



Bartlett revisited: Direct comparison of repeated reproduction and serial reproduction techniques



Henry L. Roediger III^{a,*}, Michelle L. Meade^b, David A. Gallo^c, Kristina R. Olson^d

^a Washington University in St. Louis, United States

^b Montana State University, United States

^c University of Chicago, United States

^d University of Washington, United States

ARTICLE INFO

Article history:

Received 1 February 2014

Received in revised form 12 May 2014

Accepted 20 May 2014

Available online 10 June 2014

Keywords:

Repeated reproduction

Serial reproduction

F.C. Bartlett

Errors of memory

DRM paradigm

Input-bound scoring

Output-bound scoring

ABSTRACT

Bartlett developed the procedures of repeated reproduction (the same person repeatedly recalling information) and serial reproduction (people transmitting information from one person to another). Our experiment directly compared recall accuracy across these two techniques, which has not previously been reported, using DRM word lists. Recall of the initial study list words remained constant across repeated reproductions but declined markedly across serial reproductions. In contrast, recall of associated words that were not originally studied (i.e. critical words) was steady across both conditions. Because more of the original list words were forgotten across each link of the serial reproduction chain, the proportion of critical items recalled (relative to list words) increased significantly as the list passed between people. Using output bound scoring, serial reproduction resulted in lower accuracy than repeated reproduction by the final recall trial. Our results are broadly consistent with Bartlett's (1932) informal observations: Serial reproduction produces greater forgetting of the original material than does repeated reproduction and also leads to greater distortion relative to the proportion of correct material recalled.

© 2014 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.

1. Introduction

In the classic book, *Remembering: A Study in Experimental and Social Psychology*, Bartlett (1932) argued that memory is a reconstructive process. He theorized that individuals rely on pre-existing knowledge, or schemas, to encode and remember information. Schemas can aid encoding and retrieval, but during the course of recall, they may also distort recollections of the event to be more in line with the prior schema. Bartlett noted several processes (leveling, sharpening and others) that may distort recollection of an episode through repeated retelling.

Bartlett (1932) based these conclusions on naturalistic studies that used two primary techniques: Repeated reproduction and serial reproduction. Repeated reproduction refers to repeated recall by the same person over time, whereas serial reproduction refers to a chaining technique by which one person's recall is the next person's study material, and then this second person's recall becomes

the third person's material to learn, and so on (much like the childhood game of "telephone" in the U.S.). Bartlett intended repeated reproduction to examine processes of an individual's recall of the same event over time. Serial reproduction, on the other hand, was meant to reflect the social process transmission of information across individuals.

Bartlett's (1932) observations led him to conclude that both techniques lead to error-prone constructive memories, and he implied that errors are more likely in serial reproduction. Specifically, Bartlett observed that on repeated reproduction tasks, subjects' first reproduction persisted across subsequent recall attempts, so that in spite of many errors of omission and rationalization, the narrative of the event remained relatively consistent (Chapter V). In contrast, Bartlett observed that on serial reproduction tasks, "the cumulative recall of a very few people can result in the production of a totally new event, story or representation" (Chapter VII, p. 165). Considered together, these two statements imply that serial reproduction is more error prone than repeated reproduction.

Although Bartlett's (1932) observations regarding serial and repeated reproduction are well known and are recounted in virtually all textbooks on cognitive psychology and human memory, they have never been confirmed in a direct experimental

* Corresponding author at: Department of Psychology, Box 1125, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130-4899, United States. Tel.: +1 314 935 4307.

E-mail address: roediger@artsci.wustl.edu (H.L. Roediger III).

comparison. He treated results from the two techniques in separate chapters in his book, with repeated reproduction informing how recollections change over time within the person whereas serial reproduction was used to measure the social processes involved in passing memories across people and how the accounts of events change during this process. The purpose of the current experiment was to fill the empirical gap that Bartlett left by directly comparing the error rates produced by repeated reproduction and serial reproduction techniques. In order to do so, we had to create a technique that would permit such a controlled comparison. First, however, we review what has been learned from each testing method in the years since *Remembering* was first published. We review these literatures briefly in turn.

