1 Introduction

The term "testing effect" refers to the finding that, when it comes to long-term retention of a piece of information, retrieving that information from memory trumps restudying it (Adesope, Trevisan, & Sundararajan, 2017; Glover, 1989; Roediger III & Butler, 2011; Roediger III & Karpicke, 2006a, 2006b; Rowland, 2014). It is generally recognised that testing can have two types of effects - *direct* and *indirect* (Arnold & McDermott, 2013b; Roediger III & Karpicke, 2006a). Direct effects refer specifically to the increased retention that ensues from repeatedly *successfully* retrieving the target information - a process which is, presumably, reactivated at the time of a later test. A typical demonstration of the testing effect entails an initial learning phase, followed by a period during which participants either restudy the same material, engage in a memory test involving the studied material, or are not exposed to the original material at all. Finally, after a retention interval, an additional memory test reveals that the group subjected to a memory test during the intervening period has a distinct advantage over the other two groups.

On the other hand, indirect effects are brought about by some other process or processes besides the act of taking the test (Roediger III & Karpicke, 2006a; but for a different view, see Kornell, Klein, & Rawson, 2015). For example, *unsuccessful* retrieval attempts (which are not followed by feedback) can, through subsequent repeated encoding, also generate a testing effect, namely *test-potentiated (re)learning* (Arnold & McDermott, 2013a, 2013b; Izawa, 1966, 1970; Kornell, Hays, & Bjork, 2009; Wissman & Rawson, 2018). In order to isolate the effect of unsuccessful from successful retrieval attempts, Arnold and McDermott (2013b) let participants learn a list of 25 word pairs. One group completed 9 cycles comprising a single test and a restudy session, while another completed three cycles comprising five tests and a restudy session. In the end, all participants took a final test. Results showed that, compared to taking fewer tests, taking more tests produces a greater increase in the proportion of *newly* retrieved items in a test immediately following a restudy episode - a veritable potentiation of learning.

Test-potentiated new learning

However, juxtaposed to the well established finding that attempting to recall learned material, be it successful or unsuccessful, compared to merely restudying it, facilitates the long-term retention of *that* material, stand the results of a decade of research showing that retrieving previously studied information can even facilitate the acquisition of *new* information (Chan, Meissner, & Davis, 2018; Yang, Potts, & Shanks, 2018). If each additional study episode in the paradigm used to demonstrate test-potentiated learning contains *new* materials, one still observes that taking a test on the new materials after each learning episode yields a greater number of correct responses and a decrease of proactive interference on a test administered to all subjects after the final learning episode (e.g. Szpunar, Khan, & Schacter, 2013; Szpunar, McDermott, & Roediger, 2008; Wissman, Rawson, & Pyc, 2011). Following the reasoning of Chan, Meissner, and Davis (2018), in this paper we will use the term "test-potentiated *new* learning" (TPNL) to denote this effect.

In one of the earliest studies showing the effect of TPNL, Darley and Murdock (1971) observed that, when recalling studied lists of words, participants systematically produce more prior-list intrusions when probed for a given list, if their memory of a prior list had not been tested before they proceeded to study the given list. The findings were soon corroborated by Tulving and Watkins (1974). Building on these results, Szpunar et al. (2008) conducted a study using a standard blocked design (Chan, Meissner, & Davis, 2018) wherein they told their subjects to study five lists of items in anticipation of a final cumulative test. All subjects were tested immediately after studying the final list, but they engaged in different intermittent activities between studying the first four lists. One group was tested on each list after studying it, another group restudied each list, and a third group completed a mathematical distractor task. Participants whose memory was tested after each list produced more correct responses and fewer prior-list intrusions on the immediate test administered after studying the last list, compared to the groups that were not tested. The authors explained the found benefit of testing in terms of a segregation mechanism that prevents overburdening of retrieval cues, which, in the absence of testing, causes a build-up of proactive interference. The following decade has seen a renewal of interest in TPNL (Chan, Meissner, & Davis, 2018; Pastötter & Bäuml, 2014; Yang et al., 2018), with studies mainly using the multilist learning paradigm to delineate the scope of the effect with respect to various moderating variables:

the type of study materials, varieties of study designs, and populations, to name a few.

