected from Konkle, Brady, Alvarez, and Olivia (2010) and the Psychological Image Collection at Stirling (http://pics.stir.ac.uk), respectively.

6. EXPERIMENT 1

This first experiment had "standard" retrieval practice design as it acted akin to a baseline for subsequent experiments and tested whether pictorial context encourages its retrieval (at practice and test) and helps produce a TE. The two conditions being compared were a restudy group and a retrieval practice group. Both groups received only one restudy or practice block, depending on their condition, after the initial encoding of the ensemble (i.e., context, face, and name), which corresponded to phase 3 in the previous section. Critical for this first experiment was the absence of context during the restudy and practice blocks (see Figure 3). Because this experiment was a baseline the lack of context during these blocks will let its results show the possible effects of previously associated context. That is, if context is retrieved during retrieval practice, then this condition should ultimately perform better than the restudy condition, which should be much less likely to encourage any form of retrieval because the face-name pair is fully presented. Because the first three phases (described above) not only introduce, but also help encode and retrieve context labels, plus associate it with the face-name pairs, by the time of the practice block participants in this group should be reasonably ready to be able to retrieve contexts. While it is possible for context to also be retrieved in the restudy condition this was not expected to be the case because the instructions to this condition only call for the restudying of the stimuli shown, that is, the face-name pair without any mention to their context. It was expected that if context is retrieved during practice, which should facilitate name retrieval, then this greater degree of self-reinstatement should confer better results on the final test relative to the restudy condition that does not require nor encourage any self-reinstatement. This prediction was based on the ECA, which states that context is retrieved during practice and this should contribute to the outcome of a TE. Both conditions have the same amount of contextual

variability, so their difference should only stem from the self-reinstatement of context, which the ECA states should occur during practice. Alternatively, it is possible to obtain a reversal of the TE. That is, the restudy condition would outperform the practice condition. A possible reason for this is the reduced amount of exposure to the target in the event that it is not initially retrieved during practice. If a name is not remembered, then the only opportunity to learn it comes from the corrective feedback given after each trial. The amount of feedback time might not be enough to learn the target compared to the restudy condition, which technically has more time to do so as will be described below. This TE reversal would indicate that attempting the retrieval of context may not actually confer a benefit that is greater than just restudying materials for just a *few* seconds longer.

6.1. Method

6.1.1. Participants

A total of 89 undergraduate students from Texas A&M University participated in this experiment. In exchange for their participation students received course credit. The experiment was conducted in sessions of up to 15 participants, the exact number of participants per session depended on their self-enrollment. The restudy condition contained 47 participants and the retrieval practice condition had 42 participants.

6.1.2. Design and Materials

The design of the experiment manipulated Study Condition (restudy *v.* retrieval practice) as a between-subjects variable. Correct performance on the final cued-recall test was the dependent variable. Furthermore, the retrieval practice condition was analyzed as a within-subjects variable to compare the practice block and the final test using correct performance as the dependent variable. The 20 pictorial contexts used in this experiment were taken from an

existing picture set (Konkle, Brady, Alvarez, & Olivia, 2010). These images depicted common locations, such as a parking lot, library, hallway, or a balcony. The cue-target pairs to be associated with each context were faces and names, respectively. A total of 20 male face pictures were taken from the Psychological Image Collection at Stirling (http://pics.stir.ac.uk). The 20 common male names served as the targets. All stimuli were presented on a projector screen and participants were to write their answers on sheets of paper.

6.1.3. Procedures

The first three, and last of the experiment's phases were introduced in the General Method section, but they will receive a final and detailed description here to reduce any further mention in subsequent experiments. The first phase displayed 20 context images one at a time for 10 sec. each. Participants were instructed to write on their answer sheet a one/two-word label about the context. Next, the second phase once again displayed the same context images but for 5 sec. each. On each of these trials participants had to write, on another answer sheet, the same label they had written in the previous phase. Then, for the third phase a trial consisted in displaying one of the contexts previously seen, a face, and a name. Each trial began with displaying the context for half a second before the onset of the face-name pair. The context image served a background over which a face was displayed centered on the screen with a name shown below it. Participants were instructed to study each ensemble for 10 sec. After all 20 ensembles were presented the next phase began. The fourth phase had participants either restudy the face-name pairs for 8 sec. or do practice retrieval. On this latter condition each trial would show a face cue with three question marks below it. Participants had four seconds to retrieve the name corresponding to the face and write it down on a new answer sheet. After each retrieval attempt, corrective feedback was given by displaying both face and name for 4 sec. Both

conditions had a blank background (i.e., “no” context) for all trials. Next, a maze completion filler task was given for 20 minutes. Lastly, the final cued-recall test consisted of showing a face cue (with ??? below it) and participants had 5 sec. to retrieve its corresponding name. These answers were written on yet another answer sheet. Participants were then debriefed and thanked for their participation.

6.2. Results

An independent samples t-test between retrieval practice and restudy conditions showed a significant difference in their final test performance, $t(87) = 2.361$, $p = .02$, Cohen’s $\delta = .50$ (Figure 4). A reversal of the TE was found as restudy ($M = .302$, $SD = .160$) performed better than the retrieval practice condition ($M = .229$, $SD = .136$). According to the ECA, self-reinstating context should have benefited the retrieval practice condition as the opportunities existed, at practice and test, to mentally reinstate the original contexts for the face-name pairs. This was not supported in the current results which indicate that practicing retrieval under conditions encouraging self-reinstatement are not as conducive to retention as plain restudying.

To examine learning from practice to the final test, a paired samples t-test was conducted for the retrieval practice condition, which showed a significant improvement across the phases, $t(41) = 3.783$, $p < .01$, Cohen’s $\delta = .65$. Correct performance improved from the practice block ($M = .141$, $SD = .134$) to the final test ($M = .229$, $SD = .136$). This increase in name recall is likely due to the corrective feedback given after practice trials. Despite this slight bump performance overall was low for both restudy and retrieval practice conditions, but especially so for the latter. Given that about 14% of names were retrieved at practice, arguing for a possible benefit from self-reinstatement would not be tenable.

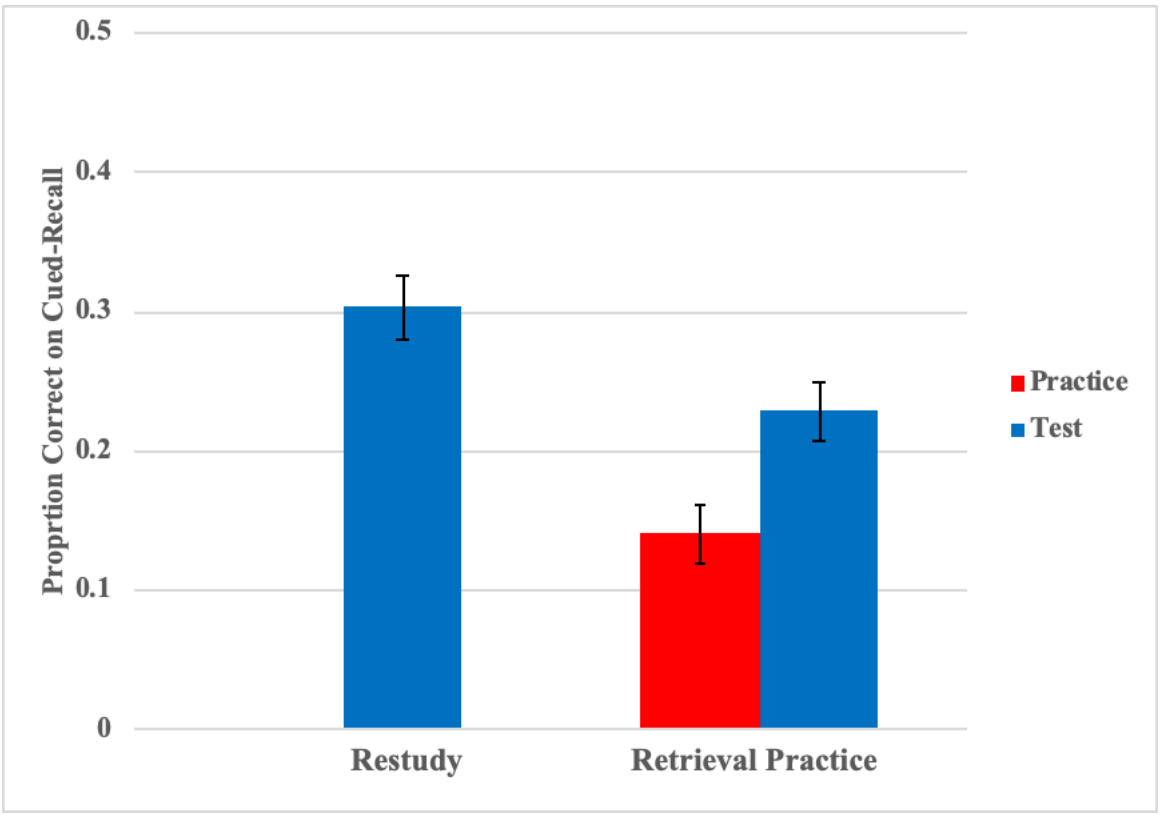

Figure 4. Experiment 1 Results. Restudy outperformed Retrieval Practice on the final test. The Retrieval Practice condition showed a significant improvement from practice to the final test. Error bars display standard errors.

6.3. Discussion

Overall, results do not suggest that self-reinstatement benefited retention on the final test. The reversed TE actually shows that any self-reinstatement that could have been done might have hurt performance, not just on the test but also during practice. If participants attempted to self-reinstate during practice trials, then the difficulty in such endeavor was too high and did not facilitate learning. Despite devoting the initial three phases to the encoding of context it still might have been too difficult to self-reinstate and try to retrieve the target names. These results, the test and practice, do not support the idea of self-reinstatement as a desirable difficulty even though the experiment had conditions which encouraged such a practice. That is, it used pictorial

contexts encoded multiple times, but this was not enough to make them retrieval routes facilitating name retrieval. It is still possible for context to have played a positive role towards learning in the restudy condition. Restudy instructions made no reference to the previously encoded context and only mentioned the face-name pairs shown needed to be studied, but it is possible for context to have been retrieved during these trials. Whether context retrieval at restudy was involuntary or intentional, this might have had the effect of allowing the full trial length to further encode the face-name pair while also remembering the context. However, while this covert or explicit context retrieval is possible, it is not considered as the sole explanation for the difference between the conditions. Technically, both conditions had the same amount of time to be able to retrieve context (i.e., 8 sec.). If this did occur both conditions had the opportunity to associate the (covertly or intentionally) retrieved context with the face-name pairs in restudy trials and during feedback in practice trials. Again, while this remains possible a more parsimonious explanation for the results is that self-reinstatement was too difficult to do, or just not attempted, and that restudy trials provided more encoding time for the face-name pairs. If self-reinstatement was not attempted it is natural to ask why. It could still be due to the difficulty in trying to self-reinstate or, more generally, context reinstatement is not strategy used at practice. Either possibility runs counter to the ECA. If too difficult, then self-reinstatement is not desirable and not beneficial. If not attempted, then the role of context is overestimated or misplaced. Both go against the second assumption of the ECA of context retrieval. The latter possibility strikes more critically against the assumption at it deals with context retrieval in general and would make the idea of self-reinstatement (as desirable) unnecessary.

Perhaps self-reinstatement, as encouraged in this experiment, might not be as conducive to retention as posited, but the use of context might still have an important part to play in the TE.

Because Experiment 1 was intended, and apparently succeeded, in creating difficult self-reinstatement conditions this does not rule out the possibility context still aids practice retrieval. Lowering the difficulty of practice trials can ease name retrieval by experimentally reinstating context. This should result in better practice and test performance by the retrieval practice condition, relative to Experiment 1. Thus, by considerably decreasing retrieval difficulty a *much* lower level of self-reinstatement is arguably present, but which might still confer a small benefit to retention. The following experiment made such an adjustment, among others, to the current design to further test self-reinstatement as a desirable difficulty, but at the low end of the difficulty spectrum.

7. EXPERIMENT 2

For Experiment 2, its design is practically identical to the previous experiment, but context is now experimentally reinstated during restudy and retrieval practice blocks. By incorporating context into these blocks this second experiment aims to better display a modulation of retrieval difficulty across conditions. Because the previous experiment failed to find a TE, and actually found its reversal, showing the original contexts should enable an easier retrieval of names for the retrieval practice condition. Because the restudy condition is also exposed to the original contexts the need for any retrieval from previous phases is minimized as much as possible. But, the retrieval practice condition should benefit from having context reinstated as it provides helpful cues which can be used for name retrieval, which means that practice should be easier relative to the same condition in Experiment 1. In removing context altogether from practice its trials might have been too difficult and name retrieval less likely, so that the opportunity to learn the name would be during feedback. Reinstating context should facilitate practice enough for it to now beneficial and result in a TE. Making trials a bit easier can help learning as names are more likely to be retrieved which gives, not only the obvious practice of retrieving that information, but also a greater amount of exposure time to the target (i.e., name) without having to rely solely on feedback. It is expected that if context reinstatement does aid learning then the outcome will show the retrieval practice condition outperform restudy. To find a reversal, as in Experiment 1, would suggest using context as cues during practice does not actually help, but might actually hinder learning as it might enable the building of context dependence. To address this possibility another variable was added to the design, namely, retention interval length. The final cued-recall test was administered immediately after the restudy/practice block or after 20 min. It is predicted a TE will be found for both retention

intervals if the benefit derived from reinstating context at practice would extend to the final test. This prediction differs a bit from what the ECA might predict which is that a condition in which retrieval is more difficult (Experiment 1) should do better on a final test than a similar condition where retrieval is easier (Experiment 2). Difficulty in this case refers to the degree of self-reinstatement which is greater in the former experiment. If a retrieval practice condition did not do well under conditions conducive to retention, why should it do well now that the same condition is easier, especially when context self-reinstatement is not needed? In other words, why should there be a TE *given the previous results*? If the ECA is not correct in its assumption that self-reinstating more context at practice is beneficial, then now that it is experimentally reinstated and eases name retrieval it can take on a supportive role of being a helpful cue. Finding a TE would provide counter evidence for context self-reinstatement as conducive to retention, at least as an informal cross-experiment comparison. Alternatively, results consistent with context dependence would be expected if a TE is found if the final test is given after a short retention interval and a reversed TE for the 20 mins condition. The same results as in Experiment 1 would be expected for the 20 mins interval condition, because the retrieval practice group would have used the reinstated context at practice to help their performance and after a "long" retention interval access to that context would have been more difficult. Whereas the restudy group would not grow dependent on context because it only has to further encode the ensemble without a need to make context a retrieval route. On the other hand, with immediate testing the retrieval practice group can benefit from the short interval as their main retrieval route is still accessible. In addition, their practice in retrieving the names, that is, becoming familiar with the test can serve as a benefit. Overall, the retrieval practice condition can outperform the restudy

condition due to these two factors. A higher performance with immediate testing, but lower after 20 minutes can signal the development of context dependence for the retrieval practice condition.

