The One-Trial Learning Controversy and Its Aftermath: Remembering Rock (1957)

HENRY L. ROEDIGER III and KATHLEEN M. ARNOLD
Washington University in St. Louis

In 1957 Irvin Rock published an article in the *American Journal of Psychology* igniting a controversy that dominated the field of verbal learning for the next 8 years before mostly burning out. Rock published 2 paired-associate learning experiments in which he compared performance of a control group that learned a constant list of pairs to the criterion of one perfect trial with an experimental group in which forgotten pairs on each trial were dropped and replaced on the next trial with new pairs. That is, on each trial for experimental subjects, pairs that were correctly recalled were maintained in the next trial, whereas pairs that were not recalled were dropped and replaced randomly with new pairs from a large pool. Surprisingly, Rock found that the 2 groups took the same number of trials to reach criterion. He concluded that learning occurred not with a gradual, incremental increase in strength of memory traces but rather in an all-or-none fashion. Rock's conclusions rocked the world of verbal learning, because all theories followed a gradualist assumption. However, Estes (1960) published research that led him to the same conclusion shortly thereafter. We recount these developments and discuss how the verbal learning establishment rose up to smite down these new ideas, with particular ferocity directed at Rock. Echoing G. A. Miller (1963), we conclude with a note of sympathy for Rock's and Estes's positions and muse about why their work was so summarily dismissed. The important question they raised—the nature of how associations are learned—remains unanswered.

In 1957, Ghana obtained its independence, the Soviet Union launched *Sputnik 2,* the Wham-O Company produced the first Frisbee, and the Ford Motor Company produced the Edsel (on what the company proclaimed was E-Day, a takeoff on D-Day 13 years earlier). The Edsel was going to make a splash, the Ford marketing people thought, and it did—just not the kind desired.

Meanwhile, in the world of experimental psychology, the *American Journal of Psychology* launched a rocket of its own, an article by Irvin Rock titled "The Role of Repetition in Associative Learning" in that same year. Rock declared independence from old ways of thinking about learning as a function of repetition, and his article received a warm reaction. Depending on one's point of view, it soared like *Sputnik,* fluttered like a Frisbee, or flopped like the Edsel (a car whose production became synonymous with failure for the next 30 years). Nonetheless, Google Scholar reported (in February 2012) that Rock's ar-

American Journal of Psychology
Summer 2012, Vol. 125, No. 2 pp. 127–143 •

ticle has been cited 200 times, so it made a splash of some sort in the ocean of psychology.

The aim of our article, which is appearing in the 125th anniversary volume of this journal, is to tell the story of Rock's article, its prequel and its sequel, and to say a few things about the man himself. We provide a retrospective look at the controversy and its aftermath. Rock's article asked a critical question—one still unanswered today—although its methods have been severely criticized. Nonetheless, the article and others on the same issue (Estes, 1960) ignited a line of research that continues to this day. Although not many current articles cite the 1957 article, some owe a debt to it.

No issue so gripped American psychologists in the 20th century as the nature of learning. From the discoveries of Pavlov and Thorndike early in the century, to the dominance of behaviorism in midcentury, to the cognitive revolution late in the century, learning was a central (perhaps *the* central) topic. All psychology students in the midcentury learned about work of the great learning theorists. Besides Pavlov and Thorndike, there were the often conflicting ideas of Guthrie, Hull and Spence, Skinner, Tolman, and Estes, to name just some of the leading figures. Most of the debates revolved around animal learning and conditioning paradigms that featured mice, rats, cats, dogs, and many other creatures that learned responses through classical and operant conditioning techniques. A rich tradition of researchers studying human learning and memory thrived too, although their enterprise was often overshadowed by (and borrowed ideas from) the animal work. Bower and Hilgard (1981) provided an excellent overview of these various traditions in both animal learning and human verbal learning (see also Deese & Hulse, 1967, for a portrait of these fields in their heyday).

The central fact in the study of learning is the learning curve, a graph showing performance (plotted on the ordinate) as a function of the number of trials on some task (plotted on the abscissa). If a researcher aggregates over many subjects learning to perform a task over many trials, the outcome is a beautifully regular, negatively accelerated curve (see Figure 1 for an example from paired-associate learning). The task for the theorist is to explain why learning is always faster on the first few trials of a task and then shows diminishing returns later in practice (even if performance is far from ceiling). Why do learning curves go up as they do? Our story about Rock's (1957) article starts with this question.

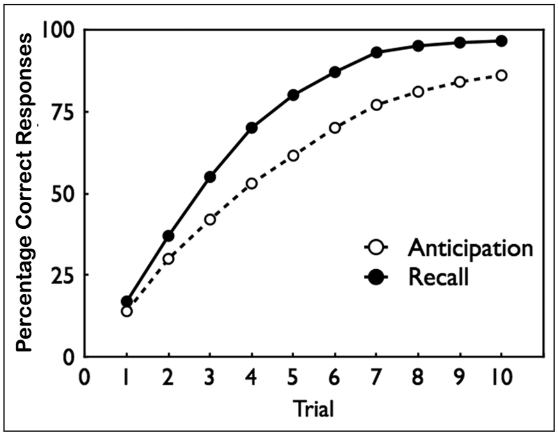

FIGURE 1. Learning curves in two paired-associate procedures. Subjects learned pairs composed of nonsense shapes as stimuli and 2-digit numbers as responses to the criterion of one perfect trial. One group of subjects learned by using the anticipation method (the stimulus was presented for 5 s with a response requested, and then the stimulus–response pair was presented as feedback). The other group learned by the recall method, in which they studied all stimulus–response pairs and then were tested by being given stimuli with responses requested. This procedure alternated until subjects reached criterion. As can be seen, the shape of the learning curve is the same in the 2 conditions, but the recall method led to better performance than the anticipation method. Data are adapted from Battig and Brackett (1961)

The Continuity–Noncontinuity Issue in Theories of Learning

Learning curves (like forgetting curves) are smooth and beautiful, and psychologists with a mathematical bent can have a field day fitting equations to them. But what processes underlie learning, and how do these operate? Because learning curves in so many domains have the same general shape, can we develop a general theory of learning that would transcend tasks and even species? This was the hope of the behaviorists in the mid-20th century, who developed general theories of learning.

One basic assumption for many researchers, from Ebbinghaus (1885/1964) to Hull (1943), is that learning reflects the gradual buildup of strength of a representation. Hull postulated "habit strength" as the basic underlying entity that was responsible for learning, and in doing so he carried forward a figure of speech embedded in our language (some memories are said to be strong and others weak). In his famous studies on the effects of repetition on savings, Ebbinghaus put the matter this way:

> These relations [between repetition and performance] can be described figuratively by speaking of the series as being more or less deeply engraved on some mental substratum. To carry out this figure: as the number of repetitions increases, the series are engraved more and more deeply and indelibly; if the number of repetitions is small, the inscription is but surface deep and only fleeting glimpses of the tracery can be caught; with a somewhat greater number the inscription can, at least for a time, be read at will; as the number of repetitions is still further increased, the deeply cut picture of the series fades out only after longer intervals. (pp. 52–53)

If we were to replace "mental substratum" with "memory traces" or "engrams" and update some other language, the statement would say that repetitions create stronger (or deeper) memory traces or engrams. The fact that the learning curve shows a gradual increase in performance is a reflection of the underlying mechanism—the buildup of strength—which is itself also gradual. The remarkable aspect of the learning curve (and the same holds for the forgetting curve) is that its general appearance is the same across astonishingly different experimental situations and dependent measures, and across species from slugs to humans. The generality of the function gave hope to the generality of mechanisms underlying the function—of the generality of laws of learning (and forgetting). This idea was an article of faith in the behaviorists' era, with their bold belief that by studying learning of nonhuman animals, they could immediately generalize the discovered findings and principles to human learning. One central assumption in most theories was that learning reflected a continuous buildup of strength of underlying physical traces of experience, or memory traces. In experiments with humans, the gradual buildup of strength was often coupled with an assumption of a response threshold; once enough strength accrued in the underlying trace, it would support the subject's responding.