Theoretical overview

Recently, Chan, Meissner, and Davis (2018) provided a meta-analytic analysis and comprehensive overview of the literature, identifying four *nonconflicting* theoretical frameworks which were put forth throughout the years as viable explanations for TPNL. *Resource theories* generally posit that testing increases cognitive resources, but they propose different mechanisms by which this is achieved: (1) proactive interference reduction (e.g. Nunes & Weinstein, 2012; Szpunar et al., 2008; Wahlheim, 2015; Weinstein, McDermott, & Szpunar, 2011), (2) restoration of encoding/attentional resources (e.g. Pastötter, Schicker, Niedernhuber, & Bäuml, 2011), or (3) alteration of mind wandering patterns (e.g. Jing, Szpunar, & Schacter, 2016; Szpunar, Khan, & Schacter, 2013; Szpunar, Moulton, & Schacter, 2013). Whereas resource theories focus on the amount of deployable cognitive resources, *metacognitive theories* emphasise the optimisation of encoding strategies induced by retrieval attempts (e.g. Chan, Manley, Davis, & Szpunar, 2018; Cho, Neely, Crocco, & Vitrano, 2017). For example, in a recent investigation, Chan, Manley, et al. (2018) found that, compared to untested groups, the group whose memory for the first three word lists was subjected to interpolated testing displayed superior semantic organisation across lists. These findigs reflect a similar pattern obtained for the testing effect, where a greater number of tests is associated with improved organisation of output displayed upon testing (Karpicke, 2012; Zaromb & Roediger, 2010).

The key idea underlying the third framework (*context theories*) is that, apart from storing the studied information per se, people store the related contextual information as well (e.g. Lehman, Smith, & Karpicke, 2014). Afterwards, the accessibility of this contextual information can affect the likelihood of successful retrieval of target information. Furthermore, the claim is that, unlike restudying, attempting retrieval causes an internal context change relative to the study context (Jang & Huber, 2008; Sahakyan & Kelley, 2002), and recalled items may be updated with contextual information from the retrieval attempt, while newly encountered information is still associated only with the study context. Therefore, recalling new-learning items is limited to only those items associated exclusively with the study context, providing them with the advantage observed upon testing. While this circumscription of separate learning episodes is at the core of both resource and

context accounts, its effect on learning is supposedly different. According to the former, isolating a learning episode through attempts at recall increases resources for subsequent learning by preventing *encoding-based* proactive interference. On the other hand, the latter place the emphasis on later *retrieval* processes, whereby isolating an earlier learning episode reduces the memory search set for retrieval.

Finally, *integration theories* advance the notion that interpolated testing facilitates the integration of the new-learning material either with its retrieval cues or with the original-learning material. On one account, testing increases the likelihood of spontaneous covert retrieval of original-learning items during the study of new items, fostering their integration, thereby increasing conceptual organisation (e.g. Jing et al., 2016) and the effectiveness of retrieval cues (Pyc & Rawson, 2010). For example, Jing et al. (2016) found that interpolated testing increased the clustering of related information that is acquired across different segments within a video-recorded lecture.

Nonepisodic recall

One of the more curious findings in the field is that TPNL can arise not only after retrieving the previously studied material (episodic retrieval), but also after retrieval of information unrelated to the studied material from semantic memory (Divis & Benjamin, 2014; Pastötter et al., 2011), or from short-term memory (Pastötter et al., 2011), although there have been unsuccessful attempts at replication (e.g. Weinstein, McDermott, Szpunar, Bäuml, & Pastötter, 2015).

Pastötter et al. (2011) let their participants learn five lists of 20 words while engaging in varied interlist activities. They either restudied the lists, recalled the words from the list, generated as many words as they could from one of four semantic categories (e.g. professions), engaged in a 2-back short-term memory task, or counted backwards from a random three-digit number. They found that all three forms of retrieval induced TPNL. In their first experiment, Divis and Benjamin (2014) adapted the procedure from Pastötter et al. (2011), using only the semantic generation and distractor (counting backwards) tasks, and found that interleaved semantic retrieval enhanced performance for final list recall. They replicated and extended these findings in their second experiment by using complex learning materials: lists of words were replaced by texts related to animals, while learning was evaluated with short-answer and multiple-choice questions.