7.1. Method

7.1.1. Participants

A total of 277 participants were recruited from the same subject pool as in Experiment 1. Participants self-enrolled into experimental sessions that held groups of up to 15 members. The retrieval practice immediate and 20 min. testing groups each contained 69 participants. The restudy immediate and 20 min. testing conditions had 68 and 71 participants, respectively.

7.1.2. Design and Materials

The design for Experiment 2 is similar to that of the previous experiment with the addition of retention interval as between-subjects variable. Thus, the design is Study Condition (Restudy *v*. Retrieval Practice) X Retention Interval (Immediate *v*. 20 min.). The materials are identical to those from Experiment 1.

7.1.3. Procedures

Experiment 2 included the customary initial three phases (see General Method) after which the fourth phase was similar to that of Experiment 1 (Figure 5). Namely, participants would either restudy or practice retrieval, but now context was reinstated during these blocks. That is, the original context that accompanied a given face (or face-name pair, for restudy) would be presented in the background on each trial. The duration of these trials, that is, restudy, retrieval practice, and feedback were identical to those from Experiment 1, but during name retrieval and restudy the contexts were shown for .5 sec. before the onset of a face and face-name pair, respectively. Next, participants would either take the final cued-recall test (immediate

testing) or complete a 20 min. maze task and then take the test. The final test and maze task were as in Experiment 1.

7.2. Results

A factorial ANOVA using Study Condition (Practice *v*. Restudy) X Retention Interval (Immediate *v*. 20 min.) both as between-subjects variables and correct test performance as the dependent variable. There was not a significant main effect of Study Conditions, $F(1, 273) = .198$, $p = .657$, $\eta_p^2 = .001$, as both restudy ($M = .264$, $SD = .210$) and retrieval practice ($M = .275$, $SD = .200$) conditions performed about the same (Figure 6). On the other hand, there was a significant difference between Retention Interval conditions, $F(1, 273) = 12.932$, $p < .01$, $\eta_p^2 = .045$. As was to be expected, immediate testing ($M = .314$, $SD = .210$) resulted in higher scores than after 20 min. ($M = .226$, $SD = .191$). More importantly, there was no TE, or reversal, as restudy and practice conditions did not differ in either retention interval. An independent samples t-test revealed restudy ($M = .308$, $SD = .208$) and retrieval practice ($M = .319$, $SD = .214$) to have scored practically the same with immediate testing, $t(135) = .298$. $p = .766$. Likewise, there was no difference between restudy ($M = .221$, $SD = .204$) and practice ($M = .232$, $SD = .177$) after 20 min., $t(138) = .333$, $p = .740$. The lack of a benefit, for either studying condition, is rather striking now that context was reinstated. This is further lack of support for the use of context as a means to improve retention. Making retrieval easier by reinstating context at practice did not seem to impact test performance much as the practice condition from this experiment did about the same as its counterpart from Experiment 1 (i.e., retrieval practice condition with 20 min. retention interval).

The experimental reinstatement of context did help performance during practice, despite this benefit not extending to the final test (Figure 7). A paired samples t-test, for the immediate

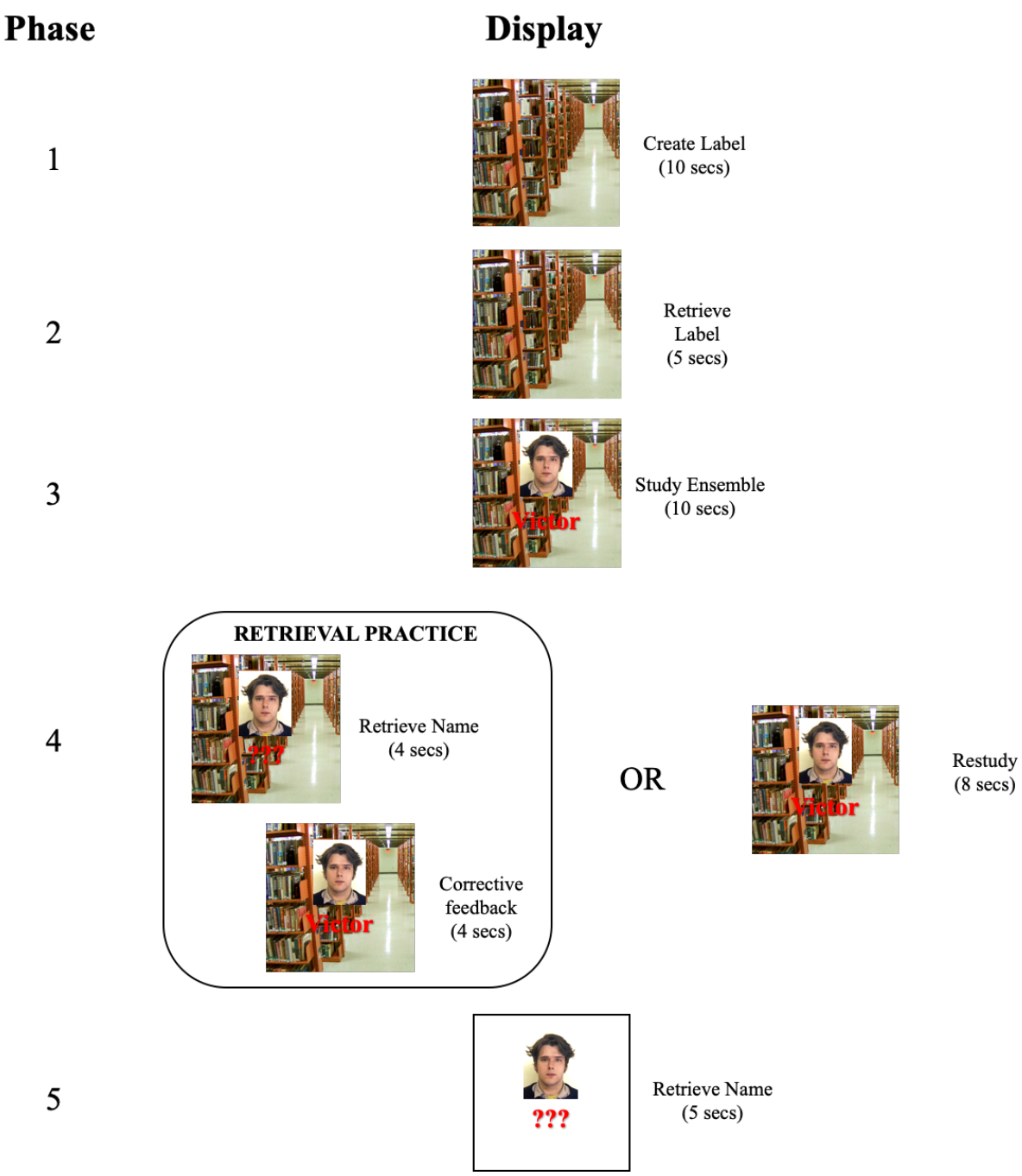

Figure 5. Phases of Experiment 2. Phases 1 and 2: the creation and retrieval of context labels, respectively. Phase 3 was the encoding of ensembles. Phase 4 was restudy or retrieval practice, which includes the experimental reinstatement of context. Phase 5 was the final cued-recall test. Context images and faces selected from Konkle et al., (2010) and the Psychological Image Collection at Stirling (http://pics.stir.ac.uk), respectively.

testing condition, showed a significant drop in performance from practice (M = .401, SD = .211) to test (M = .319, SD = .214), $t(68) = 2.715$, $p = .008$, Cohen's $\delta = .39$. Similarly, practice (M = .391, SD = .205) performance was higher than on the final test (M = .232, SD = .177) after a 20 min. retention interval, $t(68) = 5.074$, $p < .01$, Cohen's $\delta = .83$. This drop in performance is consistent with the use of context as a crutch, that is, the development of context dependence. The boost on cued-recall during practice ultimately did not translate to a benefit on the final test. These results highlight the possibility that context may not necessarily have a positive role in retrieval practice.

7.3. Discussion

The second experiment yielded another set of results that are not consistent with the idea of context use as a beneficial act in retrieval practice. The experimental reinstatement of context may have lowered the difficulty of practice trials, thereby helping name retrieval, but this was not a desirable difficulty, albeit it would have been a small one, as it did not end in better retention compared to restudy. Indeed, in neither of the retention interval conditions was practice any different from restudy. These are yet other results not supporting the ECA. Granted, this time the evidence against the ECA is not as strong as, say, from the previous experiment. The reason is that in Experiment 1, practice conditions were more ideal for obtaining a benefit from self-reinstatement as a desirable difficulty. But, in the present experiment the experimental reinstatement of context would have conferred a small boost to retention because there is little difficulty in reinstating context, so it is not too surprising that retrieval practice did not outperform restudy. At least regarding the issue of desirable difficulties these results are not necessarily a strong blow against the ECA. Although, from a more general perspective the presence of context might have helped practice performance, so it was used as a relatively

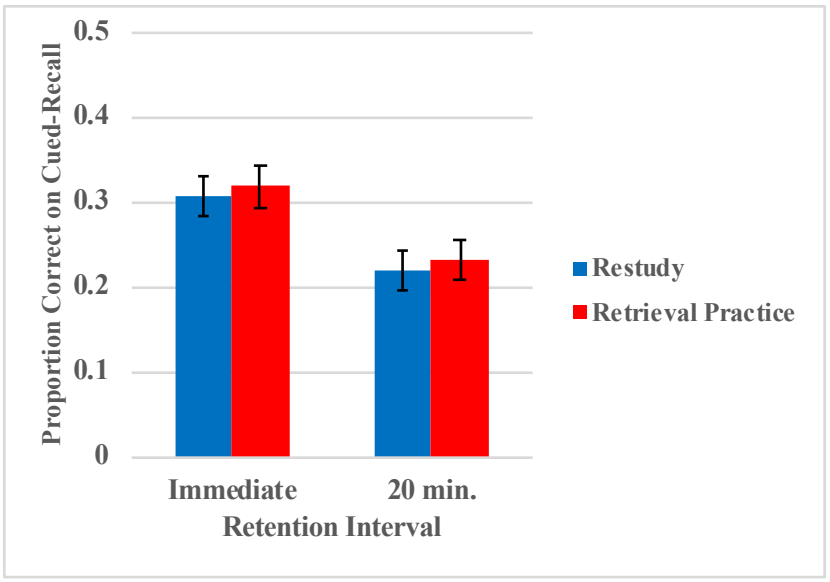

Figure 6. Experiment 2 Final Test Results. Both Retention Interval conditions did not display a TE. Overall, performance was higher if tested immediately than after 20 min. Error bars display standard errors.

