Memory & Cognition
2010, 38 (1), 102-115
doi:10.3758/MC.38.1.102

How does repetition affect memory? Evidence from judgments of recency

Douglas L. Hintzman
University of Oregon, Eugene, Oregon

Four experiments were done to investigate the effects of repetition on judgment of recency (JOR). Experiment 1 showed that repetition can make an item seem either more recent or less recent than a nonrepeated item, depending on presentation spacing. Experiments 2–4 showed that subjects are able to judge the recency of a repeated item's first presentation or of its second presentation with a high degree of independence, especially if they report that the item occurred twice. The data are more consistent with an independent-trace explanation of JOR and repetition than with a cumulative-strength account, but neither hypothesis explains how repetition can make an item seem less recent. It is proposed that the findings as a whole can be better explained by a hypothesis based on recursive reminding.

After more than a century of research on episodic memory, there is still little agreement on mechanisms that underlie the effects of repetition. The most commonly considered alternatives are the *cumulative-strength hypothesis* and the *multiple-trace hypothesis*. The first of these assumes that when an experience is repeated, the memory representation that was formed during the first such experience is strengthened. The second assumes that each experienced event leaves behind its own separate memory trace, even if the event is a repetition.

The cumulative-strength versus multiple-trace distinction goes back at least to the 19th century (Ward, 1893), but both hypotheses are prominent among modern cognitive theories of memory. Examples of cumulative-strength theories can be found in McClelland and Chappell (1998), Murdock (1982), Murdock, Smith, and Bai (2001), and Wickelgren (1972). Examples of multiple-trace theories include Bower (1967), Hintzman (1986, 1988), Lansdale and Baguley (2008), and Logan (1988). Some theorists have vacillated between the two positions. In introducing the SAM model, Gillund and Shiffrin (1984) assumed that massed repetitions strengthen a single trace, whereas spaced repetitions may produce multiple traces; Shiffrin and Steyvers (1997) implemented the cumulative-strength assumption in their REM.1 model but used the multiple-trace assumption in their REM.3 model; and Malmberg and Shiffrin (2005) discussed SAM and REM as though both were cumulative-strength models.

Repetition generally improves performance in standard recall and recognition-memory experiments, so for the purpose of explaining performances in these tasks, the cumulative-strength versus multiple-trace distinction may not much matter. Researchers investigating this issue have therefore turned to tasks that require other kinds of memory judgments. Judgments of recency (JOR) have played an especially important role in this work. There are two basic JOR procedures. In the *forced-choice JOR* or *recency-discrimination* task, the experimental subjects choose the member of a test set (usually two items) that seems more recent. In the *numerical JOR* or *absolute-judgment* task, the subjects judge the number of items that intervened since a single test item was last presented. Data from both JOR tasks suggest that apparent recency approximately follows a logarithmic function of time or actual recency (e.g., Hinrichs, 1970; Yntema & Trask, 1963).

The conclusions that researchers have reached regarding repetition's effects have depended, in part, on which type of JOR test was used in collecting the data. In an early study, Morton (1968) required subjects to discriminate the recencies of two test digits, A and B, from a short preceding list. Taking B as the correct answer, the study conditions of interest were AB and AAB. (In the second condition, A occurred two times followed by a single occurrence of B.) Recency discrimination was less accurate in the AAB condition than in the AB condition—that is, the subjects had a greater tendency to choose the incorrect item if it had been repeated. Morton concluded from this result that remembered recency is based on strength and that repetition increases the strength of the item's memory (see also Murdock et al., 2001).

Early numerical JOR experiments, however, seemed to point to a different conclusion. Peterson (1967) had subjects go through a list of words, giving numerical JORs to critical test items. Some words were tested on the second presentation (P2), and some on the third (P3). The recency or test lag varied from one to eight intervening items, but for items tested on P3, the P1–P2 spacing was constant at four intervening items. Peterson found no consistent

D. L. Hintzman, hintzman@uoregon.edu

difference in JOR between items tested on P2 and items tested on P3, suggesting no effect of repetition. In three follow-up experiments, Peterson, Johnson, and Coatney (1969) varied the spacing of P1 and P2. Mean JOR was significantly shorter for repeated than for nonrepeated items when P1 and P2 were massed, but not when they were spaced. The researchers concluded that, although JOR is based on trace strength, a single trace is strengthened only when repetitions are massed. As in the later theory of Gillund and Shiffrin (1984), it was assumed that spaced repetition yields multiple traces.