However, their protocol didn't allow for assessing the contribution of proactive interference to TPNL. Furthermore, the authors argue for the irrelevance of the level of difficulty of alternative activities (i.e. perhaps the semantic retrieval and the distractor tasks were not equal with regards to difficulty) whilst referring to the variety of tasks used by Pastötter et al. (2011), who found effects of similar magnitude for the retrieval tasks and a lower but approximately equal magnitude for the distractor task and restudy. But their argument does not follow from these findings. In other words, one cannot claim that these findings refute an explanation based on task difficulty.

The argument these authors invoke to explain their results is that nonepisodic retrieval tasks sufficiently alter participant's internal context. Because the last study session is not affected by an additional context shift, a beneficial segregation of the final study context from the previous ones is produced. However, summarising their results, Chan, Meissner, and Davis (2018) highlighted resource and integration theories as accounts which have thus far garnered more empirical support, giving a slight upper hand to integration theories, while stating that context

Feedback

Although corrective feedback is known to augment the testing effect (Roediger III & Butler, 2011), there is a paucity of research into the effect of feedback on TPNL, especially when considering studies that have implemented the standard blocked design (Chan, Meissner, & Davis, 2018). Feedback is particularly important for recognition test such as mutliple choice tests since the usual benefit testing confers might turn into a disadvantage in case the test-taker selects a lure (Marsh, Roediger, Bjork, & Bjork, 2007; Roediger & Marsh, 2005). Moreover, evidence points to the timing of feedback being a relevant variable when gauging its influence on learning, with delayed feedback showing superior effects compared to immediate feedback (Butler, Karpicke, & Roediger, 2007; Butler & Roediger, 2008; Metcalfe, Kornell, & Finn, 2009; Smith & Kimball, 2010). For example, participants in a study by Butler and Roediger (2008) read prose passages and then either took or did not take an initial multiple-choice test. If they took the test, corrective feedback was either not given, given immediately after each answer was provided, or given in bulk after the entire test. A final test administered one week after the initial test revealed (1) that taking an initial test alone tripled the success rate on the final test relative to studying, (2) that giving immediate feedback

increased performance for another 10%, and (3) that delayed feedback increased performance even further by 11%.

The variable of corrective feedback may be a fruitful avenue for research because resource and integration theories provide conflicting predictions regarding its effects on TPNL (Chan, Meissner, & Davis, 2018). Providing corrective feedback should increase the likelihood of intrusions during new learning, which are deemed beneficial from the standpoint of integration theories, but detrimental from the point of view of resource theories. Thus, feedback should reduce TPNL according to resource theories, but increase it according to integration theories.

Present study

Our study had several aims. Firstly, we sought to replicate the effect in an ecologically valid setting, by using complex learning materials and standard multiple-choice items. Secondly, acknowledging the relative dearth of studies using nonepisodic retrieval as the interpolated activity, and furthermore a lack of studies using a blocked (vs.) on the moderating influence of feedback in TPNL, guided by the analysis of gaps in the field provided by Chan, Meissner, and Davis (2018). . Secondly, we attempted to expand the existing body of literature.

Wanting to examine TPNL in a more ecologically valid setting, we...

1.1 Notes for discussion

Chan et al. (2018). across a retention interval In contrast to unrelated word lists, text passages and videos are typically written/produced in a coherent manner, which should naturally invite relational processing, so any relational processing advantage induced by prior testing is likely to be modest relative to of the strategy change account that is not tied strictly to relational processing, however, may provide a reasonable explanation for the TPNL effect with text passages and videos. In a broader sense, the strategy change account specifies that performing retrieval practice allows participants to discover the type of learning needed to ensure satisfactory performance (or conversely, to realize the type of learning that is inadequate to produce satisfactory performance, if participants are performing poorly during retrieval practice), and participants can then adjust their subsequent

encoding strategy accordingly. If we take this broader approach to strategy change, then this account can explain the TPNL effect with prose/video materials. However, we realize that the idea that "retrieval practice can improve later encoding strategies" is perhaps vaguely defined. In fact, such a broad definition of strategy change may render the account difficult to falsify. With this in mind, we believe that the strategy change account, as we currently conceive, should only be applied to explain the TPNL effect with word list type materials, for which advantageous encoding strategies can be more precisely defined (but see Jing et al., 2016 in which interspersed testing improved conceptual integration of materials across sections of a video lecture). In our opinion, application of this account to prose/video material should only be done when one clearly outlines what is considered an advantageous encoding strategy so that the hypothesis can be adequately tested.