1.1. Serial reproduction

Kirkpatrick (1932) presented serial reproduction results like Bartlett's in the same year that Bartlett published his famous book. He had students read about news items and then pass them along orally from one to another. He found "very obvious and striking distortion" (page 198) in his results and pointed to a few tendencies similar to those uncovered by Bartlett. One may be tempted to consider this as an instance of simultaneous discovery, but that is not correct because Bartlett had previously published his memory results before they appeared in his book. Bartlett (1920) reported the essence of the experiments on both serial and repeated reproduction and indeed these observations were drawn from his 1916 (unpublished) dissertation. Kirkpatrick (1932) did not cite this earlier work, but he was a sociologist and was probably unaware of Bartlett's research as he began his own project of "experimental sociology" [page 204].

Tresselt and Spragg (1941) demonstrated that subjects in serial reproduction experiments were more likely to 'assimilate' or make story elements more familiar over time, thus distorting memory for the original story. Allport and Postman (1947) used the serial reproduction technique in their famous studies of rumor, which was a critical topic during World War II in the U.S. as rumors (some true, many false) spread throughout the populace during the war. Their results confirmed the tendency to great distortion in serial reproduction. Paul (1959) published four experiments using the serial reproduction technique that also confirmed and extended Bartlett's (1932) findings and tried to answer subsidiary questions (e.g., about individual differences in recall).

Of more direct relevance to the present experiment, Deese (1961) presented subjects with word lists with words varying in associative strength to a single word. For example, the *butterfly* list contained high frequency associates (e.g., moth, insect), low frequency associates (garden, sky) or zero frequency associates (book, tutor). Deese demonstrated that intrusions increased across three serial recall trials of the word lists. We use lists of words like those Deese (1961) used in the current experiments, except that all words were associated with a critical word in a paradigm developed by Deese (1959) and Roediger and McDermott (1995), the DRM paradigm.

In sum, virtually every experiment we can find using Bartlett's serial reproduction technique confirms his observations that social transmission of information is error prone and that the more links there are in the chain, the greater the probability of error.

1.2. Repeated reproduction

The use of repeated reproduction has led to much less consistency. In fact, at least two different traditions of research exist. Ballard (1913) began one tradition by asking school children to learn passages of poetry and asking them to recall the poems soon after learning and then up to a week later. His main finding was

that children would often remember lines of poetry on later tests that they had failed to recall on earlier tests, a phenomenon he named reminiscence. In later work, Brown (1923) reported greater overall recall on a second test than for a first test with two other types of material (word lists and states of the U.S.). These studies were conducted before the widespread use of inferential statistics, but Erdelyi and Becker (1974) confirmed the observation that recall could improve across tests with pictorial materials. Others have subsequently confirmed these observations (e.g., Roediger & Thorpe, 1978; Roediger & Payne, 1982, among many others). These studies all lead to the conclusion that recall can actually improve over time with repeated testing.

Bartlett (1920, 1932) reported a different pattern of results relative to those of Ballard (1913) and others reviewed in the previous paragraph, albeit not with aggregate statistics but with sample recall protocols. During repeated reproduction of the War of the Ghosts and other materials, Bartlett reported that his subjects recalled less over time and what they did recall was full of errors. These two traditions of research lead to a puzzle of why Ballard (1913) and Brown (1923) obtained one pattern of outcome and Bartlett (1920, 1932) a different pattern. Bartlett (1932) never cited the research whose results conflicted with his own.

Wheeler and Roediger (1992) took up the challenge of explaining the different pattern of results in the two types of paradigms and partly succeeded. They showed that with short intervals between successive tests reminiscence, or gains in recall, occur over tests; however, when long intervals occurred between tests (like those Bartlett used), forgetting occurred. However, even in these cases with long intervals, Wheeler and Roediger did not observe much distortion (even using the War of the Ghosts as material; see too Wynn & Logie, 1998).

Others wondered about Bartlett's (1932) instructions to the subjects: Had he told them to faithfully remember the story as well as possible or simply to tell a story back to him? His methods were unclear on this point. Gauld and Stephenson (1967) conceptually replicated Bartlett's increased errors with repeated retellings only when subjects were not explicitly instructed to attend to the accuracy of their retelling. This outcome would seem to indicate that Bartlett's (1920, 1932) repeated reproduction experiments might not be replicable with instructions to remember accurately (as are usually employed in memory experiments). This possibility led Bergman and Roediger (1999) to attempt a direct replication of Bartlett's original study (insofar as that is possible, given Bartlett's rather sketchy account of his methods and casual reporting of his results). They used instructions to remember in their experiment and, unlike prior research, Bergman and Roediger succeeded by demonstrating that recall does indeed become distorted over repeated retellings. They showed that after a 6-month delay most recalled information was distorted. Considered together, the evidence on repeated reproduction suggests that over long delays, even with instructions to be accurate, memories become less accurate with repeated retellings. However, the results are not as clear-cut in the repeated reproduction task as in serial reproduction.