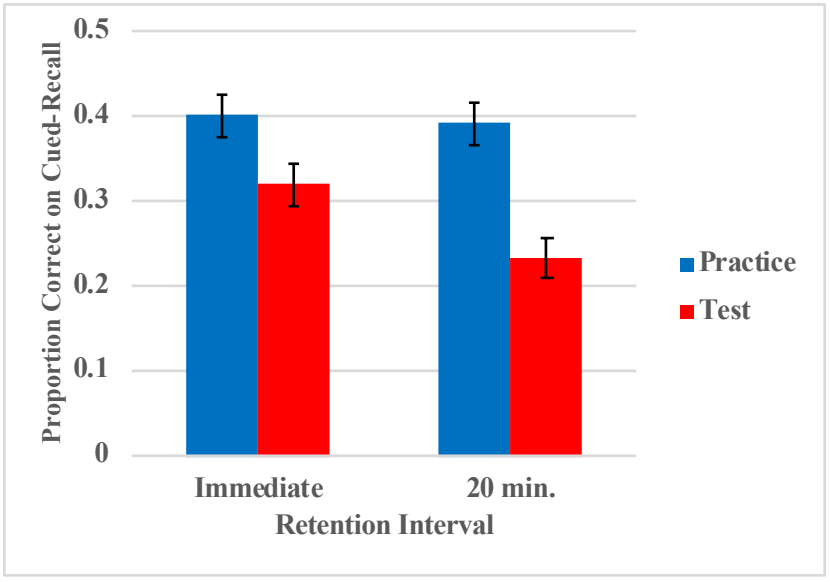

Figure 7. Experiment 2 Retrieval Practice Results. Practice conditions, in both Retention Intervals, showed a significant drop in performance from practice to the final test. Error bars display standard errors.

effective cue. This is consistent with the ECA positing context available at retrieval is used to help search memory. But, the reduction in performance on the test suggests despite having yet another phase in which context could be encoded (i.e., the initial three plus restudy/practice block) self-reinstatement during the final test still did not seem to help the retrieval practice condition. An informal comparison between this experiment's retrieval practice condition and the one from Experiment 1 suggests that after 20 min. performance hovers at around twenty percent. That is, two conditions differing only in the presentation of context at practice end up performing the same. This indicates that varying the amount of self-reinstatement needed at practice does not have an effect on retention as suggested by the idea of desirable difficulties as described by the ECA. In addition, the presence of context cues might actually foster the building of context dependence as suggested by the decrease in performance from practice to test. Support for this conclusion also comes from the restudy condition. In the present experiment, the (20 min.) restudy condition scored at about twenty percent, while its counterpart in Experiment 1 scored higher. Again, as these conditions belong to different experiments an informal comparison will have to suffice to cautiously suggest reinstating context in restudy trials might negatively impact retention. The precise reason for this decrease is not known, but as restudy trials now included another stimulus to encode (i.e., context) this could have taken away time from the encoding of face-name pairs. Likewise, the presence of an oft seen image may lead participants to mistake its fluency for how well the ensemble is known. This can reduce the effort made towards encoding the ensemble resulting in lower cued-recall. Overall, the second assumption of the ECA has not received support from the first couple of experiments. Based on these results, retrieval practice does not benefit from self-reinstatement, or even using context as a cue, at least when employing one context. The ECA's context retrieval assumption might not

be the strongest factor in determining the TE, rather updating or contextual variability might be a more critical factor. The creation of a composite memory with various elements from multiple contexts has not been assessed in the first experiments. As such, context might still play a considerable role in the TE provided its influence is not solely dependent on one contextual memory, that might be difficult to retrieve to boot. That is, contextual variability might be the more important factor in the TE.

8. EXPERIMENT 3

The first couple experiments have not shown evidence of context, be it through its possible self-reinstatement (Experiment 1) or overt use as a cue (Experiment 2) at practice, aiding final test performance. The results from Experiment 2 actually suggest using reinstated context can enable context dependence. The presentation of the original context may have encouraged an over reliance on it, but this does not completely rule out the possibility context can yet have a positive impact on retention. The effectiveness of using the original context as a practice cue may have been too strong of an incentive for participants (Experiment 2) and there was no need to either self-reinstate or have an opportunity for updating, as the context was the same. That is, the second experiment did not employ conditions as conducive to the development of the assumptions of context self-reinstatement and updating. The "presence" of both context retrieval and updating may be needed to obtain a TE. For the first two experiment, practice conditions have only allowed varying degrees of self-reinstatement (basically all or none), but no updating. To incorporate contextual updating, while maintaining the possibility for self-reinstatement, Experiment 3 modified the design of Experiment 2 by showing new contexts during practice and restudy. On each practice trial, participants can still attempt to self-reinstate context, as original context has been seen repeatedly up to that point and should be rather memorable, while also being able to update that memory with the new pictorial context. In limiting the amount of contextual variability to only two contexts the potential for decontextualization can be diminished, otherwise final test performance will reflect this phenomenon instead of the use of context. Also, increasing context variability in this manner provides two richly detailed contexts to encourage using them as retrieval routes at test. As such, the present experiment tests whether contextual updating, with the possibility for context self-

reinstatement, helps create a TE. Accordingly, a TE is expected if practicing retrieval under different context allows updating which can then benefit test performance. The addition of new context at practice will help test the updating assumption of the ECA. Alternatively, the presence of new context may hurt test performance for the practice condition, which would result in a reversed TE. As in Experiment 1, the absence of the original context might make it difficult to retrieve the names at practice, making feedback the only time the name can be associated with the new context as that is when the "new" ensemble would be shown. The short amount of time given for feedback (4 sec.) might not be enough to create an association that is useful at test, so contextual variability might not be a sufficient reason for explaining the TE, at least for similar retrieval practice designs.

8.1. Method

8.1.1. Participants

A total of 159 undergraduates from Texas A&M University participated in this experiment in exchange for course credit. The experimental sessions and self-enrollment were identical to those from the previous experiments. The retrieval practice condition had 79 participants and the restudy condition had 80.

8.1.2. Design and Materials

The design and materials for this experiment were the same as in Experiment 1, plus an additional 20 images were selected from the same source as the original set of twenty images (Konkle et al., 2010). Study condition was a between-subjects variables with two groups, restudy and retrieval practice, and correct performance as a dependent variable.

8.1.3. Procedure

This experiment proceeded exactly as did Experiment 1 with one exception, that of including context during the practice and restudy blocks. In this third experiment, the contexts shown during these blocks were new contexts not shown during any of the previous phases (Figure 8). Each face-name pair was now superimposed over a new image, so that each pair was shown with two pictorial contexts in total, the first during the first three phases and the second during practice or restudy. Retrieval practice and restudy trials were the same as in Experiment 2 (which also had context at trials). After practice and restudy blocks a maze completion task was given, which lasted 20 min. Then, the final cued-recall test was administered.

8.2. Results

An independent samples t-test, using correct test performance as the dependent variable, comparing retrieval practice and restudy did not show a significant difference between conditions, $t(157) = 1.126$, $p = .262$ (Figure 9). Although, results were in the same direction as in Experiment 1, with the restudy conditions performing slightly better ($M = .240$, $SD = .213$) than the practice condition ($M = .206$, $SD = .168$) in this present experiment. At best, results suggest a trend towards a reversed TE. Overall, both study conditions performed similarly on a final cued-recall test after practicing or restudying under conditions encouraging the building of contextual variability and updating. To state it in terms more appropriate to the current purposes, the comparable performance between conditions does not support the updating assumption of the ECA as practicing retrieval under conditions conducive to context updating did not result in better retention.

To check performance across practice and the final test a paired samples t-test was conducted on the retrieval practice condition. The dependent variable was correct performance

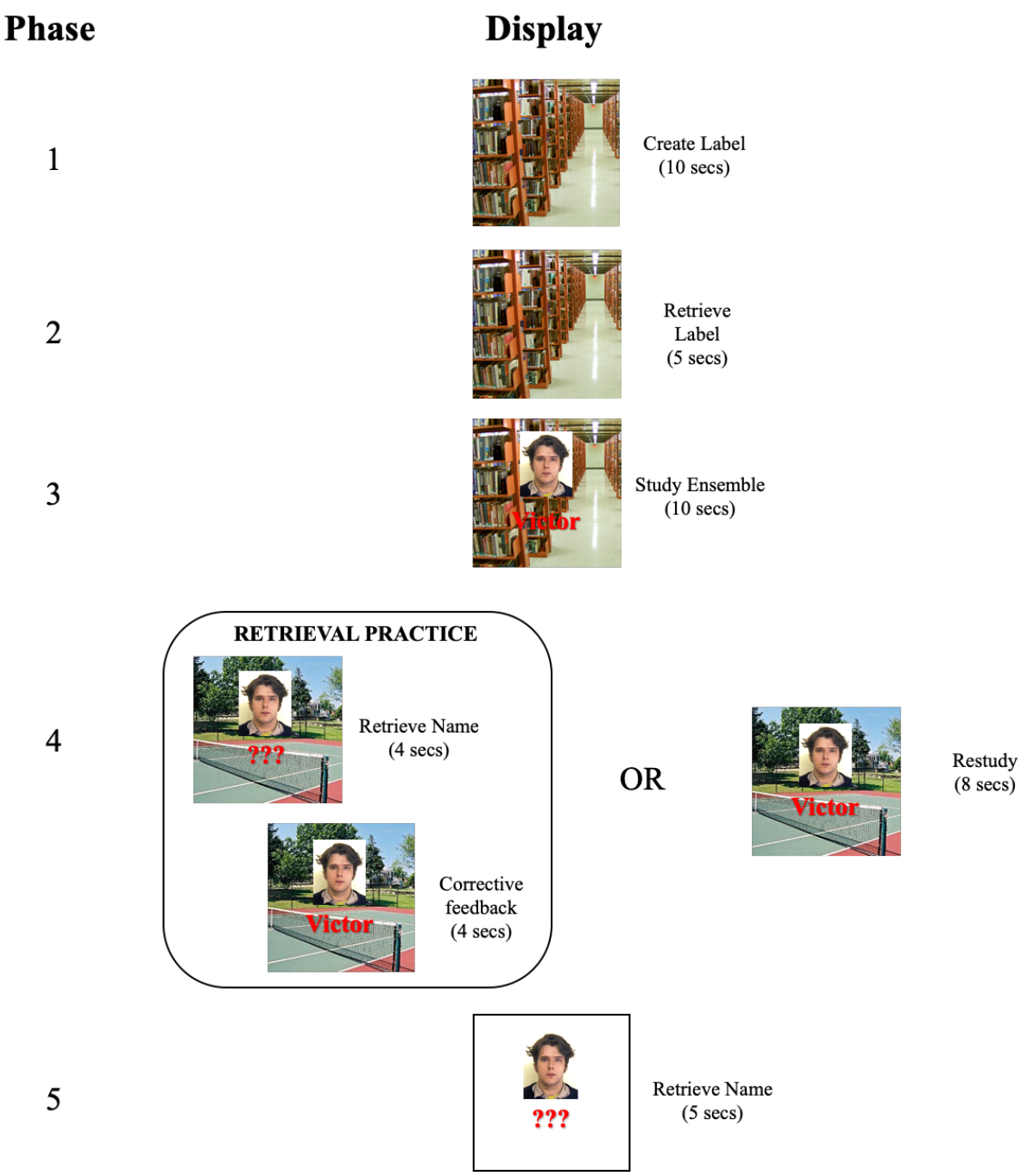

Figure 8. Phases of Experiment 3. Phases 1 – 3 involved the creation of context labels, label retrieval, and ensemble study, respectively. Phase 4 was the restudy or retrieval practice of names for a given face cue with a new context. Phase 5 was the final cued-recall test. Context images and faces selected from Konkle et al., (2010) and the Psychological Image Collection at Stirling (http://pics.stir.ac.uk), respectively.

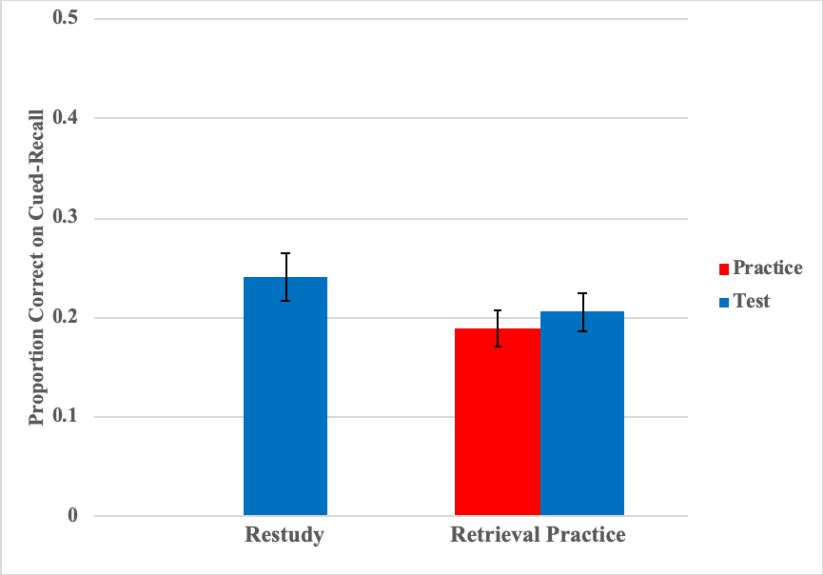

Figure 9. Experiment 3 Results. Restudy performed slightly better than Retrieval Practice on the final test. Retrieval Practice did not show a significant improvement from practice to the final test. Error bars display standard errors.

at practice and test. Interestingly, there was no significant improvement from practice (M = .189, SD = .164) to the test (M = .206, SD = .168), $t(78)$ = .808, p = .422. The inclusion of new contexts during practice did not benefit final test performance, which can be interpreted as the building of (low) amounts of contextual variability, or even updating, not having a beneficial effect on retention.

8.3. Discussion

For the third straight experiment there has not been supporting evidence for context retrieval as a desirable difficulty or, a bit more pertinent to the present experiment, context updating as factors contributing to the TE. The results suggest practice updating, even if it is just one pictorial context, does not assist in improving retention relative to restudy. This third experiment only had participants encode one context per face-name pair, in addition to its "original" context. Because pictorial contexts are richly detailed having only two is, presumably,

sufficient enough to create a composite memory of various elements from each context. Participants in the practice condition might have had, arguably, less time to encode the new context than participants doing restudying. Because the first part of practice trials (i.e., name retrieval) could be primarily spent on reinstating context (at least according to the ECA) or attempting a more direct retrieval of the name (i.e., using only the face cue), they might have not been able to encode the new context well. This means corrective feedback would have been where most of its encoding might have occurred, which is half the amount of time (4 sec) as participants had in the restudy group (8 sec). This lack of encoding time could have interfered with the successful updating of context as there was an insufficient amount of time to do so. In other words, updating might have not happened as tested under these conditions. This could be the reason why retrieval practice yielded a similar percentage score (low 20s) on the test as in the previous experiments that only used one context per face-name pair. Interestingly, the inclusion of new contexts at restudy could have been a negative effect on retention. As alerted before, comparing across experiments is merely an attempt at a helpful, informal comparison between similar conditions and the upcoming discussion stemming from it should be taken with caution. There has been lower performance for restudy conditions in experiments which presented context on restudy trials (Experiments 2 and 3). The highest a restudy condition has scored has been in the absence of it during restudy trials (Experiment 1). It is not difficult to conclude that having less information, or stimuli, to encode on these trials will allow better encoding for what is shown. That is, having pictorial contexts at restudy takes away time from the encoding of the material needed for test, face-name pairs. While not a groundbreaking conclusion, it nonetheless points out that encoding contextual elements during restudy, or practice, will divert from encoding other stimuli, which suggests that trying to explain the TE through the use of context

might be problematic for the ECA. Ultimately, paying attention to *pictorial* context could actually interfere with learning regardless of studying method. In sum, the beneficial effects of any composite, contextually updated memory absent from these results raises quite a few issues for the ECA. For one, its assumptions might not be compatible with pictorial contexts. Plus, so far experiments had employed only one block of restudying or practice which have not returned supporting evidence for the assumptions of context retrieval and updating. But, as both of these assumptions can benefit from repetitions it is possible that one block is not enough to elicit the effects proposed by the ECA. With multiple attempts, or blocks, the difficulty in self-reinstatement might become desirable. Likewise, repeated exposures to a new context might ensure successful updating.