Yet there were dissenters from this picture nearly from the start, researchers who argued that although learning curves were continuous, the underlying processes were anything but continuous. Yerkes (1916) and, more famously, Köhler (1925) observed chimpanzees that seemed to have insight into a problem and to learn its solution all at once; they did not seem to learn gradually over many trials. From a somewhat different direction, Guthrie's theory of learning assumed that "a stimulus pattern gains its full associative strength on the occasion of its first pairing with a response" (1942, p. 30). Although it takes us too far afield for present purposes, Guthrie's theory of learning could explain the apparent gradualness of learning over trials even with this strong assumption.[1]

The debate over whether learning was continuous or discontinuous continued in the animal literature, particularly in discrimination learning (learning to discriminate several stimuli and link them to responses). Researchers used clever experimental techniques and arguments that bolstered one or the other side of the argument (e.g., Krechevsky, 1938; Spence, 1945). By the early 1950s this debate had subsided, because Harlow (1949) showed in work with monkeys that discrimination learning could occur either in a seeming all-or-none fashion or in a gradual fashion, depending on the animals' level of experience. Monkeys who were inexperienced on a problem appeared to learn in a gradual, step-by-step fashion; however, monkeys who were experienced in discrimination learning by having solved hundreds of previous problems could solve a new problem in a sudden manner, often on a single trial. Thus the debate about whether learning was continuous or discontinuous subsided in a way that has often occurred in the history of experimental psychology: After a long debate about whether *X* or *Y* is the correct answer to a problem, and after the proponents of each side have waged vigorous debate on the matter, the eventual conclusion is that both are right in different situations. This verdict is what has been called the Dodo Bird Verdict from *Alice in Wonderland*. The dodo bird commissioned a race but forgot to measure speed, distance, or any other feature of performance. When asked who won, the dodo bird thought long and hard and then finally proclaimed "Everybody has won and all must have prizes." So both the continuity and noncontinuity theorists could claim victory and take their prizes, such as there are in the academic community.

Rock's Article on One-Trial Learning

Rock (1957) did not refer to any of the aforementioned debates about learning in his article (although he undoubtedly knew about them); in fact, he cited only five prior articles in his 1957 article, one of which was an unpublished BA thesis and another that was the norms of nonsense syllables from which he selected his items. Thus the question he asked seemed novel in the context of just his article, even though a long history on the general issue preceded it. An article by Estes (1960) also helped to reopen the continuity–noncontinuity debate, and we refer to it later.

Before we get to Rock's (1957) experiment, let us pause to marvel that he conducted it at all. Irvin Rock (1922–1995) received his PhD in 1952 from the New School for Social Research in New York City, studying under Hans Wallach (a psychologist interested primarily in perception). Almost all of Rock's great work in psychology over the course of his long career had to do with perception. He wrote five books and published many journal articles on the topic, making important discoveries and training many students who went on their own great achievements (Gilchrist, 1996). His articles that began the one-trial learning controversy (Rock, 1957, and Rock & Heimer, 1959) were one of his only forays into the study of learning and memory, and, as we shall see, he was roundly criticized for his work—probably too much so. The experiments themselves were examples of excellent science, for Rock took a counterintuitive idea—that learning occurs all at once on a single trial—and provided straightforward and elegant tests of the hypothesis. The criticism that eventually tarnished his work was one he was well aware of in writing the 1957 article; he simply discounted its importance. But we are getting ahead of ourselves. We first need to review his research before we get to its critics.

Rock (1957) studied paired-associate learning, the most fashionable task of his era for the study of verbal learning, echoing as it does classical conditioning (in both paradigms a stimulus elicits a response after learning). In fact, the stimulus–response terms developed in the study of classical conditioning were carried over to the paired-associates task during the verbal learning era, although development of the paired-associates task (Calkins, 1894) actually predated Pavlov's study of the conditioned response (which seems to have occurred around 1900, although news did not spread to U.S. psychologists until some years later).

Rock sketched out two positions for the learning of associations in paired-associate learning: "One possibility is that in learning a list of items, the strength of association between each pair develops gradually, with each repetition adding an increment to the bond, until it is so strong that the first item produces recall of the second" (1957, p. 186). This is the classic view of learning as a gradual process, and in this view repetition aids the *formation* of associations. However, "another possibility is that repetition is essential because only a limited number of associations can be formed in one trial. From this point of view, associations are formed in one trial, and improvement with repetition is only an artifact of working with long lists of items" (p. 186). This latter view is similar to that of Guthrie, but with a twist: Not all associations are learned on every trial but only a subset of them. That subset is learned perfectly, but the rest of the associations that were presented are not learned at all (no strength accrues to the association for those items). The "artifact" Rock referred to is essentially that of averaging across many subjects learning many lists on many trials; despite the all-or-none nature of the underlying process, the learning curve will be smooth when performance is averaged over these several parameters.

How could one test between these ideas? Rock (1957) hit on an ingenious idea. Let us use his words to tell the story:

> There are several ways of testing this assumption, but the method used in the experiments to be reported here seems most direct. A control group is given the task of learning a list of paired associates to a criterion of one errorless trial. An experimental group is handicapped by removing all pairs which S fails to get right after every trial and substituting new pairs for them. The new pairs are randomly selected from a pool of pairs prepared in advance and from which the initial lists for the two groups are also randomly selected. This means that the experimental group always has the same number of words to be learned on any given trial—the same number as has the control group—but only some of them will have been seen previ-

> ously (those already learned) and some will never have been seen previously. Training of the experimental group is also continued to a criterion of one errorless trial. For the experimental group, then, the pair is either learned the first time it is seen, or it is removed, and S does not, therefore, have what might be presumed to be the benefit of the repetition in *forming* associations. (pp. 186–187)

The prediction is clear: If associations gradually build up, the control group should learn to criterion much faster than the experimental group, because items stay constant in this condition and partial strength between associations can be formed. The experimental group, in which new pairs are constantly introduced for pairs that could not be recalled, should be penalized if learning is gradual. However, if pairs are learned in an all-or-none fashion on a single trial, there would be no penalty; no gradual associations have been built up for nonrecalled items. Thus the all-or-none theory predicts no difference between the experimental and control conditions.

Rock (1957) conducted two experiments that conformed to this logic. Subjects were presented lists of paired associates on 3- × 5-inch index cards and tested by being given the left-hand member and asked to produce the right-hand one.

In Experiment 1 the pairs were letter–number pairs such as M–38 or B–7 (thus both the stimuli and responses were highly familiar, but their association was novel). Subjects studied 12 pairs and then were tested (the backs of the 12 cards had the stimulus terms). Such separate study and test phases are typical of paired-associate learning (then and now), although other variations are possible (see the caption of Figure 1). In the control condition, the same 12 pairs were learned to criterion, but in the experimental group the experimenter sorted the pairs into two piles (right and wrong) during the test. After the test phase, the pairs gotten wrong were removed, a new set equal in number was randomly selected from the total set of remaining pairs, and the subjects studied and were tested on this new set. This process was continued until subjects reached criterion.