1.3. Direct comparison of serial and repeated reproduction

To our knowledge, only one prior study has compared repeated reproduction to serial reproduction and it focused on accurate recall. Alper and Korchin (1952) presented students with prose passages that contained controversial (in 1952) statements about admitting women to universities that were historically only open to men. They demonstrated that, across 6 repeated reproductions, veridical recall of the passages remained stable. However, when the information was passed among six subjects using serial reproduction (each subject after the first one studying and recalling the

recall protocol of the prior subject), veridical recall declined significantly. Errors in this study were examined only in relation to gender differences in the qualitative types of errors produced. The authors did not examine how errors varied between serial and repeated reproduction, which was a major focus of the current experiment.

The current study is the first to directly compare the recall of both accurate and inaccurate information across repeated reproduction and serial reproduction. The primary purpose of the experiment was to test the hypothesis, derived from Bartlett's (1932) writings, that serial reproduction results in greater errors than does repeated reproduction. Toward this end, we presented participants with the Deese-Roediger-McDermott (DRM) word lists shown to elicit memory errors (Deese, 1959; Roediger & McDermott, 1995). Briefly, DRM lists contain related words such as bed, rest, tired, dream, awake, etc., that are associated to a critical item that was not presented (sleep). Participants reliably falsely remember that the critical item was actually presented in the study list, often at high levels (Roediger & McDermott, 1995; see Gallo, 2006, 2010 for reviews). We directly compared accuracy of the DRM lists recalled repeatedly by a single individual (repeated reproduction; see McDermott, 2006) to accuracy of the same lists recalled by different people on successive tests (serial reproduction; Deese, 1961). In the latter case, each new subject studied and recalled the previous subject's recall protocol.

As in Bartlett (1932), our primary interest was the extent to which the final output accurately reflected the initially presented information, as a function of repeated or serial reproduction. In making this comparison, it is important to highlight a fundamental difference between repeated and serial recall. In both conditions, we expected that the first individual would often forget some of the list words and falsely recall the critical item, due to its strong associations with the study list (i.e., the DRM illusion). In the repeated reproduction condition, because the individual is asked to recall the original study list numerous times, the repeated recall of list words and critical items would reflect the persistence of true and false memory, respectively. In contrast, because subjects in the serial reproduction condition might have received the critical item as part of their initial study list, it is somewhat of a misnomer to refer the repeated recall of the critical item in this condition as a "false recall," at least at the level of the individual subjects. Nevertheless, at the group level, the initial generation and persistence of critical items across serial reproductions does represent a memory distortion with respect to the originally presented information.

2. Method

2.1. Subjects

The subjects were 60 Washington University undergraduates who participated in the experiment for partial fulfillment of a course requirement.

2.2. Design

The design consisted of a 2 (repeated reproduction and serial reproduction) \times 4 (test 1–4) mixed-subjects design. We employ the terminology of *recall* 1–4 for repeated reproduction and *link* 1–4 for serial reproduction (because their recalls are like links of a chain). The tests referred to as Recall 1 and Link 1 in the two situations are actually exactly the same; we use the different terminology to distinguish the procedures. Repeated reproduction was manipulated within subjects so that each participant repeatedly recalled the study lists 4 times. Serial reproduction was manipulated between subjects so that four different participants recalled each list.

Table 1

Design of repeated reproduction and serial reproduction experiment. Repeated reproduction is depicted across horizontal rows (within participants). Serial reproduction is depicted in vertical columns (between participants).

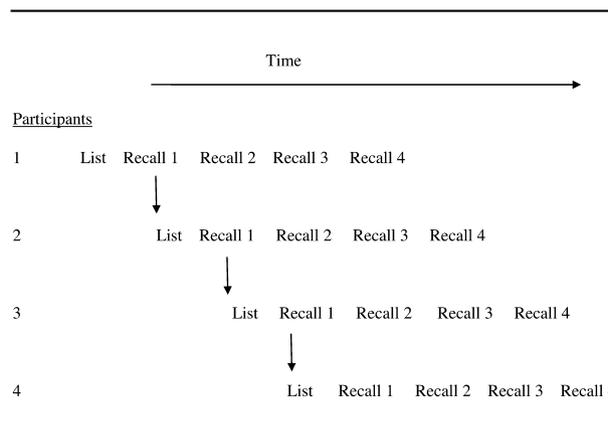


Table 1 provides a representation of the design. The first row of the table represents the repeated reproduction phase of the experiment; the rows below the first indicate the serial reproduction phases (although each of these subjects then repeatedly recalled the information. We consider only the repeated reproduction results represented by the first row in the table). The design is unusual in that one condition was manipulated within subjects (repeated reproduction) and the other variable between subjects (serial reproduction), with the added complication that the first test was the same in both conditions. The primary dependent variables were correct recall of list items and false recall of critical items.