Možda ZV intruzori nije pokazala razlike između skupina jer smo koristili recognition, a ne free recall. Context change account?

Methodological concerns. The expectation of a final test ensured the continued processing of materials across the study sequence. Chan et al. (2018): For example, in a multilist learning environment, having taken a recent memory test increases learners' expectation that they will again be tested in the immediate future, even when they are told that whether a test will follow each study list is determined randomly (Weinstein et al., 2014). Such test expectancies have been shown to significantly influence how participants approach Weinstein et al. (2014): The experiments reported here are based on the assumption that participants in the tested group may be more likely to expect a test after the fifth (last) list, having consistently received tests after previous lists. Those in the untested group, having never received a test during the experiment until the fifth list test, may therefore pay less attention or engage in lower quality encoding strategies during encoding of the fifth list. To test for this alternative explanation, we compared the two standard conditions (tested and untested) with two novel conditions that were identical to the standard conditions but included a warning before presentation of the final list to alert participants that they would be tested on the upcoming list. If attentional processes are mediating the observed release from proactive interference, warning participants in the untested group should produce the same benefit as the participants taking a test after every list

Matej [11:05 PM] E, sjetio sam se opet da bi možda bilo dobro da negdje navedemo razlike u učinku na testovima prije zadnjega...

Denis [11:09 PM] Mda, neke te stvari su mi pale na pamet, i činilo mi se kao da bi bilo zgodno spomenuti ih u raspravi. Kao, ako se radi o nekim stvarima koje bi mogle ugroziti efekte ili objasniti neznačajne

Pastötter et al. (2011): Whereas, relative to the two no-retrieval conditions, both episodic and semantic memory retrieval effectively eliminated List 1–4 intrusions during immediate List 5 recall, short-term memory retrieval led to intrusion rates that were equivalent in amount to the no-retrieval conditions. This difference in intrusion errors may suggest that the effects of short-term and long-term memory retrieval are not perfectly identical.

Yang et al. (2018): za neznačajan efekt na intruzore! the activation facilitation and enhanced encoding effort mechanisms may play important roles for complex materials whereas the release from PI mechanism is likely to play little role.29

2 Methods

These are the methods.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017, June). Rethinking the Use of Tests: A Meta-Analysis of Practice Testing. *Review of Educational Research*, *87*(3), 659–701. doi: 10.3102/0034654316689306

Arnold, K. M., & McDermott, K. B. (2013a, June). Free recall enhances subsequent learning. *Psychonomic Bulletin & Review*, *20*(3), 507–513. doi: 10.3758/s13423-012-0370-3

Arnold, K. M., & McDermott, K. B. (2013b). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 940–945. doi: 10.1037/a0029199

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, *13*(4), 273–281. doi: 10.1037/1076-898X.13.4.273

Butler, A. C., & Roediger, H. L. (2008, April). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*(3), 604–616. doi: 10.3758/MC.36.3.604

Chan, J. C. K., Manley, K. D., Davis, S. D., & Szpunar, K. K. (2018, October). Testing potentiates new learning across a retention interval and a lag: A strategy change perspective. *Journal of Memory and Language*, *102*, 83–96. doi: 10.1016/j.jml.2018.05.007

Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, *144*(11), 1111–1146. doi: 10.1037/bul0000166

Cho, K. W., Neely, J. H., Crocco, S., & Vitrano, D. (2017, July). Testing enhances both encoding and retrieval for both tested and untested items. *Quarterly Journal of Experimental Psychology*, *70*(7), 1211–1235. doi: 10.1080/17470218.2016.1175485

Darley, C. F., & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, *91*(1), 66–73. doi: 10.1037/h0031836

Divis, K. M., & Benjamin, A. S. (2014, October). Retrieval speeds context fluctuation: Why semantic generation enhances later learning but hinders prior learning. *Memory & Cognition*, *42*(7), 1049–1062. doi: 10.3758/s13421-014-0425-y

Glover, J. A. (1989). The "Testing" Phenomenon: Not Gone but Nearly Forgotten. *Journal of Educational Psychology*, *81*(3), 392–399.

Izawa, C. (1966, June). Reinforcement-Test Sequences in Paired-Associate Learning. *Psychological Reports*, *18*(3), 879–919. doi: 10.2466/pr0.1966.18.3.879

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology*, *83*(2, Pt.1), 340–344.