The data were scored two different ways, but the conclusion was the same for both methods of scoring: There was no difference between the experimental and control groups in either the median trials to criterion (TTC) or the semi-interquartile range (a measure of dispersion often used with skewed distributions. Some subjects never reached criterion, and so Rock's distributions were skewed). When all data were included (giving a value of 10 trials for subjects who never reached criterion), the median trials to criterion was 4.75 in both conditions. If the subjects who failed to reach criterion were eliminated (5 of 25 in the experimental condition and 3 of 25 in the control), the mean TTC was 4.55 for the control group and 4.35 for the experimental group. The data ran in the contrary direction to the incremental strength hypothesis, but of course this difference was not statistically significant. Still, there was no hint of support for the continuity–incremental hypothesis in the data.

The logic of Rock's (1957) second experiment was the same as the first, but the materials differed. In Experiment 1 he used digit–letter pairs to avoid subjects having to learn unfamiliar stimulus and response terms and to integrate them, which would have been required had he used nonsense syllables typical of research at the time. Repeated presentation using nonsense syllables in the control group might have given them an unfair advantage due to repeated familiarization of stimuli and responses that would be unrelated to the ability to associate them (which is the process under debate in the one-trial learning issue). Nonetheless, in Experiment 2 Rock switched to pairs of nonsense syllables, commenting, "There is also something to be said for using this more traditional material because of the greater possibility afforded for comparison with previous work" (p. 189).

Experiment 2 used the same experimental design as Experiment 1 but with nonsense syllables taken from Glaze's (1928) norms, using ones of medium associative value, by which Glaze meant how easily they could be associated with an English word (e.g., *BAJ, SOH, FAZ, WIH*). Eighty pairs were created and constituted the total set from which items were randomly selected. The list length was 8 rather than 12 pairs to start with, because of the greater difficulty of learning and associating nonsense word pairs than letter–digit pairs. Fifteen subjects reached criterion in each of the two conditions. Once again, the TTC was identical for the two groups at 8.1, and Rock said that the learning curves were the same, too. Yet again, the control group and the experimental group did not differ significantly in any way. The scientific logic of affirming the null hypothesis as evidence for a theory is always a risky business. However, most scientists agree that if a theory makes a strong prediction for a difference in performance and if well-designed experiments fail to produce such a difference, then the experiments should be taken seriously as evidence against the theory. Thus the critical question about Rock's (1957) experiments is, Were they well designed? The answer, we shall see, is "yes and no," but the "no" opinion won the day.

Before getting to that story, we need to consider Rock's (1957) discussion of his experiments. He noted that there were several reasons to expect superior learning from the control group relative to the experimental group, irrespective of the all-or-none or incremental considerations (see pp. 190–191). One is that the control group should have the advantage of general familiarity, especially in his second experiment with nonsense syllables. Repeated presentation (rather than replacement of nonrecalled pairs) should aid performance just by making the to-be-associated items more familiar. A second reason was the use of a recall test rather than some form of recognition test such as an associative matching test (which would have been difficult with the dropout procedure before computers were used in research). Recall tests give no advantage to partial knowledge (e.g., getting two of three letters correct in a nonsense syllable), whereas subjects might have been able to get the correct item in a matching test based on this partial knowledge. Rock argued that a recall test weighed against the experimental group with its constantly changing set of materials because they could not gain response familiarity for the unknown items, as could the control subjects. "That the control groups were not superior despite this possible advantage, however, makes the case against the incremental theory all the stronger" (p. 190). Finally, Rock argued that a third reason one might have expected superior performance for the control groups was due to a practice he applied to the experimental group: If a pair for the group was recalled correctly on one trial (indicating learning) but then missed on a later trial, that item was eliminated like any other missed item (despite evidence that it had been learned). This did not happen often, and it usually happened early in the learning sequence, but on such occasions the experimental group was penalized (by having the suddenly forgotten item removed despite prior evidence of learning) whereas the control group was not. Thus there were three reasons besides gradual strength accruing to associations that, Rock argued, should favor the control condition.

Rock (1957) concluded, "The present results seem to support the thesis that, in the classical multiple-item learning situation, associations are formed in one trial" (p. 190). Also, "It is as if pairs which are not retained by the time of the test leave nothing in the nervous system of any value for future use" (p. 192). He also argued, years ahead of his time, that the way subjects learn associations is by mnemonic devices. "The successful uses of such devices may mean than an idea suddenly occurs to S which enables him to link two items then and there; it has, to some extent, the character of insightful learning" (pp. 191–192). Furthermore, "The theoretical significance of the wide-spread use of such devices in rote learning experiments has not been sufficiently emphasized in the past" (p. 191). We return to these points later.

Toward the end of his discussion, Rock (1957) tried to reconcile his results with the well-known effects of overlearning. If subjects keep studying and being tested on lists after they have been learned to criterion of one perfect test trial, the greater the amount of overlearning, the better is retention when tested later. Surely repetition is having an effect here. Rock's answer was essentially that repetition does not aid in the formation of associations (which occurs in an all-or-none manner) but only aids performance afterwards. Specifically, "repetition does not seem to be of value in forming associations. Hence, it must be concluded that, in overlearning, only the repetition after the association is formed is effective in strengthening it" (p. 193). Based on more recent data, we might argue that overlearning procedures produce their benefit through repeated testing of pairs rather than through their repeated study (Karpicke & Roediger, 2008). In fact, in many contemporary studies of paired-associate learning, massive numbers of study trials lead to little or no effect on delayed tests, and the argument has been made that it is repeated testing and not repeated studying that is critical in paired-associate learning (Karpicke, 2009; Karpicke & Roediger, 2008). But that is a tale for another time.

In sum, Rock's (1957) brief (eight-page) article opened a controversy that was hotly debated in the next decade in experimental psychology. The experiments were clear, compelling, and interesting. What's not to like in such studies?

The Critics

Some scientists, maybe most scientists, in every field dislike new ideas, at least ones that are not their own. Rock's (1957) experiments challenged the wisdom of the day—the gradual, cumulative learning assumption of Clark Hull (1943, 1952)—and the leaders of the field rose up to smite him down. He was young, trained in perception, and was an outsider to the verbal learning research community. Also, he was making interesting claims that the field's leaders considered wrong, and yet his results seemed clear and persuasive. And Rock did not cite the leading figures in the field (or much of anyone else). With a few important exceptions (Estes, 1960; Estes, Hopkins, & Crothers, 1960; Miller, 1963), most articles published about Rock's (1957) work criticized every element of it but focused on one aspect in particular.

Before considering the critics, we should note that it is generally a positive feature of scientists that they tend not to believe each other's work. If some researcher makes a novel, important claim about some topic (e.g., cold fusion in physics or recognition failure of recallable words in cognitive psychology), one can rest assured that other scientists will be drawn into the fray trying to show that the claim is wrong and that the scientific world can go back to believing what it believed before. This is exactly the way science should proceed, of course. Although as scientists we like to think of ourselves as rolling back the frontiers of knowledge, we spend much of our time reviewing other researchers' work and (often) advising editors not to publish it. We also provide more stringent tests of others' ideas than we often provide of our own. Yet these factors make science appropriately conservative. Only if new ideas run the gauntlet of other scientists' replication attempts (for confirmation of the original results) and their further tests of the ideas' critical implications will those ideas be accepted into the canon of the field. It is often a messy (and sometimes vitriolic) business. That was the case in the reception of Rock's (1957) experiments.