2.3. Materials

Materials consisted of 8 lists taken from the Stadler, Roediger, and McDermott (1999) norms that were shown to elicit high levels of false recall for a critical nonpresented item (hereafter the critical items). Each list contained 15 words associated to the critical item (Roediger & McDermott, 1995). The lists were divided into two sets equated on the mean probability of false recall (Set A mean = .58 SD = 2.21; Set B mean = .57, SD = 2.21). Set A included the *smell*, *window*, *smoke*, and *rough* lists; set B included the *sweet*, *sleep*, *doctor*, and *chair* lists. Additional materials included experimental booklets containing multiplication problems and visuo-spatial puzzles that were used as filler tasks throughout the experiment.

2.4. Procedure

Subjects participated in groups of 4. All subjects were tested sequentially, so that as the first participants in each group began the study phase, the remaining participants worked on filler tasks (this aspect of the design is not shown in Table 1). In the study phase, the first subjects in each group were presented visually with sets of lists that were blocked by the themes of the list (15 items associated to *smell*, then 15 associated to *window*, etc.). Half of the subjects were presented first with set A lists; half were presented first with set B lists. Lists were always presented in the following order: Set A, *smell*, *window*, *smoke*, *rough*; Set B, *sweet*, *sleep*, *doctor*, *chair*.

Each list was presented via computer and single words were presented sequentially at a rate of 3 s/word. Subjects were told that they should pay careful attention to each word because their memories for the presented items would be tested at a later time. To ensure that subjects attended to each word, they were asked to

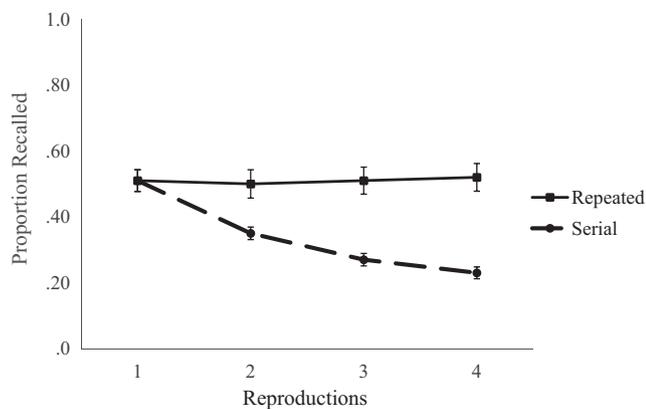


Fig. 1. Mean proportion of list items presented in original list recalled as a function of repeated or serial reproduction. Error bars represent standard error.

make a pleasantness rating (1 = not pleasant to 5 = very pleasant) for each word presented. After seeing all four lists of the first set, they were asked to complete 30 s of multiplication problems provided in the experimental booklet.

The first subjects in each group were then given 6 min to recall as many items as possible from the four presented lists by typing their responses into the computer. Typed responses remained visible on the computer screen during each recall. After completing their initial recall of the lists, subjects completed another 30-s math filler task, and were asked to recall the items again. This procedure was repeated until the subjects had recalled the lists 4 separate times. This manipulation constituted the repeated reproduction procedure.

The first recall protocol of the first subject in each group was sent electronically to the computer of the subsequent subject in the serial reproduction condition, who then underwent a procedure identical to the first subject's study phase except for the particular words studied. That is, this serial reproduction subject studied the first participant's initial recall of the list (in the order recalled by the participant), not the original set of items. This second participant then recalled her study list 4 consecutive times, with the initial recall being sent on to the third participant in the serial reproduction group. This procedure was repeated until all four subjects in the group had completed the task. After completing the repeated reproduction phase, each subject performed visual spatial filler tasks while the remaining subjects completed the entire procedure. Of course, as is clear in Table 1, the fourth serial reproduction subject was the last to finish.