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: recalling from long-term memory drives list isolation. *Journal of experimental psychology. Learning, memory, and cognition*, *34*(1), 112–127.

Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied*, *22*(3), 305–318. doi: 10.1037/xap0000087

Karpicke, J. D. (2012, June). Retrieval-Based Learning: Active Retrieval Promotes Meaningful Learning. *Current Directions in Psychological Science*, *21*(3), 157–163. doi: 10.1177/0963721412443552

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989–998. doi: 10.1037/a0015729

Kornell, N., Klein, P. J., & Rawson, K. A. (2015, January). Retrieval attempts enhance learning, but retrieval success (versus failure) does not matter. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 283–294. doi: 10.1037/a0037850

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), 1787–1794. doi: 10.1037/xlm0000012

Marsh, E. J., Roediger, H. L., Bjork, R. A., & Bjork, E. L. (2007, April). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review*, *14*(2), 194–199. doi: 10.3758/BF03194051

Metcalfe, J., Kornell, N., & Finn, B. (2009, December). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition*, *37*(8), 1077–1087. doi: 10.3758/MC.37.8.1077

Nunes, L. D., & Weinstein, Y. (2012, February). Testing improves true recall and protects against the build-up of proactive interference without increasing false recall. *Memory*, *20*(2), 138–154. doi: 10.1080/09658211.2011.648198

Pastötter, B., & Bäuml, K.-H. T. (2014, April). Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.00286

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 287–297. doi: 10.1037/a0021801

Pyc, M. A., & Rawson, K. A. (2010, October). Why Testing Improves Memory: Mediator Effectiveness Hypothesis. *Science*, *330*(6002), 335–335. doi: 10.1126/science.1191465

Roediger, H. L., & Marsh, E. J. (2005). The Positive and Negative Consequences of Multiple-Choice Testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1155–1159. doi: 10.1037/0278-7393.31.5.1155

Roediger III, H. L., & Butler, A. C. (2011, January). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. doi: 10.1016/j.tics.2010.09.003

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. Retrieved from `http://www.jstor.org/stable/40212166`

Roediger III, H. L., & Karpicke, J. D. (2006b, March). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17*(3), 249–255. doi: 10.1111/j.1467-9280.2006.01693.x

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. doi: 10.1037/a0037559

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of experimental psychology. Learning, memory, and cognition*, *28*(6), 1064–1072.

Smith, T. A., & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay–retention effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 80–95.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013, April). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, *110*(16), 6313–6317. doi: 10.1073/pnas.1221764110

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1392–1399. doi: 10.1037/a0013082

Szpunar, K. K., Moulton, S. T., & Schacter, D. L. (2013). Mind wandering and education: from the classroom to online learning. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00495

Tulving, E., & Watkins, M. J. (1974). On Negative Transfer: Effects of Testing One List on the Recall of Another. *Journal of Verbal Learning and Verbal Behavior*(13), 181–193. doi: https://doi.org/10.1016/S0022-5371(74)80043-5

Wahlheim, C. N. (2015, January). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition*, *43*(1), 27–38. doi: 10.3758/s13421-014-0455-5

Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011, April). Testing protects against proactive interference in face–name learning. *Psychonomic Bulletin & Review*, *18*(3), 518. doi: 10.3758/s13423-011-0085-x

Weinstein, Y., McDermott, K. B., Szpunar, K. K., Bäuml, K.-H., & Pastötter, B. (2015, November). Not All Retrieval During Learning Facilitates Subsequent Memory Encoding..

Wissman, K. T., & Rawson, K. A. (2018, April). Test-potentiated learning: three independent replications, a disconfirmed hypothesis, and an unexpected boundary condition. *Memory*, *26*(4), 406–414. doi: 10.1080/09658211.2017.1350717

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011, December). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review*, *18*(6), 1140–1147. doi: 10.3758/s13423-011-0140-7

Yang, C., Potts, R., & Shanks, D. R. (2018, April). Enhancing learning and retrieval of new information: a review of the forward testing effect. *npj Science of Learning*, *3*(1), 8. doi: 10.1038/s41539-018-0024-y

Zaromb, F. M., & Roediger, H. L. (2010, December). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, *38*(8), 995–1008. doi: 10.3758/MC.38.8.995