It should be recognized that the cumulative-strength versus multiple-trace issue interacts with the nature of the cue (or cues) theorized to underlie JOR. Hinrichs (1970) proposed that decaying trace strength underlies JOR, without assuming that strength accumulates over repetitions. If strength decays as a function of time, this seems like a plausible hypothesis. However, direct comparison of JOR with recognition confidence—another purported measure of strength—suggests that JOR is too tightly correlated with time for strength to be the primary cue to recency (Hintzman, 2005). Other factors that have been proposed to underlie JOR include the degree of consolidation of the memory trace (Wickelgren, 1972, 1974) and the degree of match between the retrieved context and the test context (Hintzman, 2002; Hintzman, Block, & Summers, 1973). A special version of this last hypothesis holds that JOR is based on changes in time-specific contextual elements characterized as oscillator readings (Brown, Preece, & Hulme, 2000). At present, there seems to be no compelling evidence in favor of any of these alternatives. Typically, theorists taking a multiple-trace perspective have avoided the issue by referring simply to *time tags*, without specifying the nature of the temporal information.

Although cumulative-strength and multiple-trace theories of episodic memory coexist in the current literature, the distinction has essentially disappeared as an empirical question. One reason for this disappearance may be that for the past several years, theoretical efforts have been focused on recognition memory, which is not diagnostic to the issue. Another possible reason is that memory-judgment experiments published more than three decades ago may have appeared—to observers at the time—to settle the issue in favor of the multiple-trace hypothesis. Two of these studies, one using judgments of serial position (Hintzman & Block, 1971) and the other using recency discrimination (Flexser & Bower, 1974), provide essential background for the present experiments.

Hintzman and Block (1971) asked whether a repeated item has more than one remembered serial position. Within a 50-item list, they independently varied the positions of P1 and P2 of words that were repeated. The subjects were then asked to judge the presentation positions of a series of test words, giving two position judgments for any word that they thought had appeared twice in the list. The results showed that first-position judgments for such words were affected by the position of P1, and second-position judgments were affected by the position of P2, but neither judgment was reliably affected by the nontarget position.

Although these results support a multiple-trace view of repetition, they are not of direct relevance to JOR. When a study list is followed by a randomly ordered test list, as in the position-judgment task, the serial position of an item in the study list is only weakly correlated with the study trial's actual recency at the time of the test. Moreover, delaying the test list by several minutes appears to mostly affect memory for positions near the end of the list (Hintzman et al., 1973). The most accurate position judgments are for items near the beginning of the list, and recency, per se, has little effect on position judgments over this range. This result suggests that serial position judgments are determined primarily by changes in cognitive context, which evolves most rapidly just after the start of the list, and that they are affected only secondarily, if at all, by remembered recency. Put simply, position judgments are made in relation to two temporal landmarks—the beginning and end of a previous list—whereas recency judgments are made in relation to the moment of the test. Seen in this way, position judgments and recency judgments are quite different tasks, despite their superficial similarity.

Flexser and Bower (1974) extended the multiple-trace hypothesis to the forced-choice recency-discrimination task. As in the pioneering experiments of Yntema and Trask (1963), subjects went through a long word list in which recency-discrimination test pairs were interspersed; however, as in the study by Morton (1968), many of the pairs required choosing a repeated versus a nonrepeated item. Crucial test pairs, in addition to the AB control, were of the AAB, BAB, and ABB varieties. Half of Flexser and Bower's subjects were asked to judge whether A or B was more recent, and—consistent with the findings of Morton—these subjects tended to choose a test item more often if it had been repeated. The other half of the subjects judged whether A or B occurred earlier in the list (judgments of *relative distance*). These subjects also tended to choose the repeated test item. Most notable, B was chosen as the earlier member of the test pair more often in the ABB condition than in the AB condition. Qualitatively, recency judgments and distance judgments produced mirror-image outcomes.

To explain their findings, Flexser and Bower (1974) proposed a model in which a repeated item leaves two independent traces, each with an associated contextual time tag, sampled from a distribution of times or apparent recencies. The traces are said to be independent, because they differ only in their time tags. Neighboring study trials have time-tag distributions that overlap (see Figure 1). Repeated items tend to be judged more recent than nonrepeated items, because they have two traces and two such time tags, either of which might seem more recent than the single time tag of the item that occurred once. In agreement with Flexser and Bower's findings, the same mechanism predicts a bias in the opposite temporal direction when subjects are asked to judge relative distance (i.e., the mirror image of Figure 1).