Once subjects had completed the procedure with one set of lists, the entire procedure was repeated using a new set of lists (set B if set A was previously presented; set A if set B was previously presented). Subjects were tested in the same order for both sets of list (i.e., if a subject were the third subject in the sequence for one set of materials, she or he retained the same spot for the second set of materials). At the end of the experiment, all subjects were thanked and fully debriefed.

3. Results

3.1. List words

The mean proportion of list items recalled from the original study lists are presented in Fig. 1. Recall remained steady in the repeated reproduction procedure across the four tests taken by the first subject, but it declined across the four successive serial reproduction tests of four different subjects. For the latter subjects, we

scored them on the proportion of items in the original lists, not the proportion of the list of items they studied.

Separate one-way ANOVAs were conducted on repeated reproduction and serial reproduction (repeated tests could not be treated as a factor in the experiment because repeated testing was manipulated within subjects in the repeated reproduction condition and between subjects in the serial reproduction condition). The one-way ANOVA conducted on repeated reproduction data showed that recall did not differ significantly across the four successive tests, $F(3, 42) = 1.64, p > .05$. It was only when the list was passed across subjects that the number of items recalled decreased across the reproductions, $F(3, 56) = 27.00, MSE = .008, p < .001$. The first subject in each chain of the serial reproduction task recalled a significantly greater proportion of the originally presented list items ($M = .51$) than the second subject ($M = .35$), $t(28) = 4.10, SEM = .03, d = 1.50$, who in turn recalled a greater proportion than the third ($M = .27$), $t(28) = 2.86, SEM = .02, d = 1.05$. However, recall of the third ($M = .27$) and fourth ($M = .23$) subjects in each chain did not differ significantly, $t < 1.4, p > .05$. By the fourth recall trial, subjects in the repeated reproduction conditions recalled a greater proportion of the originally presented list items than those in the serial reproduction condition, $t(28) = 6.32, SEM = .05, p < .001, d = 2.31$. These findings are consistent with those of Alper and Korchin (1952) who demonstrated that veridical recall remained stable across repeated reproductions but declined across serial reproductions.

Fig. 1 presents data in terms of the proportion of words recalled from the original lists, i.e., the proportion recalled of the 60 words that were studied. However, using this criterion provides a misleading impression of performance for subjects in the serial reproduction condition, because the second through fourth subjects never studied the entire set of items but only those recalled by the prior subject. Another way to analyze the data for serial reproduction subjects is using the mean proportion of list items produced as a proportion of the number of items that each subject studied. For the first subject in each chain, those who were presented with the original study list, the data point is the same as in the previous analysis shown in Fig. 1. However, in a new analysis, subjects 2 through 4 in the serial reproduction condition were scored on the proportion of the items that they could correctly recall of those that they had studied (i.e., what proportion could subject 2 recall of those items that had been recalled by subject 1 and hence studied by subject 2?). When measured in this manner, recall increased across links from .51 to .88 from link 1 to link 4. A one-way ANOVA revealed that as the study list got shorter across links in the serial reproduction task, subjects were able to produce a greater proportion of studied items, $F(3, 56) = 27.10, MSE = .01, p < .001$. Follow up t -tests revealed that the proportion of studied items increased significantly with each serial recall, $M = .51$ for subject 1; $M = .67$ for subject 2; $M = .77$ for subject 3; $M = .88$ for subject 4, t 's $> 2.39, p$'s $< .025, d$'s $> .87$. This outcome provides another demonstration of the list length effect: The shorter the list of items studied, the greater the proportion that were recalled (Crowder, 1976).

3.2. Critical Items

The mean proportion of critical items recalled (out of 8) across recall trials is presented in Fig. 2. False recall remained relatively stable across the tests, although there appears to be a slight upward trend. Separate one-way ANOVAs were computed on the proportion of critical items recalled in both repeated and serial reproduction tasks and no significant effects emerged, F 's $< 1.4, p$'s $> .05$. The proportion of critical items produced did not vary across repeated retrieval attempts within subjects (repeated reproduction) or across subjects (serial reproduction). On the 4th test, no significant difference emerged between critical items produced on tests of repeated reproduction and serial reproduction, $t < 1, p > .05$.