9. EXPERIMENTS 4-6

Across three experiments, a TE has not been found and most results have indicated restudy to be the better studying method relative to practice conditions enabling context self-reinstatement and updating. That is, there has not been evidence supporting either of the assumptions of the ECA. Experiment 1 and 2, which modified difficulty in self-reinstatement, and Experiment 3, which enabled context updating (self-reinstatement), showed a "reduced" practice schedule to not be an effective learning condition. In using only one practice block, it is possible for the ensembles to not be well learned. Overall test performance, for both practice and restudy conditions, was low across all three experiments suggesting only one block is an insufficient amount of acquisition time. Increasing the number of blocks not only better mimics real-world studying behavior but can also better exploit any benefit from retrieval practice. As it is, context self-reinstatement and updating may not be beneficial under such conditions. Self-reinstatement is not yet a desirable difficulty as retrieval, of either context or target, is too difficult. Similarly, updating might have not been beneficial because the second context was not learned well enough, that is, the creation of contextual variability might have not been successful. Self-reinstatement might need to develop into a *desirable* difficulty and updating needs more trials to adequately encode new context and build variability. In the remaining experiment the number of practice and restudy blocks were increase to three, instead of one. Experiments 4 and 6 were similar to the previous experiments with the addition of two extra practice/restudy blocks. In total, there were three consecutive practice or restudy blocks. Experiment 5 had a modified, streamlined design with the removal of the initial two phases common to the rest of the experiments, but still had three practice/restudy blocks. Finally, the final test was taken immediately after the final practice/restudy block. In sum, the next

experiments follow up testing the two assumptions of the ECA in extended designs as self-reinstatement (as a desirable difficulty) and updating might require an iterative learning process to produce the TE.

10. EXPERIMENT 4

In Experiment 4 both context self-reinstatement, as a desirable difficulty, and context updating were present in practice conditions. This experiment had two retrieval practice conditions that varied the amount of context self-reinstatement needed and also provided the opportunity for updating. During each practice trial, one of these conditions fully reinstated the original context and the other did not show any context (i.e., a blank background). Then, for both conditions, corrective feedback with a new context was displayed after each trial. That is, the retrieval of the name (as the face cue is shown) was under conditions requiring either no self-reinstatement or allowing for self-reinstatement. And, feedback allowed for contextual variability and updating. To rephrase it a bit, one condition (i.e., no context at name retrieval) had conditions that should have allowed both assumptions of the ECA to take place and the other only encouraged the development of updating (i.e., reinstatement of context). The comparison between these two practice conditions will help determine how important self-reinstatement is to the TE. Plus, a restudy condition is included to test whether either practice condition obtains a TE. This final condition will act as a sort of baseline to see whether a TE is obtained by either of the practice conditions. If context self-reinstatement and updating needed several practices to develop as a desirable difficulty and contextual variability, respectively, then it is expected that the practice condition with both factors should outperform the other two conditions, restudy and practice with experimental reinstatement. But, if self-reinstatement is an encumbering task then the condition removing this need by presenting the original context will perform better than the other two conditions. Given that previous work consistently shows a TE when multiple practices are included (e.g., Butler, 2010; Roediger & Karpicke, 2006b), a reversed TE is not expected. Overall, a TE is expected if the ECA is supported in regard to both assumptions, or even one.

10.1. Method

10.1.1. Participants

A total of 186 undergraduate students from Texas A&M University participated in this experiment for course credit. The size of the experimental sessions was determined by self-enrollment with a maximum of 15 participants per session. The restudy condition had 62 participants, while the context experimentally reinstated condition and the no context condition had 64 and 60 participants, respectively.

10.1.2. Design and Materials

The design was the similar to the first set of experiment with Study Condition (Restudy *v*. Practice with context *v*. Practice with no context) as a between-subjects variable. Correct performance was the dependent variable for analyses of retrieval practice and final test. The materials were the same as those from Experiment 3.

10.1.3. Procedures

This experiment also included the initial three phases (see General Method) after which three consecutive practice or restudy blocks were given. These blocks were done in a similar fashion as the ones from the first experiments, but with a few minor modifications (Figure 10). In both practice conditions, participants had 5 sec. to retrieve the name of the presented face cue. The face was superimposed over a blank background or its original context. Then, in both conditions corrective feedback with a new context was given for 5 sec. Restudy trials were 5 sec. each and all face-name pairs were shown with a new context. Immediately after the practice/restudy phases participants were given the same final cued-recall test as in the previous experiments.

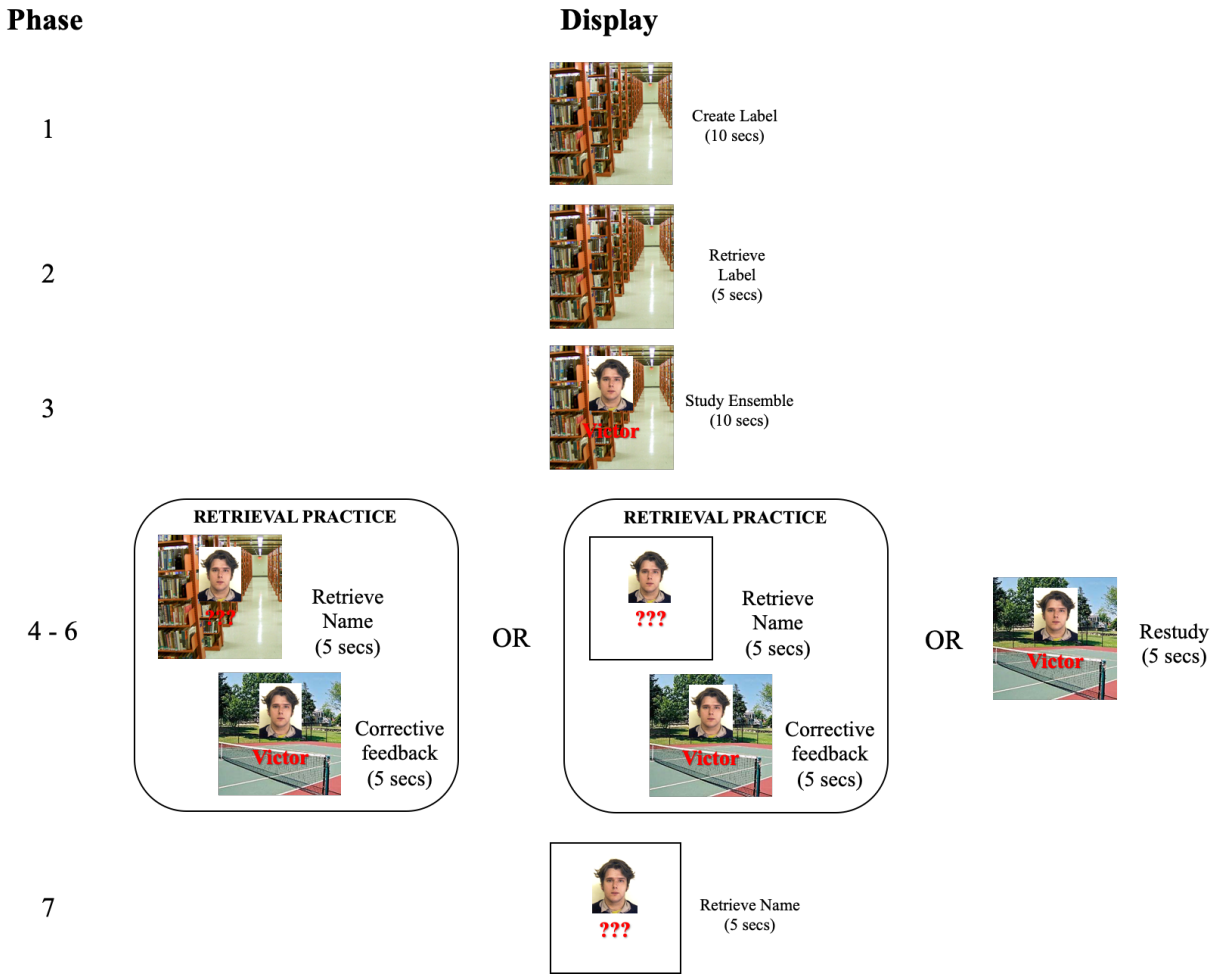

Figure 10. Phases of Experiment 4. Phases 1 – 2 included the creation and retrieval of context labels, respectively. Phase 3 was the study of the ensembles. Phases 4 – 6 involved the restudy (right) or retrieval practice of target names. One practice condition experimentally reinstated a given face's context (left) and the second practice condition did not show any context (middle). Both practice conditions showed a new context at feedback. The final task was a cued-recall test. Context images and faces selected from Konkle et al., (2010) and the Psychological Image Collection at Stirling (http://pics.stir.ac.uk), respectively.

10.2. Results

A one-way ANOVA was conducted using Study Condition (Restudy *v*. Practice with context *v*. Practice with no context) and correct test performance as a between-subjects variables and dependent variable, respectively. The analysis yielded a significant main effect of study

condition, $F(2, 183) = 13.141$, $p < .01$, $\eta_p^2 = .126$ (Figure 11). Comparisons between groups, using independent samples t-tests using the same dependent variable as above, resulted in restudy ($M = .386$, $SD = .232$) significantly underperforming practice with context ($M = .558$, $SD = .223$), $t(124) = 4.231$, $p < .01$, Cohen's $\delta = .75$, and practice with no context ($M = .588$, $SD = .253$), $t(120) = 4.604$, $p < .01$, Cohen's $\delta = .83$. Thus, for the first time a TE has been found once multiple blocks are built into the structure of studying. An additional result of note is the parity in performance between both practice conditions, $t(122) = .714$, $p = .477$. Specifically, conditions varying the amount of self-reinstatement during practice did not differ in their retention for the final test. This does not support the idea that the degree of self-reinstatement at retrieval confers a boost to later performance. Taken singly, these results do not support the desirable difficulties facet of the context retrieval assumption of the ECA. Practicing retrieval under the favorable conditions in which context is experimentally reinstated showed the same amount of retention as doing retrieval under purportedly more difficult conditions (no context shown).

To test the trajectory of learning across all practices and final test a mixed ANOVA was conducted with Study Condition (Practice with context *v*. Practice with no context) as a between-subjects variable and Phase (Practice #1, #2, #3, and test) as a within-subjects variable. The dependent variable used was correct performance. Because the assumption sphericity was violated, the following results are reported using the Greenhouse-Geisser correction. Results show a significant main effect of phase, $F(1.813, 221.184) = 97.551$, $p < .01$, $\eta_p^2 = .444$, and interaction, $F(1.813, 221.184) = 16.619$, $p < .01$, $\eta_p^2 = .120$. Plus, there was a main effect of study condition, $F(1, 122) = 17.152$, $p < .01$, $\eta_p^2 = .123$. The significant effect of phase was due to the continuous increase in performance practice after practice. This improvement continued

from the last practice block to the final test for the practice with no context condition, but not when practice included context (Figure 11). The latter condition actually had a significant drop

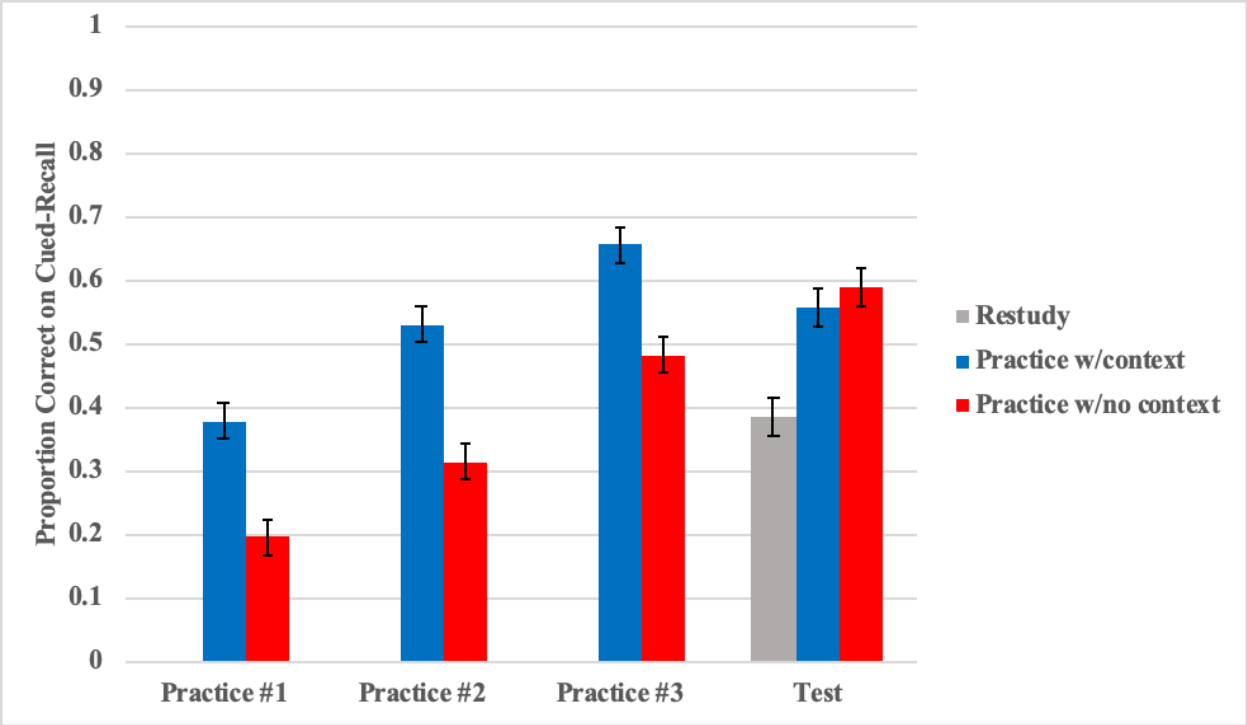

Figure 11. Experiment 4 Results. Both practice conditions outperformed restudy on the final test. Although practice conditions did not differ from each other in their test performance. The Practice with context conditions showed a marked decrease in performance from its last practice to the final test. The Practice with no context condition had significant improvement from the third practice to the final test. Error bars display standard errors.