During the era in which Rock published his learning results, two of the biggest names in traditional verbal learning research were Benton J. Underwood and Leo J. Postman. Both went after Rock's (1957) experiments with guns blazing, but they had plenty of company. By 1962, the traditionalists had regrouped and had done their research, and at least five articles appeared that year criticizing Rock's results and conclusions (Battig, 1962; Postman, 1962; Underwood & Keppel, 1962; Underwood, Rehula, & Keppel, 1962; Wollen, 1962). The criticisms were varied, but here we will stick to the main one: the claim that an item selection artifact clouded his results. Today we would probably say that the experiments involved a possible subject-by-item selection artifact, which is even worse than a simple item selection artifact because it is impossible to fix.

The basic idea is that the experimental condition was compromised in Rock's (1957) design by this artifact. The essence of the argument is that the pairs subjects could not recall after presentation on one trial—by definition, hard pairs—were replaced with pairs that subjects could recall (eventually, if they met criterion). Therefore, the new pairs might be systematically easier to learn and recall than the old pairs. If this claim is true, then subjects in the experimental condition received a set of pairs that was generally easier than those in the control condition. This factor would represent a confounding of item difficulty and experimental condition and therefore possibly undermine Rock's conclusions.

Who first noticed this possible difficulty in the experimental design? The answer is that Rock did in 1957, and he provided a thorough paragraph about it in his original article. He laid out the logic well and commented, "It is not easy to deal with this objection experimentally if difficulty is defined idiosyncratically, because the only way of finding out about a pair is to present it to that *S* for learning. Work on this problem is now in progress" (p. 191). Rock also noted that he thought the possible criticism was unlikely to be true because the materials he used were arbitrary pairs, were homogeneous, and were selected randomly.

The work referred to as in progress in 1957 came to fruition in an article by Rock and Heimer (1959, also in the *American Journal of Psychology*, as was much of the rest of this debate). In Experiment 4 of this later article, Rock and Heimer attempted to equate item difficulty across the control and experimental conditions by doing exactly what Rock (1957) said was necessary to determine idiosyncratic differences: They presented all the items to the subjects for learning. This procedure also helps to overcome the subject-by-item selection artifact, to which Rock alluded in the quote earlier. Specifically, Rock and Heimer had subjects study 7 two-digit number pairs (e.g., 49–75) one time before taking an initial test and study 12 pairs four times before taking an initial test. The lengths of the lists were chosen so that a roughly equal percentage of the items would be retrieved on each initial test for the two lists. In this way, Rock and Heimer argued, the unlearned items in each list would be of equal difficulty: "If on the average roughly one-third of the list of seven pairs is learned in one trial and roughly one-third of the list of 12 pairs is learned in four trials, then it can be presumed that those unlearned in both lists are of about equal difficulty. In both cases, the unlearned pairs would constitute the most difficult two-thirds of the list" (p. 10). In this way, the unlearned pairs in each list would be of roughly equal difficulty; one third of each list was dropped, so the remaining two thirds were assumed to be the same difficulty because all had been randomly selected at the outset. Thus, two groups of similar items could be compared, one group having been presented four times and one group only once.

After taking the initial tests, subjects again studied and were tested on eight previously studied but unlearned items. This new list was created by randomly choosing four unlearned items from each original list. Because Rock and Heimer (1959) argued that there were no differences in item difficulty between the unlearned items from each original list, the only remaining difference between the items was the number of times they had been studied. According to the incremental theory, items that were studied four times should have built up more associative strength than items studied only once and should therefore have been easier to learn on the final study trial. In contrast, the all-or-none theory would argue that no association strength had built up for any of the items in either original list, so therefore the number of previous presentations should have no effect on learning on the final trial.

Rock and Heimer (1959) conducted this experiment twice. In both experiments a "considerable number" (p. 11) of subjects were eliminated because they did not learn enough items during the initial study trials or because they did not learn any items in the final study trial. In the first experiment, the remaining 20 subjects learned 33% of the once-studied items and 44% of the repeated-studied items during the initial study trials. According to Rock and Heimer, even though these percentages were not identical, "they are sufficiently close to warrant the statement that on the average, for both lists, the easiest third (approximately) of the pairs was eliminated" (p. 11). On the final test, there was no significant difference between the number of once-studied ($M = 0.80$) and repeated-studied ($M = 1.05$) items retrieved, supporting the predictions from the all-or-none theory.

Yet even though there was no significant difference in final recall between the groups, there was a trend that suggested that if more subjects had been tested, an advantage might have emerged for repeated study items. For this reason, Rock and Heimer (1959) decided to repeat the experiment with a new set of 20 subjects. This replication was identical except that, whenever possible, only pairs that were incorrect on the initial study trial because of an error of omission (i.e., no response was given) were included in the final study trial. This was done because errors of commission (i.e., incorrect responses) during the initial study trial may have resulted in erroneous associations, which would presumably make learning the correct associations more difficult. In this replication, the percentage of items learned from each list during the initial study trials was more even; 34% of the once-studied items and 37% of the repeated-studied items were retrieved on the initial tests. Unlike the first replication, on the final test there was no hint of an advantage for the repeated study items; for both the once-studied and repeated-studied items, the mean number of items retrieved was 1.1 out of 4. When the two experiments were combined, there was still no significant difference between once-studied ($M = 0.95$) and repeated-studied ($M = 1.08$) items. These results showed no advantage for items that were repeatedly studied, suggesting, contrary to the incremental theory, that no associative strength accrued during the initial study trials.

It is worth noting that two other sets of experimental reports replicated Rock's (1957) results (and both were in the *American Journal of Psychology*).

Wogan and Waters (1959) used Rock's original procedure and found no difference between the experimental and control groups in either the number of trials needed to reach criterion or the number of errors made before reaching criterion. In addition, Wogan and Waters brought their subjects back to the lab a week later and had them relearn the list of pairs. The experimental group relearned the final list with which they had reached criterion. In this task, the experimental group actually took significantly fewer trials than the control group to reach criterion. The authors saw two possibilities for interpreting these results. First, they suggested that the results "might lead us to go even further than Rock and conclude that frequency neither facilitates the formation of connections nor strengthens connections already formed" (p. 613). Alternatively, the experimental group may have needed fewer trials to relearn their list because their list was made up of easier pairs. In other words, they suggested that Rock's procedure might have been subject to item selection effects.

Clark, Lansford, and Dallenbach (1960) also replicated Rock (1957). Using his procedure, they also found no differences between the experimental and control groups. After replicating the basic effect, Clark et al. also investigated the role item selection effects may have played in Rock's procedure. Like Rock and Heimer (1959), they assumed the best way to determine idiosyncratic differences in difficulty was to have subjects attempt to recall the items. The items a subject failed to recall were assumed to be difficult for that subject. In order to test for the effect of item selection, they first repeated Rock's procedure. The first 12 items the experimental group failed to recall were set aside. These subjects returned to the lab a week later and relearned these 12 presumably difficult items to the criterion of one perfect correct trial. For the control group, after subjects learned a list to criterion in the normal way, they studied and were tested once on two additional lists of 12 items. The first of these lists was set aside and used a week later. The second list was designed to mask learning from the first list. A week later, the control group relearned the first additional list, which was presumably made up of a mix of both easy and difficult items. Clark et al. assumed that if item difficulty played a role in determining the number of trials needed to reach criterion, the control group should relearn their easy and difficult list faster than the experimental group could relearn their list of only difficult pairs. However, both the experimental and control groups took exactly the same mean number of trials to reach criterion ($M = 2.9$), suggesting that differences in item difficulty did not affect the results in Rock's procedure.