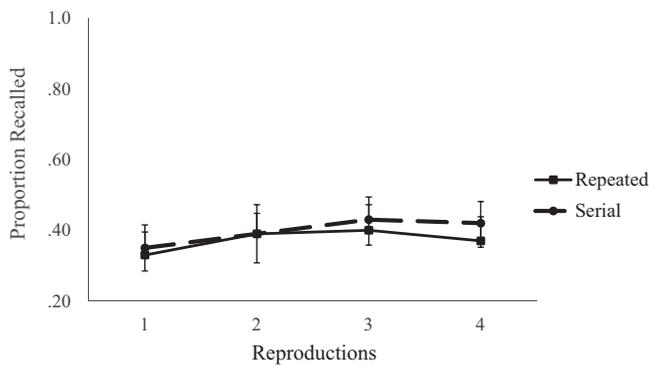


Fig. 2. Mean proportion of critical items recalled as a function of repeated or serial reproduction. Error bars represent standard error.

The data in Figs. 1 and 2 have used some form of input-bound scoring (Koriat & Goldsmith, 1994) in which proportion recalled is plotted as a function of the total possible material recalled from the entire set of materials. However, as Koriat, Pansky, and Goldsmith (2011) have pointed out, another useful way of conceiving of the recall is output-bound scoring. In this method of scoring, the issue is what proportion of recalled material is correct and what proportion is in error; that is, if 3 items were recalled incorrectly out of 10 items produced, the error rate would be .30. This technique is especially useful across delayed tests when total recall is dropping because it permits one to measure what proportion of recalled material is accurate and what proportion is in error. To follow with the example above, after a delay subjects might recall only 5 items. If 3 were errors, then the proportion of errors would be .60 by output-bound scoring. Dallenbach (1913) used such a procedure in experiments over 100 years ago, although of course he did not use this name.

Fig. 3 presents our false recall data using output bound scoring, presenting the proportion of critical items recalled as a proportion of total recall across the tests. This analysis answers the question: Of the items produced on the test, what proportion were critical items? A one-way ANOVA revealed that the relative proportion of critical items produced on the repeated reproduction task remained relatively constant within subjects, $F < 1.9$, $p > .05$. However, as recall was passed from one subject to another in serial reproduction, the proportion of critical items produced increased relative to the total output, ($M = .06$ link 1; $M = .12$ link 4), $F(3, 56) = 5.17$, $MSE = .003$, $p = .003$. Thus, as the list was transmitted from one subject to the next, overall accuracy was reduced when assessed by this measure. On recall 4, serial reproduction resulted in a significantly greater proportion of false recall than repeated reproduction, $t(28) = 3.02$, $SEM = .02$, $p = .01$, $d = 1.10$. In sum, we have shown in output-bound scoring that serial reproduction leads to more errors than repeated reproduction in recall. As noted above, however, the reference point for calling a critical item an error in

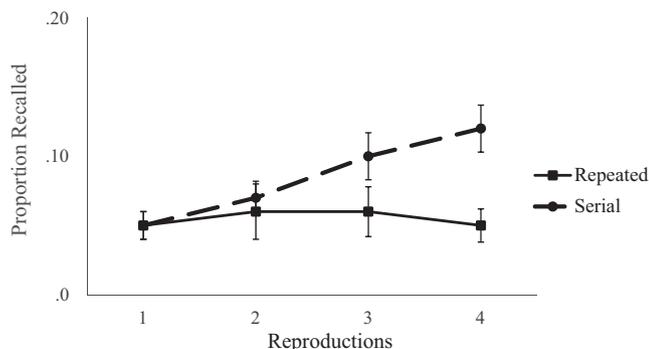


Fig. 3. Mean proportion of critical items recalled as a function of total items recalled (output bound scoring). Error bars represent standard error.

the serial reproduction experiment is the initial study list (not the immediately preceding study list for subjects 2–4 in the procedure). On some occasions, a subject will produce a critical item as an error, the next subject will study the recall protocol of the previous subject and will then “correctly” recall the critical item. Still, for purposes of comparison between repeated and serial reproduction, the scoring of errors with regard to the original studied material is appropriate.

We examined other intrusions besides critical items, but subjects made very few errors (a mean of 1 or fewer per reproduction) and no systematic variation could be observed due to scale attenuation.