in performance from the third practice (M = .658, SD = .237) to the final test (M = .558, SD = .223), $t(63) = 3.183$, $p = .002$, Cohen's $\delta = .41$. On the other hand, the practice with no context condition had a significant improvement from practice (M = .484, SD = .207) to test (M = .588, SD = .253), $t(59) = 5.025$, $p < .01$, Cohen's $\delta = .44$. This contrast between conditions was behind the interaction as up to the final test both conditions had improved in parallel, granted the

practice with context consistently performing higher across all practice. The latter pattern reflects the main effect of studying condition. These results are rather interesting as they suggest two different phenomena as responsible for these results. The first is the development of context dependence in the practice with context condition as its performance drops considerably once context is removed on the final test. The second is consistent with the idea of self-reinstatement as beneficial to retention. If the results from the final test (reported above) were the only analyses done, then it could be reasonably concluded that different degrees of self-reinstatement might not actually influence retention. But, a more comprehensive analysis on the practice conditions, across all practices and test, shows a pattern in agreement with the ECA.

10.3. Discussion

The fourth experiment was the first to use multiple blocks of practice/restudy and its results can be interpreted as congruent with the idea of self-reinstatement, and possibly updating, helping form a TE. The updating aspect of this experiment does not figure much in its results and conclusions as it was common to all conditions. Technically, the group which enabled both the most self-reinstatement and updating performed better than restudy, which supports the ECA, at least in a design with multiple practice/restudy blocks. But as updating was constant across all conditions there is not much to conclude from its inclusion. Instead, the results emphasize self-reinstatement as a factor that produces a TE. But, this finding needs to be considered in tandem with the practice condition that reinstated context. The latter condition was scoring substantially higher (approximately twenty percent) across all practices except in the final test where its performance dropped to the level of the other practice group. In the end, experimentally reinstating context helped practice recall as much as possible self-reinstatement. The learning trajectories for each condition are consistent with the use of context at practice, albeit in different

ways. Reinstating context at practice seemed to encourage participants to use it as cues to aid name retrieval but may have been a crutch ultimately hurting retention. Whereas enabling self-reinstatement made practice more difficult but always improved across practices. That is, context dependence raised performance high enough that despite its cost, at test, it placed equal to a condition purposefully made more difficult to maximize retention. This difficulty may have been unnecessary as an easier method ultimately led to comparable results. Even though practicing with no context obtained a TE, it did not uniquely benefit test performance. In other words, if self-reinstatement does aid retention, should it be done as there might be an easier, equally effective strategy (i.e., fully reinstating context at practice)? Thus, these results support the ECA, but may have also illustrated an easier way to use retrieval practice. As long as the learner is willing to accept his/her practice performance to be an overestimate of what will eventually be retained which is, granted, not ideal. Overall, multiple practices were able to establish a TE, but it is still not clear which of the assumptions, context retrieval as a desirable difficulty or context updating, is the most critical. Testing the relative importance of each assumption will help find which leads to the optimal retrieval practice set up.

11. EXPERIMENT 5

The fifth experiment was a slight detour of sorts as it only included retrieval practice and no restudy condition. Without the latter a TE could not be investigated. But, this experiment aimed at partially replicating Experiment 4 and further testing self-reinstatement as a desirable difficulty. Experiment 4 only included practices in which updating was viable, so that a comparison between groups differing only as to the degree of self-reinstatement (i.e., with no updating) was not possible. The present experiment included three practice conditions, one in which the original context was reinstated during name retrieval and feedback, a second in which no context was given until the original was shown at feedback, and a third condition with no context during name retrieval and a new context at feedback (i.e., identical to the Practice with no context condition from Experiment 4). Including these three conditions within one experiment allows a comparison between different degrees of the ECA's assumptions. The two conditions (with no updating) differing only in terms of self-reinstatement will give a clearer view of the importance, if any, of context retrieval as a desirable difficulty. The third condition will test the relative importance of retrieval practice trials incorporating conditions favorable for the development of the context retrieval (as a desirable difficulty) and updating, or a "fuller" account of the ECA. This latter condition is expected to do the best if the ECA is correct as to both of its assumptions. That is, self-reinstatement and updating lead to the best results among other types of practice.

11.1. Method

11.1.1. Participants

A total of 182 undergraduate students from Texas A&M University self-enrolled into the experiment in exchange for course credit. Up to 15 participants were included in each

experimental session. The no context reinstated, context reinstated, and no context reinstated plus updating conditions included 59, 50, and 73 participants, respectively.

11.1.2. Design and Materials

The design was Practice Condition (no context reinstated *v*. context reinstated *v*. no context reinstated plus updating) as a between-subjects variable. Plus, Phase (Practice #1, #2, #3, Test) was included as a within-subjects variable. The dependent variable for all analyses was correct performance, be it for practice of final test.

11.1.3. Procedures

The procedures were exactly as in Experiment 4 with the exception that out of the customary three initial phases, only the last one is present. This means the creating and recalling context labels tasks were dropped for this experiment. The first phase was that of studying the ensembles, which corresponds to phase 3 described in the General Method section. Afterwards, everything proceeded as in Experiment 4. During all practice blocks, one condition always reinstated the original context for a given face-name pair, a second only reinstated original context at feedback, and the third only showed a new context during feedback (Figure 12). The timing for trials within all phases were identical to Experiment 4. Then, the usual final cued-recall test was given immediately after the last practice block.

11.2. Results

A mixed ANOVA with Practice Condition as the between-subjects variable (no context reinstated *v*. context reinstated *v*. no context reinstated plus updating) and Phase as the within-subjects variable (Practice #1, #2, #3, Test) was conducted using correct performance as the dependent variable. There was a significant main effect of Phase, $F(2.038, 364.859) = 302.177$,

$p < .01$, $\eta_p^2 = .628$, and interaction between Phase and Practice condition, $F(4.077, 364.859) = 17.538$, $p < .01$, $\eta_p^2 = .164$ (results reported with Greenhouse-Geisser correction). There was also

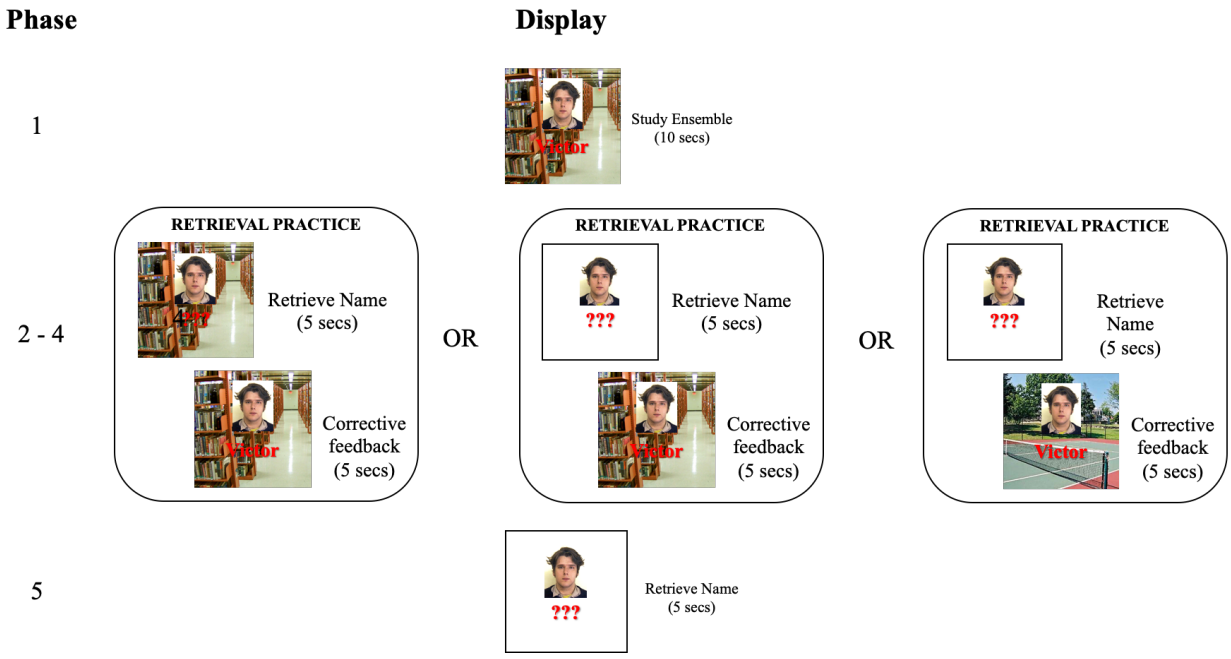

Figure 12. Phases of Experiment 5. The first phase is the study of the ensembles. Phases 2 – 4 are the multiple blocks of retrieval practice. Practice conditions reinstated original contexts through the entirety of each trial (left), or only during feedback (middle), or showed a new context at feedback (right). The final phase was a cued-recall test. Context images and faces selected from Konkle et al., (2010) and the Psychological Image Collection at Stirling (http://pics.stir.ac.uk), respectively.

a significant main effect of Practice condition, $F(2, 179) = 24.062$, $p < .01$, $\eta_p^2 = .212$. On average, the context reinstated condition performed higher than the other two conditions (p's < .01), and the no context reinstated with and without context conditions did not differ ($p = .150$). These differences mainly reflect the much higher performance the first group had across all practices, whereas the latter two groups were quite similar throughout those same phases (Figure

13). This changed on the final test (the source of the interaction) where all conditions did not differ from each other. Final test performance was a bit higher for the context reinstated condition (M = .611, SD = .207) than the no context reinstated plus updating condition (M = .562, SD = .235), $t(121) = 1.200$, $p = .232$, and slightly lower than the no context condition (M = .625, SD = .264), $t(107) = .313$, $p = .755$. The latter two conditions also did not differ from one another, $t(130) = 1.466$, $p = .145$. In sum, two conditions varying only the amount of allowed context self-reinstated (i.e., none *v.* full) and a third enabling self-reinstatement plus updating did not show any considerable differences. That is, self-reinstated and updating did not appear to provide any unique benefit to the final test results. The results suggest the inclusion of both context self-reinstatement (as a desirable difficulty) and updating does not aid retention any more than low amounts of self-reinstatement and no updating. This does not support the ECA stance that the degree of self-reinstatement or updating helps retrieval practice.

As in the previous experiment, there was a similar transition between the last practice and the final. There was a decrease in performance for the condition reinstating context from practice (M = .768, SD = .236) to test (M = .611, SD = .207) as shown by a paired samples t-test, $t(49) = 4.787$, $p < .01$, Cohen's $\delta = .71$. In further keeping with the last experiment's results, the two conditions not reinstating context significantly improved their scores. The no context reinstated conditions improved from the last practice (M = .492, SD = .228) to final test (M = .625, SD = .264) , $t(58) = 6.627$, $p < .01$, Cohen's $\delta = .53$. As did improve the no context reinstated plus updating group by a similar amount from the third practice (M = .453, SD = .195) to the test (M = .562, SD = .235), $t(72) = 5.842$, $p < .01$, Cohen's $\delta = .49$. This reinforces, and conceptually replicates, the results from Experiment 4. Namely, even though all practice condition had similar final scores, their learning trajectories were characterized by how much self-reinstatement was

possible during practice. Once again, even though final test results do not quite support the ECA, taking a more comprehensive view at performance shows a sizeable increase in learning when self-reinstatement is encouraged. And, if context is repeatedly (experimentally) reinstated, dependence on it can lead to underperforming relative to prior practice performance. Overall, practices allowing greater degrees of self-reinstatement appeared to better prepare participants for similar conditions (i.e., no context shown) at test.