The judgment task used by Flexser and Bower (1974) was more complex than just described, however. Prior to making the choice between A and B, the subject gave a

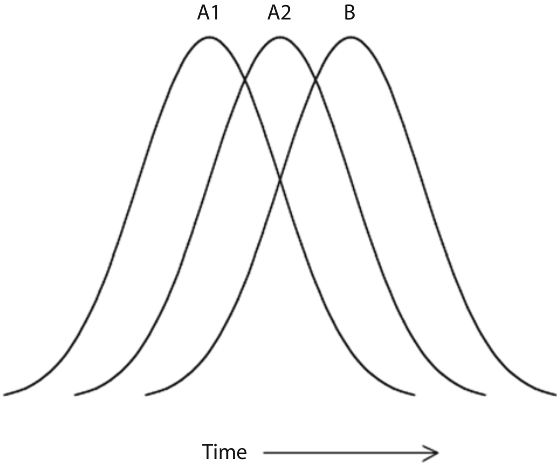

Figure 1. Distributions of apparent recencies or time tags, according to the independent-trace hypothesis. Item A is presented twice, and item B is presented once. What is important is the overlap, not the shapes of the distributions.

judgment of presentation frequency (JOF) for A and a JOF for B. The idea was to determine whether the subject was really retrieving two traces of the repeated item. In accordance with the reasoning behind their model, Flexser and Bower found that repetition effects in both the recency- and distance-discrimination tasks were strongest when the subjects correctly gave JOF = 1 for the once-occurring item and JOF = 2 for the repeated item.

Although no one seems to have noticed it at the time, this result is also consistent with a different interpretation of Flexser and Bower's (1974) results: Judgments of relative recency and relative distance could have been biased by memory for frequency. Consider a case in which the subjects know that A occurred twice and B occurred once but know nothing of their recencies. If there are three events—A1, A2, and B—they could have occurred in any of six orders. In four of the six orders A is first, and in four of the six orders A is last. In the absence of reliable recency information, therefore, the most rational choice in either task would be to pick the repeated item. Thus Flexser and Bower's basic outcome—interpreted as evidence that a repeated item has two recencies—could have been an artifact of strategic guessing.

This suspicion is reinforced by the findings of Galbraith (1976), who compared recency discrimination in the AB, AAB, and ABB conditions, using both massed and spaced presentations of the repeated items. Consider in particular the AAB condition. According to the independent-trace model of Flexser and Bower (1974), the tendency to incorrectly choose A should be greater if the A1–A2 spacing is short than if it is long, because, in the former case, the A1 time-tag distribution has more overlap with the B time-tag distribution. Galbraith's subjects, however, chose A more often when A1 and A2 were spaced than when they were massed—exactly the opposite of this prediction. The result is consistent with the hypothesis that a preference for the repeated item is a strategic-guessing artifact, because spaced repetitions are more likely than massed repetitions to be given JOF = 2 (e.g., Hintzman, 1969). Galbraith discussed his result as showing that a frequency attribute may contribute to recency judgments, but he did not point out that it would be entirely rational for subjects to use remembered frequency in this way in the recency-discrimination task, especially if reliable recency information were lacking.

If remembered frequency can bias responding in the forced-choice task, experiments using this task (e.g., Flexser and Bower, 1974; Morton, 1968; Murdock et al., 2001) may not be appropriate for studying the effect of repetition on apparent recency. It is less obvious whether remembered frequency might also bias the numerical JOR task, but that possibility also needs to be considered.

The present experiments were motivated by questions about repetition and JOR that the just-reviewed literature leaves unanswered. The task was a running recognition and JOR paradigm using very long lists to minimize contamination by associations with temporal landmarks (e.g., Hintzman, 2001, 2002, 2003). Experiment 1 was essentially an extension of the earlier paradigm of Peterson (1967; Peterson et al., 1969), which suggested that only massed repetitions affect JOR. Compared with those studies, Experiment 1 employed a wider range of recencies and a wider range of P1–P2 spacings, but none of the repetitions were massed (spacing = 0). In Experiments 2–4, I applied the independent-manipulation logic of the Hintzman and Block (1971) position-judgment experiment to numerical JORs, to determine whether a repeated item has more than one remembered recency. In these experiments, the recencies of P1 and P2 were manipulated separately, and subjects either judged the recency of a test item's first occurrence (JOR first) or of its last occurrence (JOR last). In Experiments 1 and 2, each JOR was preceded by an *old–new* recognition judgment, but in Experiments 3 and 4, it was preceded by a frequency judgment. Because the intended manipulations of Experiment 1 were different from those of Experiments 2–4, the crucial spacings or lags are most conveniently defined in different ways. Figure 2 illustrates the relationships among the definitions.

As will be seen, the results of these experiments considered as a whole were not as expected from either the

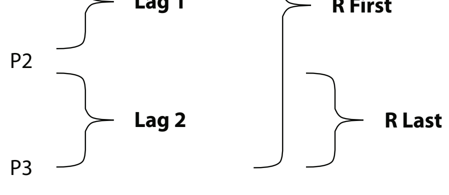

Figure 2. Definitions of the terms Lag 1, Lag 2, R first, and R last. P1, P2, and P3 are different presentations of the item.

cumulative-strength or the independent-trace perspective. To explain the data, I will invoke the view that repetition results in *recursive reminding*—a hypothesis that has been proposed to explain how people remember presentation frequency (Hintzman, 2004). For the purposes of exposition, presentation of the recursive-reminding hypothesis will be deferred to the General Discussion section.