4. General discussion

The current experiment was the first to directly compare both veridical and false recall in repeated and serial reproduction paradigms. Both recall of list items and recall of critical items remained stable across repeated reproductions (or repeated recalls) of the list. Subjects produced fairly stable recall protocols in terms of overall list recall with repeated reproduction, with no change across tests with these short intervals between tests. However, when information was transmitted between individuals in the serial reproduction task, a different pattern emerged. In this case, subjects recalled fewer of the original list items as the list was passed from subject to subject. Yet because fewer items were transmitted in each link of the chain in serial reproduction, the proportion of list items recalled relative to the number studied actually increased across the four links.

For critical items, the overall proportion recalled did not vary across serial reproduction tests using input-bound scoring (Fig. 2). On the other hand, using output-bound scoring, the proportion of critical items recalled increased significantly as the list passed between subjects, because the list became shorter with each successive transmission and thus recall of critical items as a proportion of all items recalled increased (Fig. 3). Thus, by the fourth recall test, serial reproduction resulted in lower output-bound accuracy than repeated reproduction.

The finding that veridical recall remained constant across repeated reproduction and declined with serial reproduction is consistent with previous work that examined repeated and serial reproduction separately. Our results are broadly consistent with Bartlett's (1932) conclusion that repeated reproduction results in a “persistence of form” such that there is relative stability across repeated retellings. Our results are also consistent with Alper and Korchin's (1952) findings showing that serial reproduction led to greater forgetting of prose passages than repeated reproduction. Research in other traditions confirms this basic pattern. For example, in research on flashbulb memories, people remember the core event quite well, but often do not correctly remember the moment of occurrence. That is, over repeated retellings, some rememberers report a different initial moment of receiving the news of the event even though they are confident that it is correct (Hirst et al., 2009; Talarico & Rubin, 2003).

The primary purpose of our experiments was to attempt to confirm the observation that serial reproduction leads to greater distortion of the originally presented material than repeated reproduction, implied if not explicitly stated in Bartlett's (1932) writings. We did show this using output-bound scoring, which is also the way Bartlett considered his results (albeit again through examples of protocols). Bartlett noted that over repeated retellings, the amount of correct information dropped while the proportion of erroneous information increased. Bergman and Roediger (1999) replicated this finding in repeated reproduction using the War of the Ghosts as material and with a retention interval of six months

before a final test. We did not obtain that outcome in the current experiment with repeated reproduction, probably because we used short delays between tests. However, we did obtain greater distortion with serial reproduction under these same conditions, and therefore obtained greater distortion from serial than repeated reproduction. Finding increased errors across serial reproduction is consistent with prior work using this task. Across a range of stimuli from word lists (Deese, 1961), news articles (Kirkpatrick, 1932), rumors (Allport & Postman, 1947), and prose passages (Bartlett, 1932), serial reproduction has been shown to consistently increase errors across the retellings. However, we should note that with very long intervals between tests, repeated reproduction may become more like serial reproduction with many errors introduced relative to the amount correctly recalled (Bergman & Roediger, 1999).

The current experimental finding that errors did not increase with repeated reproduction (using either input-bound or output-bound scoring) is especially interesting in relation to previous literature. Most notably, the finding seems inconsistent with Bartlett's original claim that repeated reproduction increases memory distortion. However, as noted above, the most likely explanation for why the current experiment failed to obtain increased errors across repeated reproductions is that we used short intervals between tests (Wheeler & Roediger, 1992). As noted above, with long intervals between tests like those Bartlett (1932) used, Bergman and Roediger (1999) found increased distortion with repeated tests.

Besides confirming the observation that serial reproduction leads to greater forgetting and more errors than repeated reproduction, the other novel feature of our experiment is in designing an experimental method to directly compare Bartlett's two methods of testing (Table 1). As noted, our solution to the vexing problem of comparing these two methods is not perfect, because it produces results that do not lead to straightforward statistical analysis (e.g., employing a 2×4 ANOVA to analyze the results). Nonetheless, we hope the design will prove useful in future studies, and indeed Nunes (2012) has already replicated and extended the results we report here in a novel use of the paradigm involving impressions of personality from lists of traits.

In sum, we have provided results that confirm Bartlett's (1932) observations concerning the different effects of repeated and serial reproduction, at least insofar as we found greater forgetting and distortion from the latter. In addition, we have provided a paradigm for directly comparing repeated and serial reproduction that should be useful for future work. Bartlett emphasized the role of schema in explaining his results, whereas our materials (word lists) may seem resistant to this concept. However, subjects are clever in finding (or creating) organizational structures for word lists even when the words are unrelated (Mandler, 1967; Tulving, 1962). Our lists contained related words that were organized around the critical nonpresented words and thus they had a thematic, if not schematic, structure. Thus we believe that Bartlett's ideas (again, broadly speaking) can be applied in the case of our experiments, too. People organize their experiences in order to enable later retrieval, and such organizational processes (whether via schemas or associative structures) can both enhance accurate recall and lead to schema- or associative errors in recall (Roediger & DeSoto, 2014).