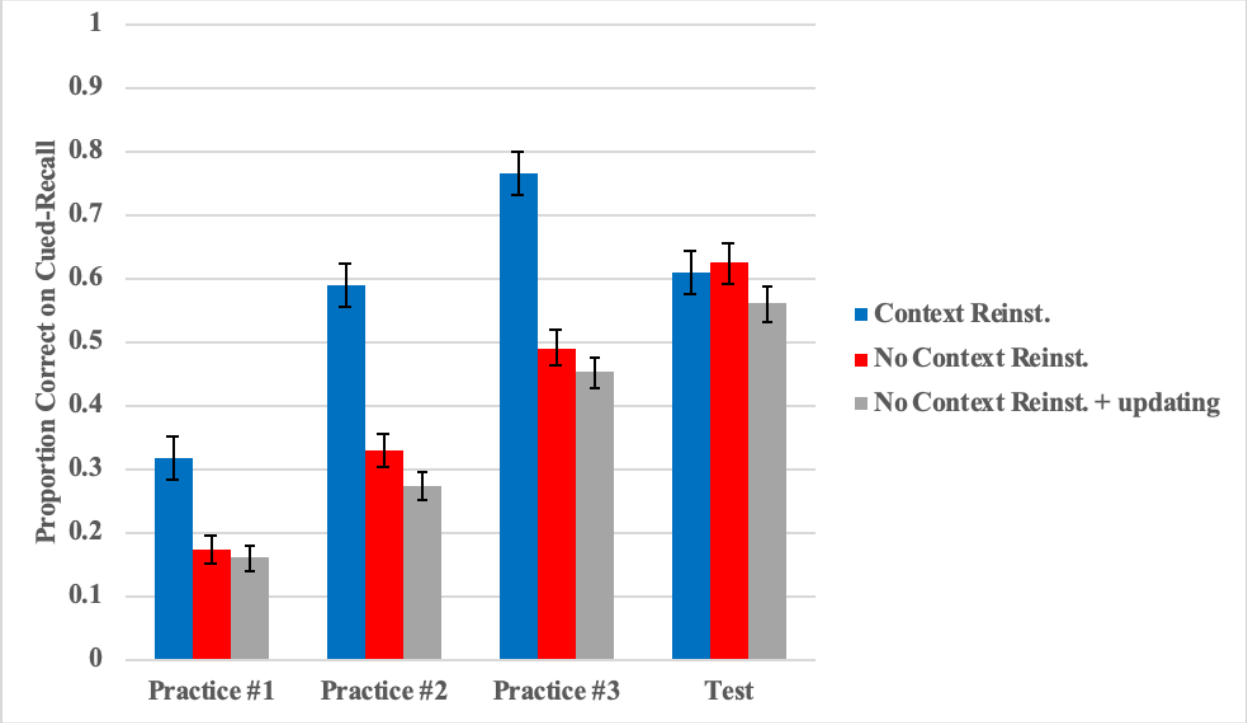

Figure 13. Experiment 5 Results. All practice conditions performed similarly on the final test. All conditions showed improvement across practices, including up to the final test with the exception of the Context Reinstated condition, which showed a significant decrease in its performance from the third practice to the final test. Error bars display standard errors.

11.3. Discussion

The fifth experiment partially replicated results from the previous experiment, and more importantly, extended those findings by comparing several practice conditions varying the presence and degree of self-reinstatement and updating. The analyses showed, for a second time, that different practice conditions did not confer any bonus benefit for retention. That is, all practice conditions had pretty close scores on the final test. No matter the possible amount of self-reinstatement or updating all conditions did about the same. This is actually not that big of a "setback" for the ECA as at least one of these conditions showed a TE in the previous experiment. And, let's remember that the purpose of the ECA is to help explain retrieval practice as it pertains to the TE and in that regard the results from this, and the prior, experiment are actually consistent with its assumptions. One condition (i.e., no context with updating) in the current and prior experiments gave very similar scores both times. Given the replication, this condition enabling both assumptions of the ECA is not unreasonably expected to consistently show a TE when using multiple practices. If the mere result of this TE were solely considered it would reflect favorably on the ECA. But, as seen by now, there is more to this picture because practice conditions allowing the "full deployment" of both assumptions are not giving any extra benefits as when they would be at their lowest, that is, with practically no self-reinstatement or updating. This puts into question the importance of these factors within retrieval practice. Do either self-reinstatement or updating matter much if there are multiple rounds of practice?
The extent to which self-reinstatement matters appears to be in the steady and reliable improvement across phases. Even though self-reinstatement conditions have always had lower practice scores than experimentally reinstated conditions their changes from practice to test have been significantly better each time. Their means for reaching the same destination, or retention,

are quite different and their long-term effects are not clear as these multiple-practice experiments have only given immediate tests. But, before planning a new research line on testing how different types of practices hold up in the long-term, it would be useful to first settle one issue still left to be assessed, the benefit of self-reinstatement without updating over restudy. That is, what is the magnitude, or importance, that the degree of self-reinstatement gives the TE?

12. EXPERIMENT 6

The final experiment returned to the same type of design as in Experiment 4 by including two practice and a restudy condition but eliminated contextual updating. Specifically, a face-name pair was shown with its original context at practice during (1) name retrieval and feedback, or (2) only during feedback, or (3) during restudy. In short, the present experiment removed the encoding of new contexts as it did not seem to provide much of an advantage over the other conditions that only modified the degree of context self-reinstatement. The lack of a benefit from encoding a new context, and being able to create a composite memory, has only received partial support. While Experiment 4 showed a TE for conditions enabling updating, one of these conditions (i.e., self-reinstatement and updating) in Experiment 5 did not perform differently from conditions only changing the degree of self-reinstatement at practice. The present experiment did not require the encoding of "new" stimuli, or context, as the assumption of updating might not be critical to the effect of the TE. Instead, two degrees of context self-reinstatement in practice was retained to test whether it is enough to obtain a TE. Because Experiment 5 did not seem to show a benefit for the updating condition over the two that did not have updating, a simpler explanation for the TE might be the use of context at practice. That is, using context as a cue, be it generated from self-reinstatement or experimentally presented, might be more important contributors to the TE than updating. As such, this experiment will test whether self-reinstatement is a desirable difficulty that helps create a TE. If the degree to which context self-reinstatement is needed aids retention, then the condition showing no context during name retrieval should outperform the other two conditions. Namely, it should do better than restudy and experimentally reinstating context.

12.1. Method

12.1.1. Participants

A total of 244 undergraduates from Texas A&M University participated in the experiment. Students self-enrolled into experimental sessions that had up to 15 participants. The self-reinstatement, experimental reinstatement, and restudy conditions had 81, 85, and 78 participants, respectively.

12.1.2. Design and Materials

The design was similar to that of Experiment 4 with two practice conditions and a restudy condition. Overall, the design had Study Condition (self-reinstatement *v*. experimental reinstatement *v*. restudy) as a between-subjects variable. The dependent variable for analyses of practices and final tests was correct performance.

12.1.3. Procedures

The procedures were similar to those in Experiment 4 which included the three phases described in the General Method section. Among the few differences was that participants were placed into one of three conditions (Figure 14). In one condition retrieval practice trials consisted of showing the original context while showing its face cue for the retrieval of its corresponding name. In a second practice condition, there was no context shown (a blank background) during name retrieval. Both conditions gave corrective feedback after each trial which included the original context for a given face-name pair. The third condition was a restudy group in which the entire ensemble, including the original context, was displayed on each trial. Then a final cued-recall test, as in the previous experiments was given immediately after the last practice or restudy block. The duration of each trial in all phases was the same as in Experiment 4.

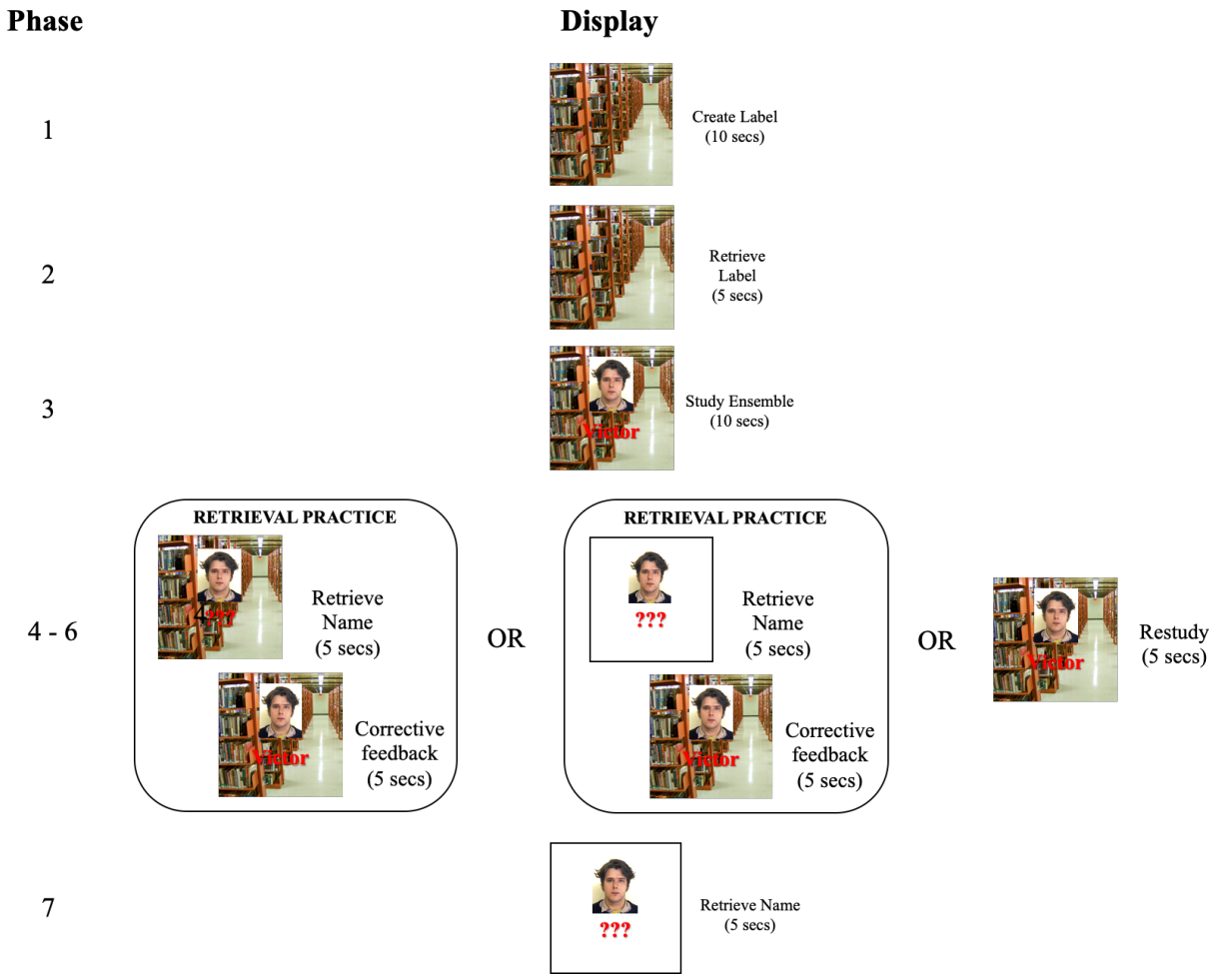

Figure 14. Phases of Experiment 6. Phases 1 – 3 were the creation of context labels, label retrieval, and ensemble studying, respectively. Phases 4 – 6 consisted of restudying (right) or practicing retrieval. Practice trials experimentally reinstated context during name retrieval and feedback (left) or only during feedback (middle). Phase 7 was the final cued-recall test. Context images and faces selected from Konkle et al., (2010) and the Psychological Image Collection at Stirling (http://pics.stir.ac.uk), respectively.

12.2. Results

A one-way ANOVA was conducted using Study Condition (self-reinstatement *v.* experimental reinstatement *v.* restudy) and correct test performance as a between-subjects and dependent variable, respectively. The analysis showed a significant main effect of Study Condition, $F(2, 241) = 18.403$, $p < .01$, $\eta_p^2 = .132$ (Figure 15). An independent samples t-tests

revealed the experimental reinstatement condition (M = .635, SD = .249) scoring significantly higher than restudy (M = .465, SD = .258), $t(161) = 4.279$, $p < .01$, Cohen's δ = .67. Likewise, the self-reinstatement condition (M = .694, SD = .232) scored higher than restudy, $t(157) = 5.878$, $p < .01$, Cohen's δ = .93. But both practices did not differ from each other, $t(164) = 1.567$, $p = .119$. The two practice conditions achieved TEs but neither practice condition conferred a greater benefit. The degree of self-reinstatement encouraged during practice did not make a difference, at least in the final test. Although, just as in previous experiments, the learning trajectories were a bit different for each condition.

A mixed ANOVA was conducted using Practice (self-reinstatement *v.* experimental reinstatement) and Phase (Practice #1, #2, #3, and Test) as a between-subjects and within-subjects variable, respectively. The dependent variable was correct performance. Results are reported with Greenhouse-Geisser corrections. Overall, there was significant main effect of Phase, $F(1.749, 286.811) = 336.427$, $p < .01$, $\eta_p^2 = .672$, and its interaction with Practice, $F(1.749, 286.811) = 63.989$, $p < .01$, $\eta_p^2 = .281$. There was also a significant main effect of Practice, $F(1, 164) = 57.512$, $p < .01$, $\eta_p^2 = .260$. This set of results reflects the continual improvement across phases and the superior performance of the experimental reinstatement condition across all phases, except the final test (Figure 15). As in previous results, experimentally reinstating context helps practice performance, but is followed by a substantial drop in the last phase. A paired samples t-test showed a decrease in performance from practice (M = .829, SD = .197) to the final test (M = .635, SD = .249), $t(84) = 7.221$, $p < .01$, Cohen's δ = .85 . Self-reinstatement showed improvement from the last practice (M = .546, SD = .196) to the test (M = .694, SD = .232), $t(80) = 9.465$, $p < .01$, Cohen's δ = .67. This last improvement was also found in past experiments for conditions requiring self-reinstatement. Overall, final test

scores were the culmination of continuously increasing performance for the self-reinstatement condition. On the other hand, the experimentally reinstated condition was subject to a drop in performance when final cued-recall conditions did not have context cues.