The critics were not appeased by this research, as the spate of articles in 1962 shows. Let us consider a few of them. Underwood and Keppel (1962) complained about several features of Rock's (1957) experiment on various grounds: his presentation rate, his use of acceptance of the null hypothesis, his not taking response learning into account, and others. (Regarding the response learning issue, Rock did discuss it as his rationale for using easy, familiar stimuli in Experiment 1—letters and digits—that did not have to be integrated). Underwood and Keppel reported their own experiments, which they interpreted as supporting the incremental theory, but even they did not seem convinced: "We are under no illusions that these experiments settle the issue in the sense that they are crucial experiments" (p. 11). They also essentially argued that the incremental theory is untestable, so it is hardly a wonder that even they failed to be convinced by their own results. They wrote, "Certain theories are not capable of disproof. Certain aspects of the incremental theory seem to be of this nature" (p. 12). Concerning Rock's (1957) efforts, Underwood and Keppel wrote, "We have taken the position that the data which have been used to support one-trial learning postulates have come from experiments with faulty methods" (p. 11).

Postman (1962) noted that Rock's subjects had simply studied the pairs during learning, and he wondered whether this mattered. In two experiments using nonsense syllables as in Rock's (1957) Experiment 2, Postman had subjects spell out pairs letter by letter (both the stimuli and responses) in order to prevent them from attending to only a few pairs at a time (said to be possible under Rock's procedure). The other main difference was that Postman used two control groups. One was Rock's original control group: eight pairs repeatedly studied until learned to criterion. However, the other control group learned the lists that the experimental group eventually wound up learning. That is, because the experimental group in Rock's procedure involved repeated dropping and adding of pairs, the eventual list learned to criterion was different from that of the control group. If the item selection criticism is true, then a control group (called Control Group II by Postman) that learned the eventual lists of the experimental group should learn them faster than the original control group (Control Group I). If this result was found, it could be interpreted to mean that the lists the experimental group eventually learned were easier than the lists Control Group I had to learn. In this case, the experimental group would have been able to learn their list in as many trials as the original control group, not because learning is all-or-none but because their list was easier. However, the results in this regard were not particularly impressive; in one of Postman's experiments, no difference at all was apparent, and in the other experiment, the difference was fairly small (but significant). To rescue the situation, Postman used a cumulative recall analysis that better supported what he clearly hoped to find: a difference between the two control groups that was not much apparent by the customary analysis.

Kristofferson (1961), like Postman (1962), had subjects spell out responses to items rather than pronounce them in experiments that were critical of Rock's (1957) experiments. Rock and Steinfeld (1963) tried replicating these studies, comparing conditions in which subjects had to pronounce response items (say the word aloud) or had to spell them out (letter by letter). They showed that the spelling procedure seemed flawed and underestimated what subjects could report under standard (pronouncing) conditions. In addition, they used a paired-associate matching procedure in one of these experiments and concluded that the dropout procedure produced equivalent performance to the standard condition even in a matching task. The Rock and Steinfeld (1963) article that rebutted criticisms of Rock's (1957) earlier research was rarely cited in later works.

The study often cited in textbooks (e.g., Crowder, 1976) as the coup de grâce for Rock's experiment is one by Williams (1961). She used the same essential logic as Postman's two control conditions just described but had better luck finding the evidence for the item selection artifact Postman had sought. (These researchers were probably working independently from one another and unaware of each other's work; Williams's article was based on her PhD dissertation at Yale). Williams used letter–number pairs, as in Rock's (1957) Experiment 1, and had the same three groups Postman used: an experimental group that learned the items using Rock's dropout procedure, a control group that repeatedly studied the same pairs, and a second control group that repeatedly studied the pairs that subjects in the experimental group wound up learning. Williams replicated Rock's finding that there was no difference between the experimental group and the traditional control group in number of trials needed to learn the list to criterion. However, the control group that learned the experimental group's eventual list took significantly fewer trials to reach criterion, suggesting that item selection effects were a confounding factor in Rock's procedure. In addition, Williams had a third control group that provided further evidence for an item selection artifact. To determine this group's stimuli, Williams had another group of subjects learn pairs using a different kind of dropout procedure. Rather than dropping missed items, she dropped items that were answered *correctly* from the list (see also Kristofferson, 1961). This procedure continued until the entire pool of stimuli had been studied. The items on which the subjects made the most errors were used as stimuli for the third control group. This group took many more trials to reach criterion than the other three groups, indicating that this dropout procedure had produced a list of more difficult items.

Williams (1961) concluded that Rock's experiments could not be used as evidence for all-or-none learning because the item selection effects inherent in the procedure gave the experimental group an unfair advantage. This advantage was probably even larger than demonstrated in Williams's experiment because "of the likelihood of selection on the basis of idiosyncratic differences in difficulty" (p. 628).

This last phrase refers to the dreaded subject-by-item selection artifact. The argument is that what items are "difficult" is only partly normative across subjects but is also idiosyncratic to subjects. To illustrate, if long lists of categorized items were composed of earth formations, stars, and birds, geologists would do well in recalling on the first list, astronomers the second, and ornithologists the third. In a normal experiment, the researcher would have no way of understanding why some subjects did so well on some material and poorly on other material. The argument for subject-by-item artifacts in designs such as Rock's

(1957) dropout experiment is that there is always an unknown item selection artifact of this kind inherent in the procedure. Even with lowly digit-number and nonsense syllable pairs, each subject will find some easy and some hard, and items belonging to these hard and easy categories may differ widely across subjects. Thus, even using the control procedure of having one subject study the experimental list of another subject may not control for the effect (which may be why Postman, 1962, found no difference in one of his experiments). Of course, Rock and Steinfeld's (1963) procedure may surmount this problem, but no one ever followed it up.

The studies we have cited here are not all that exist on this matter by any means; others include Brackett and Battig (1963), Jones (1962), Kintsch (1963), Lockhead (1961), and Schwartz (1963). Postman (1963) wrote a chapter on the controversy that included both the procedure Rock (1957) used and that of Estes (1960), discussed briefly later in this article, and concluded that both types of experimental procedures were flawed. On the other hand, Kintsch's article is notable in that it argued that Rock might be right in assuming one-trial learning.

Tulving (1964) provided a theory of free recall learning, and he made a point in passing about the one-trial learning controversy: "Recent evidence shows clearly that a small unit of well-integrated verbal material, or an association between two such units, can practically always be recalled immediately following its presentation. . . . When we use probability of response as a measure of learning, therefore, we must conclude that learning of a small unit of material is always complete on a single trial. Learning, in this sense, is neither incremental nor all-or-none, it is always 'all'" (p. 221). The problem then becomes one of intratrial forgetting, of why an association retained perfectly immediately after study is recalled so imperfectly by the end of the list (see Bernbach, 1965). In a footnote, Tulving (1964, p. 221) remarked that Rock (1957) had considered this idea—that learning is always "all," with the real problem being forgetting between learning and testing—in his original article. Tulving wrote, "He [Rock] rejected this interpretation for reasons that are not entirely clear to this writer" (p. 221). The same is true for these writers. As noted earlier, Guthrie (1942) made this argument a centerpiece of his theory. Of course, at least in the realm of verbal learning, one could argue that immediate recall showing learning of the pair arises from a short-term store (Waugh & Norman, 1965) and that the issue in which Rock was interested was formation of associations that would support recall over the long term. Perhaps this is the reason, though never stated, that Rock dismissed the notion that all pairs are learned perfectly on their first presentation. In commenting on other work, Rock and Steinfeld (1963) said that another procedure (dropping correct items from lists across trials) was flawed because "Towards the end the list will contain only a few items which perhaps can be retained in immediate memory" (p. 823).