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgment

We thank Angela Wong for help designing the computer program and testing subjects. The research was presented at the annual meeting of the Psychonomic Society in 2008.

References

- Allport, G. W., & Postman, L. (1947). *The psychology of rumor*. New York: Holt.
- Alper, T. G., & Korchin, S. J. (1952). Memory for socially relevant material. *The Journal of Abnormal and Social Psychology*, 47, 25–37.
- Ballard, P. B. (1913). Oblivescence and reminiscence. *British Journal of Psychology Monograph Supplements*, 1, 1–82.
- Bartlett, F. C. (1920). Some experiments on the reproduction of folk-stories. *Folk-Lore*, 31, 30–47.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge England: Cambridge University Press.
- Bergman, E. T., & Roediger, H. L., III. (1999). Can Bartlett's repeated reproduction experiments be replicated? *Memory & Cognition*, 27, 937–947.
- Brown, W. (1923). To what extent is memory measured by a single recall? *Journal of Experimental Psychology*, 6, 377–382.
- Crowder, R. G. (1976). Principles of learning and memory. Hillsdale, NJ: Erlbaum.
- Dallenbach, K. M. (1913). The relation of memory error to time interval. *Psychological Review*, 20, 323–337.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58, 17–22.
- Deese, J. (1961). Associate structure and the serial reproduction experiment. *Journal of Abnormal and Social Psychology*, 63, 95–100.
- Erdelyi, M. H., & Becker, J. (1974). Hypermnnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology*, 6, 159–171.
- Gauld, A., & Stephenson, G. M. (1967). Some experiments related to Bartlett's theory of remembering. *British Journal of Psychology*, 58, 39–49.
- Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. New York, NY: Psychology Press.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, 38(7), 833–848.
- Hirst, W., Phelps, E. A., Buckner, R. C., Cue, A., Gabrieli, D. E., & Johnson, M. K. (2009). Long-term memory of the terrorist attack of September 11: Flashbulb memories, event memories and the factors that influence their retention. *Journal of Experimental Psychology: General*, 138, 161–176.
- Kirkpatrick, C. (1932). A tentative study in experimental social psychology. *American Journal of Sociology*, 38, 194–206.
- Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, 123, 297–316.
- Koriat, A., Pansky, A., & Goldsmith, M. (2011). An output-bound perspective on false memories: The case of the Deese–Roediger–McDermott (DRM) paradigm. In A. Benjamin (Ed.), *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork* (pp. 301–332). London, England: Psychology Press.
- Mandler, G. (1967). Organization and memory. In K. W. Spence, & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press.
- McDermott, K. B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition*, 34, 261–267.
- Nunes, L. D. (2012). *Conditional activation of associative semantic structures: Forming and transmitting impressions of personality* (Doctoral Dissertation). University of Lisbon.
- Paul, I. H. (1959). *Studies in remembering: The reproduction of connected and extended verbal material*. *Psychological Issues* (Vol. 1) (issue 2).
- Roediger, H. L., & DeSoto, K. A. (2014). Confidence and memory. Assessing positive and negative correlations. *Memory*, 22, 76–91.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 21, 803–814.
- Roediger, H. L., & Payne, D. G. (1982). Hypermnnesia: The role of repeated testing. *Journal of Experimental Psychology: Learning Memory and Cognition*, 8, 66–72.
- Roediger, H. L., & Thorpe, L. A. (1978). The role of recall time in producing hypermnnesia. *Memory & Cognition*, 6, 296–305.
- Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, 27, 494–500.
- Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science*, 14, 455–461.
- Tresselt, M. E., & Spragg, S. D. S. (1941). Changes occurring in the serial reproduction of verbally perceived materials. *The Journal of Genetic Psychology*, 58, 255–264.
- Tulving, E. (1962). Subjective organization in the free recall of unrelated words. *Psychological Review*, 69, 344–354.
- Wheeler, M. A., & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, 3, 240–245.
- Wynn, V. E., & Logie, R. H. (1998). The veracity of long term memories: Did Bartlett get it right? *Applied Cognitive Psychology*, 12, 1–20.