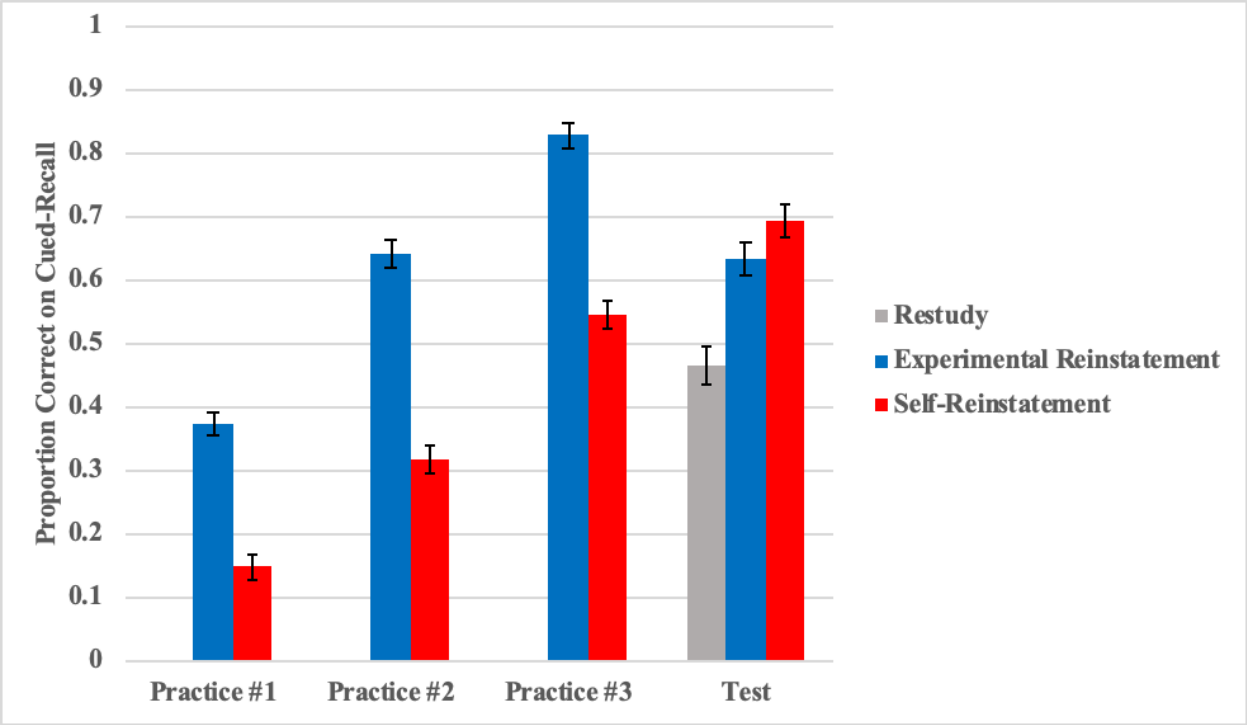

Figure 15. Experiment 6 Results. Retrieval Practice conditions performed higher than restudy on the final test. Across practices 1 – 3, both practice conditions showed improvement in performance. But, practicing with experimentally reinstated context showed a significant decrease in performance from the third practice to the final test. The self-reinstatement condition had a significant improvement across the same interval. Error bars display standard errors.

12.3. Discussion

The results from the final experiment converged with those of past multiple-practice experiments. Namely, practice conditions produced TEs, but neither self-reinstatement nor experimental reinstatement yielded a unique boost to retention. And, learning trajectories for

practice conditions depended on the degree of possible self-reinstatement during practice. The first of these patterns was also found in Experiment 4 and the latter was found in Experiments 4 and 5 as well. Because the degree of self-reinstatement did not have much of an impact on final test retention these results are not consistent with it being a desirable difficulty. This makes it the third experiment (among the latter set of experiments) in which conditions differing in terms of self-reinstatement have not shown any unique effects. The distinguishing characteristic between such conditions is the build-up or learning trajectory towards the final test. Also as seen for the third time, experimentally reinstating context gives the benefit of significantly higher cued-recall across practices but not on the final test. Self-reinstatement has a lower but constant phase-to-phase increase, but ultimately its more difficult condition does just as well as a considerably easier condition (experimentally reinstated context). Without an extended retention interval it is not known which trajectory would show to be more stable or have better long-term retention. As such, there does not appear to be any benefit of self-reinstating context under immediate testing conditions. While this part of the results does not agree with the desirable difficulties aspect of the ECA's second assumption, they are consistent with it in general. The self-reinstatement condition actually produced a TE, so it is technically possible for self-reinstatement to help during retrieval practice albeit it in a way that is reproduceable by an easier condition. In sum, enabling self-reinstatement did help retention as it yielded a TE, which is consistent with the second assumption of the ECA, but whether the benefit derived from context use depends on its *degree* of reinstatement is not supported by these results. That is, whether context is mentally reinstated or selected from available cues does not appear to make a difference. Self-reinstating to a greater degree still amounts to using context as cue, which is also done when it is experimentally shown. The degree is not as important, if at all, as that of using context to guide

retrieval, regardless of its source. Finally, it should be emphasized again that learning trajectories for practices conditions were different enough to warrant future work investigating their long-term effects. This would deliver more conclusive evidence for self-reinstatement being, or not, a desirable difficulty. As it is, the current results do not support it.

13. SUMMARY AND CONCLUSIONS

Across six experiments the extent to which context benefited learning was based on the number of practices and context cue availability at retrieval. The results could be summed as follows:

1. Experiments 1-3- showed one practice block was insufficient to obtain a TE.
2. Experiments 4 and 6 showed multiple practice blocks were necessary for a TE.
3. In Experiment 5, enabling the updating of context did not yield any particular benefit to retrieval practice.
4. Allowing greater a degree of context self-reinstatement did not confer a benefit over using experimentally reinstated context cues.
5. Conditions experimentally reinstating context showed learning trajectories consistent with context dependence.

The first conclusion stems from Experiments 1-3 in which a TE was not found. In two experiments restudy actually outperformed retrieval practice (granted, in one experiment the TE reversal was not significant) and a third did not show any differences between the two conditions. Once multiple practices were built-in the study designs then a TE emerged from all retrieval practice conditions in Experiments 4 and 6. The practice conditions from these experiments differed as to the amount of self-reinstatement possible during practice and none showed a significant advantage over the other. Thus, the degree of self-reinstatement allowed did not seem to be a desirable difficulty, meaning this aspect of the ECA's context retrieval assumption (#2) is not presently supported. The updating assumption also did not gather much evidence either as whenever new context was instructed to be encoded at practice, which enables the creation of a composite memory of previous and current contexts, there was no singular

benefit from this act. The results do not support the idea context updating positively influences retrieval practice nor that self-reinstatement is a desirable difficulty. But context retrieval, be it experimentally or self-reinstated, does help during retrieval practice as such conditions obtained TEs. This broader support for the second assumption of the ECA was accompanied by interesting results as learning trajectories for those practice conditions differed quite starkly as they progressed to the final test. These differences could signal the drawbacks and benefits of using context at retrieval. And, it should be noted these retrieval practice results apply to immediate retention intervals and multiple practices, as such were the conditions in which TEs were found. These conditions are key to future research into self-reinstatement and updating.

An unexpected result was the lack of a TE in any of the first three experiments. Upon further consideration perhaps it should not have been too surprising given that some of the "standard" findings of the TE have been with the use of multiple practices (e.g., Roediger & Karpicke, 2006b). Although, TEs are not uncommon for material receiving only one practice block (e.g., Finn and Roediger, 2013). As such, it is possible the stimuli used in these experiments, context-face-name ensembles, may have been a bit too much to learn after one study phase. Because other experiments tend to use stimuli of one format (e.g., single words, or passages) and not a mixture, as was the case in these experiments, it is possible the lack of a TE was influenced what the material itself. But, Finn and Roediger's (2013) second experiment which also used face-name pairs, albeit without pictorial contexts, were able to find a TE after one practice in a similar design, but also included the encoding of new information (i.e., the profession for a given face-name pair). Thus, in a design using similar stimuli and, up to a point, similar task demands (i.e., encoding new information/context) a testing effect appeared. But, a critical difference was their retention interval as they tested after 24 hours and current

Experiments 1 – 3 tested after 20 min. In their previous experiment, Finn and Roediger did not find a TE with immediate testing, which can suggest that even 20 mins. might not be a long enough interval to fully display the effects of a TE. That is, the long-term benefits to retention were not actually reflected in Experiments 1 – 3. And, the retention interval might not have been the only factor behind the lack of a TE. Because the first three experiments had low test performance, if the retention interval would have been longer (e.g., 24 hours) then floor effects might have shown a lack of difference between restudy and practice groups. Solely increasing the retention interval might not be the best way to modify Experiments' 1 – 3 designs so that they more accurately reflect the effects of context in retrieval practice. A second factor that would need to be changed is that of practice time. Specifically, Finn and Roediger gave participants as much time as needed during practice trials (plus a few seconds of feedback), something that was different in the present experiments. The name retrieval portion of these experiments was 4 sec. (plus 4 more of feedback) which limits the amount of time participants are exposed to the stimuli (e.g., face cues and context). If participants would have had unlimited time to retrieve the name, as in Finn and Roediger, then the exposure to these stimuli could have well increased and participants could further encode them. This can enhance their memory which could make them more effective cues. As such, the failure to find a TE could be due to a couple of main factors, retention interval and practice trial time.

The first three experiments aimed to replicate the conditions under which the ECA posited context would be used in certain ways. In Experiment 1, the absence of context at practice was meant to allow the self-reinstatement of repeatedly seen (and hence easily retrieved) context, and this resulted in a reversed TE. The single practice block, as well as the two factors mentioned in the previous paragraph, may have well contributed to this effect. But even if such

reasons were responsible for the result, the lack of a TE also shows the (reasonable) limits of the ECA's assumptions. Namely, material in the early stages of learning might not necessarily make use of context in the manner prescribed by the ECA. Using available context cues at practice is also not very helpful as final test performance tended to be below or equal to restudy in such a case (Experiment 2). Likewise, encoding new contexts during practice did not help (Experiment 3). The reliance on context might not begin until the to-be-learned stimuli is learned beyond a certain level. From the results it is unclear what this "level" might be, but presumably it is after whatever learning can occur after one practice and is (consistently) present after three practices, as seen from Experiments 4 and 6. The ECA's use of context might require amending as there were signs that context dependence might actually be hurting final test performance. In Experiment 2, the experimental reinstatement of context helped the practice conditions reach a performance level (approx. 40%) that was quite close to restudy from Experiments 4 and 6 (approx. 39% and 47%, respectively). Performance for these restudy conditions were after three such blocks and the practice conditions (in Experiment 2) were able to do about as well simply by being able to use context cues. But, once those cues were unavailable at the final test, their performance dropped to among the lowest among all experiments (approx. 20%). The salience and effectiveness of these pictorial context cues might have encouraged their use and may had led to the development of context dependence. While the ECA does discuss being able to break out of such dependence, that is, decontextualization, after repeated retrieval attempts (Karpicke, Lehman, & Aue, 2014), the overreliance on context as part of retrieval practice still needs to be fully addressed. The second experiment is quite pertinent towards this end. The "crutch" of context might be more alluring in early learning stages as it can provide an effective retrieval route, but may incur a cost at retrieval if the same context is not available. Again, it is quite

possible that this particular type of context might be more susceptible to such dependence than other types, but further work will be needed to reinforce this conclusion. As it is, these experiments suggest that using salient, and richly detailed contexts may not be the best cues to rely on if practices are not extended beyond a single block. A similar cautionary note might also apply if contextual variability is employed to help retrieval success by incrementing retrieval routes. In Experiment 3 the inclusion of new contexts at practice did not give any benefit over restudying. Attempting to build contextual variability and updating early on might also not be conducive to the TE. Overall, factors such as exposure time, retention intervals, context dependence, and number of practices may have contributed to the chipping away and reversal of a TE, but the latter two factors might be the most responsible for it. Dependence developed an overreliance on context that hampered learning and having one practice was not enough to familiarize participants with the process of self-reinstatement for target retrieval.

A TE was finally obtained with the expansion to three retrieval practices, but meaningful support for self-reinstatement as a desirable difficulty or context updating was still not present in the latter three experiments. By having participants go through three practices it was expected that any steep difficulties arising from inadequate encoding would be eliminated. Multiple practices were to gradually facilitate self-reinstatement as up to the beginning of practices this had not actually been done by participants. In doing so, self-reinstatement has the possibility of becoming a desirable difficulty instead of just a difficult process. And, multiple practices were to allow the encoding of new context as the first set of experiments (i.e., 1 – 3) might have made this too difficult as there just were not enough opportunities to do so. In Experiments 4 – 6, this "practice expansion" helped in getting overall final test performance much higher. The TEs found (Experiment 4 and 6) were for both practice conditions and were not different from each

other in terms of test performance. Practice conditions in Experiment 4 had the encoding of new context (shown at feedback) to enable updating and only differed in how much self-reinstatement was possible during name retrieval (i.e., the first portion of practice trials). As such, one condition made it possible for self-reinstatement and updating to take place and was intended to represent practice circumstances in which both assumptions of the ECA were present. This was expected to yield the greatest benefit to a TE. Results showed these "optimal" conditions for retention did not provide a benefit over experimentally reinstating context and updating. Because it was possible for updating to be doing the heavy lifting, meaning that the TE was mainly derived from the benefit of updating and not from the use of context (be it self-reinstated or not) the next experiment decided to look into the relative benefit of each of those factors. Experiment 5 had only practice conditions which differed in their amount of self-reinstatement and updating. In other terms, they differed based on which assumption was present, which were as follows: (1) no self-reinstatement, (2) self-reinstatement, and (3) self-reinstatement and updating. The third condition was expected to do the best, but that was not the case. Instead, all conditions performed rather similarly. If a condition with both assumptions "present" did not give a benefit over the other two, which only had the second assumption (or to be specific, only an aspect of it), then it might not play a role in retrieval practice or an easier way exists to obtain similar results. That is, why include context updating in retrieval practice when there are simpler practice conditions which help learning just as much? Experiment 6 removed context updating from its practice conditions and retained conditions 1 and 2 from Experiment 5. That is, it tested the possible benefit given by self-reinstatement on retrieval practice. Once the again, the parity between both practice conditions did not provide support for self-reinstatement being a desirable difficulty as posited by the ECA. This was despite having multiple practices for self-

reinstatement to become a desirable difficulty. In sum, the latter set of experiments did not find much, if any, evidence that could be interpreted as supporting self-reinstatement or updating as providing unique benefits to retrieval practice. But, the broader aspect of the ECA's second assumption (i.e., context use) did find support among these results.