Restle (1965) published another review of and commentary about the one-trial learning controversy that was more favorable. He helpfully pointed out that there were some six different possible versions of what theorists meant by *all-or-none learning,* and he further argued that some all-or-none postulates were probably right and others were certainly wrong. For example, a claim that is true (according to Restle) is "There exists at least one task that is learned all-or-none," and a postulate that is false is "All learning is all-or-none" (p. 323). Other cases were more in doubt. Restle's article, arguing again (like Harlow, 1949) that everyone in the debate seemed partly right and partly wrong, seemed to finish off the all-or-none versus incremental debate in verbal learning and retention. Considering the issue later, Tulving and Madigan (1970) stated that "the issue of all-or-none versus incremental learning was successfully pronounced dead by Restle (1965)" (p. 450). They seem to have been right (even though they advocated bringing it back to examine a related issue).

The Aftermath of the All-or-None Debate

In retrospect, the all-or-none brouhaha begun by Rock's (1957) experiments seems a strange episode in the history of the study of learning. Rock performed straightforward experiments testing an interesting idea. Estes (1960) championed the same point of view from a separate type of experiment called a reinforcement–test–test (RTT) experiment, in which pairs were studied (reinforced, in the language of the day) and then tested twice in succession. Estes's work will be considered briefly later in this article, but suffice it to say for now that Estes also argued for an all-or-none position. Although his work also attracted critics, many of the same critics in fact, they seemed more respectful to Estes. William Estes was already a leading figure in the field, and Irvin Rock was then young, brash, and an outsider (and to add insult to injury, his 1957 article did not cite much of the establishment's work).

Reading between the lines of some of the articles authored by critics, one can see vitriol seeping in. For example, in footnote 16 of the article in which he intended to replicate Rock's (1957) experiment, Postman (1962) wrote, "It is not clear whether correction [feedback] was given on the test trials in Rock's first experiments. No reference to a correction was made in the original report. However, in describing a study that is said to have followed the original procedure in all respects, Rock and Heimer state that S was told whether he was right or wrong" (p. 377). The implication is that Postman could not be bothered to pick up the telephone to seek the answer from Rock or even write him a letter. After Irvin Rock passed away in 1995, Gilchrist (1996) cited Rock's many contributions to the psychology of perception. He then noted that Rock was an iconoclast and never hesitated to oppose entrenched interests. Gilchrist wrote, "During the one-trial learning controversy, vigorous and effective efforts were made by influential learning theorists to suppress Rock's work" (p. 497).

Rock's (1957) experiment and the all-or-none debate he began in verbal learning was a hot topic in the 1960s. Slamecka (1967) published a highly regarded book of readings that has 10 sections of readings on the critical topics of the day, one of which was the incremental versus all-or-none debate. Yet after the 1960s, the debate died away. The argument against Rock's work has a curious feature. Rock originally pointed out several other factors that would lead subjects to perform better in the control than in the experimental condition, yet he found no difference. Critics said that this was because there was a negative force—more difficult items—facing the experimental condition and driving it down. Thus, to explain Rock's results, the critics needed to suppose that there were positive effects (gradual accumulation of associative bonds and other factors) that were exactly balanced by the negative effect of item selection (more difficult items in the experiment). This seems a difficult argument to sell, especially when some researchers had failed to find evidence for item selection. Nonetheless, it worked. The one-trial learning debate ground to a halt in 1965 or shortly thereafter.

Leading textbooks published in the mid-1970s finished off whatever whiff of interest there might still have been. Kausler (1974) devoted two pages to Rock's work and its aftermath. After ably reviewing the basic experiments in a paragraph, he wrote, "Other investigators, however, have largely refuted the contention that associative strength does not accrue prior to the first correct recall of a given pair" (p. 154). The criticisms went on for a page and a half. Crowder (1976, p. 268) praised Rock on two grounds but dismissed his experiments. He wrote, "Although Rock's study served the important function of rearousing interest in the nature of the acquisition process among workers in verbal learning . . . and although he anticipated more modern attitudes about mediation and problem solving in paired-associate learning . . . his experiments were, strictly speaking, irrelevant." Ouch.

At least Kausler's (1974) and Crowder's (1976) texts covered Rock's work. Other texts of the time (Baddeley, 1976; Hall, 1971) did not mention the issue. More recent textbooks, including all current ones, have followed suit. Thus, in many ways the one-trial learning issue raised by Rock may be damned with the faint praise of being "merely of historical interest." We will contest this assertion later in this article, but first we turn to a brief description of William Estes's (1960) research (also see Estes et al., 1960) that raised the same issue.

William Estes and the New Mental Chemistry

Estes's (1960) article was titled "Learning Theory and the New 'Mental Chemistry'" and is perhaps the only article in the history of psychology to begin with the word "Hovering." As noted earlier, Estes introduced the RTT paradigm (which today would be called an STT paradigm, for study–test–test). A complete account of his logic with regard to one-trial learning would take this article too far afield, but the basic idea is straightforward. Consider performance of subjects after a single presentation of a list on a single trial and imagine it to be 50%. According to all-or-none and incremental theories, this value arises in two totally different ways. As Crowder (1976, p. 269) put it, "According to the incremental theory the .50 success rate means that all the items, for all subjects, have been hovering at around the same intermediate strength of 50% and, because of some oscillation principle or random error, this strength has resulted in correct recall half the time. The all-or-none view is that half the items are fully learned and half are not at all learned."

The prediction is for what should happen on the second test in the RTT paradigm, which is given without an intervening study trial. Estes's (1960) logic led to the prediction that if the all-or-none principle is correct, subjects should never get the item correct on the second trial (or at least no more than chance) if they did not get it on the first trial, because it is an unlearned state with zero strength. On the other hand, if all items are oscillating around 50% strength, then there should be a good deal of fluctuation in which items are recalled between two trials even without additional study. Estes's research (1960; Estes et al., 1960) showed that recall on the second trial in paired-associate learning was very low and close to chance levels, hence confirming the all-or-none learning position. Others were not convinced, however (Postman, 1963; Underwood & Keppel, 1962). Still, critics were generally much kinder in the tone of their criticisms of Estes than of Rock. Bower (1962) also proposed a model that assumed one-trial learning, but Bower's own results disconfirmed his model, so it never gained much traction.

Unlike Rock's (1957) dropout paradigm, Estes's (1960) RTT paradigm led to much future research. Izawa (1966, 1971, and many other articles; for a review, see Izawa, 1992) used variations of the paradigm to study what she called test-potentiated learning. This term refers to the fact that taking a test before a study trial leads to greater learning from that study trial (relative to plausible control conditions). This research is now seeing a resurgence of interest because of an increase in interest on the effects of testing memory on later retention (Arnold & McDermott, 2012; McDermott & Arnold, 2010; see Roediger & Karpicke, 2006, for an overview of how retrieval practice during tests affects retention). However, whether learning is incremental or all-or-none is not part of this test-potentiated learning discussion. Still, research originally directed at the one-trial learning issue lives on, indirectly, in this work and in related work on the effects of tests.

We should note that the incremental versus all-or-none debate was played out in paired-associate learning. Research using the STT paradigm in free recall has led to very different conclusions. Tulving (1964, 1967) showed that when two free recall test trials are given back to back, there is a great deal of fluctuation between trials; there is item loss (forgetting) and item recovery (reminiscence). The latter finding of recovery between trials is quite different from the usual case in paired-associate learning that Estes (1960) studied. Erdelyi and Becker (1974) showed that reminiscence or recovery between trials in free recall is often greater than forgetting, leading to overall improvements in recall. They named this effect—overall greater recall on a later test than on an earlier test—hypermnesia, and it has been replicated many times (e.g., Mulligan, 2006; Roediger & Thorpe, 1978; see Payne, 1987, for a historical review). The general point, noted by Restle (1965), is that the incremental versus all-or-none debate was specific to a particular task in both Rock's (1957) and Estes's (1960) work. Because both researchers were interested in how associations are formed, paired-associate learning was the natural task to use.

Contributions of Rock's All-or-None Work to the Field

Rock raised an important issue that is central to psychology but that is often dodged: How does learning occur? How are associative bonds formed? Aristotle asked the same questions, and they are enduring ones. Rock (1957) provided a strong answer for which there was historical precedent (Guthrie, 1942), but his view was not accepted. To the extent that anyone worries about this topic today, the gradualist assumption still reigns as psychologists and neuroscientists speak of the strength of memory traces. The idea of graded strength of associations is found in associative memory models such as human associative memory (HAM; Anderson & Bower, 1973) and search of associative memory (SAM; Raaijmakers & Shiffrin, 1981) and in connectionist models (Rumelhart & McClelland, 1986). However, most models simply assume that strength of associations is graded without citing strong evidence for this assumption. As Deese and Hulse (1967) commented in their text about Rock's critics, "Those who are unhappy with the basic assumptions of the all-or-none theory have not been as rigorous about their own views. For example, what do we mean by the term strength?" (p. 313). At least Rock made his ideas explicit enough to be tested, which those in the other camp in the early 1960s never did. Recall that Underwood and Keppel (1962) pronounced this assumption nearly untestable, accepted simply as an article of faith, which is a curious stance for experimental psychologists. We should note in passing that other models of memory (Hintzman's MINERVA model, 1988; and the "one shot" hypothesis of context storage in the REM model, Malmberg & Shiffrin, 2005) make assumptions about learning that are more like the all-or-none position. The issue remains with the field, even thought Rock's (1957) and Estes's (1960) research are rarely cited in these more contemporary contexts.

Another contribution of the one-trial learning debate was to alert researchers to item selection artifacts and the subject-by-item selection artifact. This pernicious issue has been a consideration in much research in the 50 years since Rock's (1957) research was criticized on these grounds. The problem crops up in many types of research, but most experimentalists now know how to examine the criticism and, where possible, overcome it. Rock (1957) understood the problem, but his research led him to conclude that it was not serious (Rock & Heimer, 1959; Rock & Steinfeld, 1963).

Finally, Rock (1957) also called attention to the idea that paired-associate learning can be facilitated by what he called mnemonic devices or what today would be called mediators. The idea is that in learning some pair (e.g., *lion–stripes*), if subjects thought of remembering the series *lion* → *tiger* → *stripes* they might remember better. Recently, Pyc and Rawson (2010) developed a mediator effectiveness hypothesis and applied it to paired-associate learning, arguing that subjects try one mediator, and if it does not work to prompt recall on the next trial, they will switch to another mediator. The flavor of this idea is in line with Rock's early suggestion of one-trial learning. Once a subject hits on a successful mediator, the item has been learned. Carpenter (2011) provided additional evidence for the idea. The mediator effectiveness hypothesis may be the closest concept today to one-trial learning as Rock (1957) envisaged it, although Pyc and Rawson's (2010) formulation of the issue is rather different.

Final Reflections

What are we to make of the one-trial learning controversy 50 years later?

In one sense, it is dead and gone. No one asks whether associations are learned in an all-or-none fashion or incrementally, and perhaps it is simply a wrong question to ask. Conceptions about learning and memory have greatly changed in the past 50 years. Tulving (1964) argued that all learning of pairs occurs at the time they are presented, in the sense that an alert subject can always repeat the last pair perceived after a brief delay. The problem then becomes one of explaining why an item learned well enough to be produced immediately with a probability of 1.0 is recalled with a much lower probability a short time later on a test at the end of the list (depending on list length and other list characteristics). The problem of retaining associations is then recast as one of intratrial forgetting rather than of means of initial learning (Tulving, 1964). As noted earlier, the notion of learning always occurring in the short term may still lead to the question of how associations are formed in the long term.

On the other hand, many researchers still use concepts such as memory strength in both the psychology and neurobiology of learning and memory. Is it not reasonable to ask by what process this strength grows, which was the question Rock (1957) addressed? Rock asked the question with one task in particular in his focus—paired-associate learning—and that task is still of great current interest. We predict that, like most fundamental questions, the issue of how strength accrues will come to the forefront again, albeit through methods other than those tried previously.

In our review of the literature on this topic, perhaps the most interesting (and certainly the funniest) piece we encountered was a commentary by George Miller (1963). The occasion of his contribution was for the Second Conference on Verbal Learning and Verbal Behavior, held in June 1961. Luminaries of the day were invited to speak on major issues, and other equally distinguished researchers were asked to comment on the eight main presentations. One of the featured presentations was by Leo Postman on the one-trial learning debate. Postman (1963, the year the book reflecting the 1961 conference proceedings appeared) examined the issue thoroughly and concluded that both Rock (1957) and Estes (1960) were wrong. He said that it was good of them to raise such an interesting issue but that the field could rest assured that learning occurred in a gradual, incremental fashion and that their new idea was simply wrong. (Neither Rock nor Estes attended the conference.)

Miller (1963) had the task of commenting on Postman's (1963) chapter. He began by noting that verbal learning was not really his area and that he relied on his friends to keep him abreast of interesting developments. One of those close friends was Leo Postman, so "I arrive at the task of criticizing his ideas with a certain amount of embarrassment. Since he is one of my major sources of information, it would be impertinent, not to say reckless, for me to take issue with him. So let me begin by saying that I endorse all he says in his paper. If Postman says one-trial learning is nonsense, then it is nonsense and that is all there is to it. Clark Hull is still our leader" (1963, pp. 322–323).

After these sentences, Miller (1963) went on to rescue the all-or-none theory, to say that Postman's criticisms assume a very strong all-or-none position that does not apply to Rock's (1957) claims. Miller then developed a new theory—the junk box theory—that shows how all-or-none learning can be reconciled with various sorts of data that do not look as if an all-or-none assumption could give rise to them. In short, while not exactly criticizing Postman—he repeats that he would not do that—he shows how Postman could be wrong. We will not spell out the argument here, but if our article commemorating Rock's work has not choked off all interest in the reader, we suggest Miller's (1963) commentary as a good place to look for insight into the issue. We suspect that the whole topic is due for a comeback.