The apparent ineffectiveness of updating was unsurprising given previous work on the difficulty of encoding new information during retrieval practice. As mentioned before, Davis and Chan (2015) proposed the *borrowed time hypothesis* in which encoding time is allocated to a given stimulus, for example, if perceived as being more task important than other stimuli. As relevant to Experiment 4 (and 5), practice (task) demands might have made participants pay more attention to the face-name pairs during feedback, which was exactly when new context was shown. Because the former might have seemed more important, considering they had just attempted to retrieve that same information, participants may have allocated more time and effort to further learn the face-name pairs. Despite being instructed to study the "new ensemble" at feedback, participants may have succumbed to trying to learn what was perceived as having more importance. The inclusion of new contexts could have had limited influence as a well learned composite memory might have not been formed. Without the contextual variability of a composite memory the impact of updating may have been diminished. Interestingly enough, Davis and Chan (2015) found that incentivizing the learning of new information helped its recall performance and obtain a TE for the names from face-name pairs. In their experiment, they had participants encode face-name pairs and then either restudy or practice retrieval. During restudy trials they would show an occupation to be encoded along a given face-name pair. For practice trials, name retrieval would be attempted, feedback given, and finally shown the occupation to be associated with the face-name pair. By giving an incentive to participants to learn the new

information they were able to eliminate the cost to new learning. But, the incentive was basically the early dismissal from the experiment and not merely instructing participants to emphasize certain stimuli (e.g., name or occupation). Apparently, a pilot study done by Davis and Chan in which the latter set of instructions were given to participants resulted in impairment for new learning. The instructions from (the current) Experiment 4 could have amounted to about the same, that is, by telling them to encode the new ensemble they still could have borrowed time from encoding the context to focus on the face-name pairs. Granted, the experiments from Davis and Chan do not use pictorial context as the new information that is learned and tested. Further research is needed to see whether these effects of impaired new learning apply to pictorial contexts. Plus, in their experiments names and occupations are tested separately. In the present experiments, the retrieval of new information (context) was done in the same trial as the retrieval of the names. Finn and Roediger (2013) had similar designs and materials (i.e., face-name-occupation triads) as Davis and Chan, and each final test trial required the retrieval of both names and occupations for a given face cue. They also found impaired new learning after testing (i.e., for a retrieval practice condition). Based on the findings of Experiment 4 the updating of pictorial context does not seem to play a role during retrieval practice. And, given prior work, the possibility of updating happening at practice seems doubtful if pictorial context can be categorized as new learning.

If updating does not have much, if any, of an influence on retrieval practice, then the findings from Experiment 5 can make a bit more sense. Because the final test did not reveal a singular benefit to the sole condition including updating, its comparable performance to the other two practice conditions could have been due to the use of context. Specifically, using context cues, either by self-reinstating them or selecting them from their experimental reinstatement, can

be a key part of what makes retrieval practice effective. In this experiment two practice conditions varying only the amount of self-reinstatement allowed were compared for the first time. Because these two groups did not differ, the issue of self-reinstatement as a desirable difficulty was not supported but context in general played a significant role. This shows the value of context cues, internal or external, as effective retrieval aids. Although, it needs to be noted that Experiment 5 did not include the first two phases, included in the rest of the experiments, involving the creation and retrieval of context labels. If these phases help familiarize participants with the contexts, they may help self-reinstatement during practice blocks and bolster final test performance. Would including those two initial phases actually reveal a difference between self-reinstatement conditions? Rather, does self-reinstatement require context to be well known to activate its role as a desirable difficulty? Well, the answer appears to be no as described below.

Experiment 6 had a full design, which included the aforementioned initial two phases, just as the fourth experiment and despite showing a TE for both practice conditions their different levels of possible self-reinstatement did not make a difference on final test performance. The greater familiarity the two initial phases were intended to provide did not significantly impact self-reinstatement, but it still slightly influenced the conditions. Because the two practice conditions in Experiment 6 were also present in Experiment 5, their retrieval practice and test performances could reflect an influence from greater context familiarity stemming from creating and retrieving context labels (done in Experiment 6). The no context self-reinstatement, or experimentally reinstated, condition in both Experiment 5 and 6 had similar scores at 61.1% and 63.5%, respectively. The extra context exposure did not seem to impact test performance. The self-reinstatement conditions actually had a slightly higher gap in scores. Self-reinstatement test scores went from 62.5% to 69.4% across Experiments 5 and 6,

respectively. This increase in performance shows a modest, at best, influence of two phases devoted to encoding context. Improved accessibility of context did not aid self-reinstatement much. As mentioned before, the benefit of context cues might not really differ be them internally generated or experimentally reinstated. In general, results point toward this conclusion as self-reinstated conditions actually performed just a bit higher than experimentally reinstated conditions. These differences were never significant, but across two of the three final experiments the trend persisted. Again, any benefit from self-reinstatement might be negligible based on these results.

This series of experiments raised further questions about context self-reinstatement as a desirable difficulty. Because the only two levels of self-reinstatement used throughout the experiments were either "high or low," future work would need to probe deeper into this issue by using levels that are not on the extremes of difficulty. That is, the practice conditions could partially reinstate contexts during each trial and vary the amount of elements that are shown to better manipulate the degree to which self-reinstatement is needed. The use of pictorial context may be useful with such work as it may be easier to control which elements are experimentally reinstated. Previous research has shown diminishing the number of cues at each successive practice can be an alternative to a "standard" retrieval practice. For example, Fiechter and Benjamin (2018) had participants learn cue-target word pairs and across multiple blocks they had a retrieval practice condition, that is, present only the cue to retrieve its target. Or, they would show the cue and target, but with each practice a letter would be dropped from the target word, so these practices would diminish the number of cues available for target retrieval. The authors found evidence for the latter condition performing better than a restudy condition and retrieval practice (when no feedback is given). By beginning with easy retrieval conditions (i.e., target

fully shown) and slowly making its retrieval harder there is a gradual increase in task/retrieval difficulty which was held to help retention. Future work could attempt to replicate this benefit but as applied to context and not the diminishing of target cues.

Among the most intriguing findings, in the latter three experiments, were the learning trajectories for practice conditions. Specifically, experimentally reinstated conditions consistently outperformed self-reinstatement conditions across all practices until the final test. The lower test performance for the former group is quite possibly due to the effects of context dependence (Smith & Handy, 2016). After repeated practices in which context is easily, and effectively, used as a cue it becomes a "crutch." Its overreliance as the primary retrieval route was disrupted on the final test as those trials did not have any context cues, which resulted in a marked decrease in performance from the last instance of context availability (Practice #3) to the test. This negative consequence of context use could be further delved by testing its presence among other types of context and how to avoid it or even decontextualize memories. Because the pictorial contexts used in these experiments were salient and richly detailed it is possible that may have contributed to the inducement of context dependence. Other types of contexts that do not share these characteristics, at least to the same degree, might not be as conducive to context dependence. Likewise, with enough practices or exposure to many contexts target memories may become decontextualized as suggested by the design and results from Smith and Handy (2016, see the Current Experiments section for a description). As such, future work could build contextual variability to a greater extent than in Experiments 4 – 6 to determine how any dependence on one, or a few, contextual associations can be reduced and decontextualize memories.

Overall, six experiments did not find results supporting context self-reinstatement as a desirable difficulty or contextual updating within retrieval practice. Single block practices/restudy designs did not show a TE (and indicated restudy might be slightly more beneficial), suggesting multiple rounds of practice are needed before it becomes beneficial towards retention. Upon the expansion to three practices, TEs were found but results still did not support either the desirable difficulties aspect of the ECA's context retrieval assumption or its updating assumption. Rather, it seems context use regardless of source (internal or external) can have a comparable beneficial effect on retention. These results show pictorial context can be employed during retrieval practice to aid learning and further elucidates its role within the TE.

REFERENCES

Balota, D. A., Duchek, J. M., & Logan, J. M. (2007). Is Expanded Retrieval Practice a Superior Form of Spaced Retrieval? A Critical Review of the Extant Literature. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 83-105). New York, NY, US: Psychology Press.

Bäuml, K. H. T. (2019). Context retrieval as a critical component in selective memory retrieval. *Current Directions in Psychological Science*, *28*(2), 177-182.

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive psychology, 61*(3), 228-247.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123-144). Hillsdale, NJ: Erlbaum.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, *2*(59-68).

Bjork, R. A., & Richardson-Klavehn, A. (1989). On the puzzling relationship between environmental context and human memory. In *Current issues in cognitive processing: The Tulane Flowerree Symposium on Cognition* (pp. 313-344).

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, *106*(3), 849.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118.

Butler, A. C., & Roediger III, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*(4-5), 514-527.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & cognition*, *34*(2), 268-276.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin*, *132*(3), 354.

Davis, S. D., & Chan, J. C. (2015). Studying on borrowed time: How does testing impair new learning?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1741.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological review*, *62*(3), 145.

Fiechter, J. L., & Benjamin, A. S. (2018). Diminishing-cues retrieval practice: a memory-enhancing technique that works when regular testing doesn't. *Psychonomic bulletin & review*, *25*(5), 1868-1876.

Finn, B. (2017). A framework of episodic updating: An account of memory updating after retrieval. In *Psychology of Learning and Motivation* (Vol. 67, pp. 173-211). Academic Press.

Finn, B., & Roediger III, H. L. (2013). Interfering effects of retrieval in learning new information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1665.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269-299.

Howard, M. W., Jing, B., Rao, V. A., Provyn, J. P., & Datey, A. V. (2009). Bridging the gap: Transitive associations between items presented in similar temporal contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 391.

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology*, *83*(2p1), 340.

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, *101*(3), 621.

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 530.

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review*, *27*(2), 317-326.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772-775.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In *Psychology of learning and motivation* (Vol. 61, pp. 237-284). Academic Press.

Karpicke, J. D., & Roediger III, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of experimental psychology: learning, memory, and cognition*, *33*(4), 704.

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, *62*(3), 227-239.

Kliegl, O., & Bäuml, K. H. T. (2016). Retrieval practice can insulate items against intralist interference: Evidence from the list-length effect, output interference, and retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 202.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological science*, *21*(11), 1551-1556.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P.E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). New York: Academic Press.

Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2008). Test-enhanced learning in medical education. *Medical education*, *42*(10), 959-966.

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1787.

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *Journal of memory and language*, *64*(3), 249-255.

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological review*, *122*(2), 337.

Martin, E. (1968). Stimulus meaningfulness and paired-associate transfer: an encoding variability hypothesis. *Psychological review*, *75*(5), 421.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory.

Mensink, G. J. M., & Raaijmakers, J. G. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology, 33*(2), 172-186.

Miller, J. F., Lazarus, E. M., Polyn, S. M., & Kahana, M. J. (2013). Spatial clustering during memory search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 773.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological review*, *116*(1), 129.

Psychological Image Collection at Stirling. (n.d.). [Face set]. Retrieved from http://pics.psych.stir.ac.uk/.

Putnam, A. L., Sungkhasettee, V. W., & Roediger III, H. L. (2016). Optimizing learning in college: tips from cognitive psychology. *Perspectives on Psychological Science*, *11*(5), 652-660.

Richardson, J. T. (1985). The effects of retention tests upon human learning and memory: An historical review and an experimental analysis. *Educational Psychology*, *5*(2), 85-114.

Roediger III, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on psychological science*, *1*(3), 181-210.

Roediger III, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, *17*(3), 249-255.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432.

Schwoebel, J., Depperman, A. K., & Scott, J. L. (2018). Distinct episodic contexts enhance retrieval-based learning. *Memory*, *26*(9), 1291-1296.

Siegel, L. L., & Kahana, M. J. (2014). A retrieved context account of spacing and repetition effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 755.

Smith, S. M. (2013). Effects of environmental context on human memory. *The SAGE handbook of applied memory, SAGE*, 162-182.

Smith, S. M., & Handy, J. D. (2014). Effects of varied and constant environmental contexts on acquisition and retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1582.

Smith, S. M., & Handy, J. D. (2016). The crutch of context-dependency: Effects of contextual support and constancy on acquisition and retention. *Memory*, *24*(8), 1134-1141.

Smith, S. M., Handy, J. D., Angello, G., & Manzano, I. (2014). Effects of similarity on environmental context cueing. *Memory*, *22*(5), 493-508.

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

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*(2), 176-199.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*(9), 641.

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, *38*(2), 244-253.

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.

Whiffen, J. W., & Karpicke, J. D. (2017). The role of episodic context in retrieval practice effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1036.