NOTES

Writing of this article was supported by a Collaborative Activity Grant from the James S. McDonnell Foundation, a grant from the Cognition and Student Learning Program in the Institute of Education Sciences in the Department of Education, and a grant from Dart Neuroscience, all to the first author. The second author was supported by a training grant from the National Institute of General Medical Sciences in the National Institutes of Health.

Address correspondence about this article to Henry L. Roediger III, Department of Psychology, Box 1125, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130-4899 (e-mail: roediger@wustl.edu).

1. Tulving (1964) made a similar assumption, as we shall note later in this article.

REFERENCES

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston & Sons.

Arnold, K. M., & McDermott, K. B. (2012). *Test-potentiated learning: Distinguishing between direct and indirect effects of tests*. Manuscript submitted for publication.

Baddeley, A. D. (1976). *The psychology of memory*. New York, NY: Basic Books.

Battig, W. F. (1962). Paired-associate learning under simultaneous and nonrepetition conditions. *Journal of Experimental Psychology, 64,* 87–93.

Battig, W. F., & Brackett, H. R. (1961). Comparison of anticipation and recall methods in paired-associate learning. *Psychological Reports, 9,* 59–65.

Bernbach, H. A. (1965). A forgetting model for paired-associate learning. *Journal of Mathematical Psychology, 2,* 128–144.

Bower, G. H. (1962). An association model for response and training variables in paired-associate learning. *Psychological Review, 69,* 34–53.

Bower, G. H., & Hilgard, E. R. (1981). *Theories of learning*. Englewood Cliffs, NJ: Prentice Hall.

Brackett, H. R., & Battig, W. F. (1963). Method of pretraining and knowledge of results in paired-associate learning under conditions of repetition and non-repetition. *American Journal of Psychology, 76,* 66–73.

Calkins, M. W. (1894). Association. *Psychological Review, 1,* 476–483.

Carpenter, S. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1547–1552.

Clark, L. L., Lansford, T. G., & Dallenbach, K. M. (1960). Repetition and associative learning. *American Journal of Psychology, 73,* 22–40.

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Deese, J., & Hulse, S. J. L. (1967). *The psychology of learning* (2nd ed.). New York, NY: McGraw-Hill.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York, NY: Dover. (Original work published 1885; translated 1913)

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory of pictures but not words in multiple recall trials. *Cognitive Psychology, 6,* 159–171.

Estes, W. K. (1960). Learning theory and the new "mental chemistry." *Psychological Review, 67,* 207–223.

Estes, W. K., Hopkins, B. L., & Crothers, E. J. (1960). All-or-none and conservation effects in the learning and

retention of paired associates. *Journal of Experimental Psychology, 60,* 329–339.
Gilchrist, A. (1996). Irvin Rock, 1922–1995. *Perception, 25,* 495–497.
Glaze, J. A. (1928). The association value of non-sense syllables. *Pedagogical Seminary and Journal of Genetic Psychology, 35,* 255–269.
Guthrie, E. R. (1942). Conditioning: A theory of learning in terms of stimulus, response, and association. *Yearbook of the National Society of Education, 41,* 17–60.
Hall, J. F. (1971). *Verbal learning and retention*. Philadelphia, PA: J.B. Lippincott.
Harlow, H. F. (1949). The formation of learning sets. *Psychological Review, 56,* 51–65.
Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.
Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory.* Oxford, England: Appleton-Century.
Hull, C. L. (1952). *A behavior system: An introduction to behavior theory concerning the individual organism*. New Haven, CT: Yale University Press.
Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports, 18,* 879–919.
Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology, 8,* 200–224.
Izawa, C. (1992). Test trials contributions to optimization of learning processes: Study/test trials interactions. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes* (pp. 1–33). Hillsdale, NJ: Erlbaum.
Jones, J. E. (1962). All-or-none versus incremental learning. *Psychological Review, 69,* 156–160.
Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138,* 469–486.
Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319,* 966–968.
Kausler, D. H. (1974). *Psychology of verbal learning and memory*. New York, NY: Academic Press.
Kintsch, W. (1963). All-or-none learning and the role of repetition in paired-associate learning. *Science, 140,* 310–312.
Köhler, W. (1925). *The mentality of apes*. New York, NY: Harcourt, Brace.
Krechevsky, I. (1938). A study of the continuity of the problem solving process. *Psychological Review, 45,* 107–133.
Kristofferson, A. B. (1961). Repetition and paired associates learning. *Science, 134,* 2036–2037.
Lockhead, G. R. (1961). A re-evaluation of evidence of one-trial associative learning. *American Journal of Psychology, 74,* 590–595.
Malmberg, K. J., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 322–336.
McDermott, K. B., & Arnold, K. M. (2010, November). *Test taking facilitates later learning*. Talk presented at the annual meeting of the Psychonomic Society, St. Louis, MO.
Miller, G. A. (1963). Comments on professor postman's paper. In C. N. Cofer & B. S. Mugrave (Eds.), *Verbal behavior and learning: Problems and processes* (pp. 321–329). New York, NY: McGraw-Hill.
Mulligan, N. W. (2006). Hypermnesia and total retrieval time. *Memory, 14,* 502–518.
Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin, 101,* 5–27.
Postman, L. (1962). Repetition and paired-associate learning. *American Journal of Psychology, 75,* 372–389.
Postman, L. (1963). One-trial learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes* (pp. 295–321). New York, NY: McGraw-Hill.
Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330,* 335.
Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.
Restle, F. (1965). Significance of all-or-none learning. *Psychological Bulletin, 64,* 313–325.
Rock, I. (1957). The role of repetition in associative learning. *American Journal of Psychology, 70,* 186–193.
Rock, I., & Heimer, W. (1959). Further evidence of one-trial associative learning. *American Journal of Psychology, 72,* 1–16.
Rock, I., & Steinfeld, G. (1963). Methodological questions in the study of one-trial learning. *Science, 140,* 822–824.
Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210.
Roediger, H. L., & Thorpe, L. A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition, 6,* 296–305.
Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
Schwartz, M. (1963). Transfer from failed pairs as a test of one-trial vs. incremental learning. *American Journal of Psychology, 76,* 266–273.
Slamecka, N. J. (Ed.). (1967). *Human learning and memory: Selected readings*. New York, NY: Oxford University Press.
Spence, K. W. (1945). An experimental test of the continuity and non-continuity theories of discrimination learning. *Journal of Experimental Psychology, 103,* 204–210.

Tulving, E. (1964). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6,* 175–184.
Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6,* 175–184.
Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology, 21,* 437–484.
Underwood, B. J., & Keppel, G. (1962). One trial learning? *Journal of Verbal Learning and Verbal Behavior, 1,* 1–13.
Underwood, B. J., Rehula, R., & Keppel, G. (1962). Item-selection in paired-associate learning. *American Journal of Psychology, 75,* 353–371.
Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72,* 89–104.
Williams, J. P. (1961). Supplementary report: A selection artifact in Rock's study of the role of repetition. *Journal of Experimental Psychology, 62,* 627–628.
Wogan, M., & Waters, R. H. (1959). The role of repetition in learning. *American Journal of Psychology, 72,* 612–613.
Wollen, K. A. (1962). One-trial versus incremental paired-associate learning. *Journal of Verbal Learning and Verbal Behavior, 1,* 14–21.
Yerkes, R. M. (1916). The mental life of monkeys and apes, a study of ideational behavior. *Behavioral Monographs,* No. 12.