EXPERIMENT 1

In this experiment, subjects went through a long list of words, indicating whether each word was old or new and then making a numerical JOR to each correctly identified old item. Most words occurred three times. Repetitions occurred at three different spacings (5, 10, and 30 items), which were assigned orthogonally to the P1–P2 and P2–P3 lags.

Method

Subjects and Materials. Twenty-five University of Oregon students participated for course credit. They were tested individually.

The stimuli were selected at random from a master list of 300 low-frequency English nouns. Mean word length was 6.7 letters (*SD* = 1.8), mean Thorndike–Lorge (1944) frequency was 1.2 per million (*SD* = 1.0), and mean concreteness was 5.3 (*SD* = 1.5) on a scale of 1–7.

Design and Procedure. The experimental procedure was patterned after those of Hintzman (2001, 2002, 2003), except that many words were presented three times in the list. As can be seen in Figure 2, the difference in serial position between P1 and P2 is referred to as Lag 1, and the difference between P2 and P3 is referred to as Lag 2. The 620-item list was arranged so that there were three different values of Lag 1 (5, 10, or 30 intervening items), crossed with three different values of Lag 2 (also 5, 10, or 30 items). This resulted in 13 different types of trials: one type of P1 (new items), three types of P2 (defined by Lag 1), and nine types of P3 (defined by Lag 1 × Lag 2).

The subjects were seated before a computer and told that they would see a long list in which many words would be repeated—some more than twice. For each word, the subjects' first task was to decide whether the word was new or old in the list and to respond using the "Z" or "?/" key of the keyboard. If they decided that the word was old, they would also be asked to judge how far back in the list was the word's most recent occurrence. They were told that a word could be repeated after 5, 10, 15, 20, 25, or 30 intervening items, and that the JORs, or lag judgments, were to be made using the 1–6 keys of the numerical keypad, which had been relabeled with these numbers. The subjects were told that there was no reasonable strategy for keeping track of word order, so they should simply rely on their intuitive feelings of recency.

To generate the 620-item list, words were selected at random from the pool without replacement, as was needed. The experimental program generated the list by stepping through trials $i = 1$–620, in sequence. For each value of i, the program first checked whether a word was already assigned to that position. If not, P1 of the current word was assigned to i, one of the repetition lags (Lag 1 = 5, 10, or 30) was chosen at random, and the program checked whether a word was assigned to position i + Lag 1. If that position was occupied, the word was not repeated; otherwise, P2 of the word was assigned to i + Lag 1, and a second lag (Lag 2 = 5, 10, or 30) was selected at random. P3 of the word was assigned to position i + Lag 1 + Lag 2 only if that location was also unoccupied.

A 500-msec blank interval preceded the onset of each test word, which was then displayed in 48-point, lowercase Helvetica text in the center of the screen. When the word appeared, the subject was first prompted to indicate whether it was old or new. The response alternatives ("Z = new" and "?/ = old") were shown on the screen in a single line beneath the word. If the subject responded *old*, this instruction was immediately replaced by the two lines: "How many items back?" and "5 10 15 20 25 30." Each display remained on the screen until the subject responded on the keypad. Illegal keypresses resulted in a warning display and a repeated request for a response. The subjects were told to go through the list at their own pace, which typically took about 30–35 min—about 3 sec/item.

Results

To ensure that there was no confounding of lag condition with serial position, the first 62 trials were treated as practice and dropped from all analyses. This resulted in an average of about 56 observations per subject in each P2 condition and 17 in each P3 condition. The data of 2 subjects who had recognition hit rates under 80% were dropped from further analysis. All of the analyses reported below are therefore based on an *N* of 23.

Recognition. Discrimination between new and old items was quite accurate. The mean percentages of *old* judgments on P1, P2, and P3 were 3.1%, 92.6%, and 98.9%, respectively. These means were qualified by three reliable ($p < .05$) linear trends: P2 performances declined as a function of Lag 1 (93.1%, 93.5%, and 91.2%), and P3 performance declined with both Lag 1 (99.4%, 99.1%, and 98.0%) and Lag 2 (99.5%, 99.3%, and 97.8%). There was no reliable Lag 1 × Lag 2 interaction.

JOR. Figure 3 shows mean JOR conditional on recognition. As was expected, the judgments on P2 increased with Lag 1, and the judgments on P3 increased with Lag 2. In addition, at each level of Lag 2, mean P3 judgments increased as a function of Lag 1. To quantify these effects, each subject's mean JORs were correlated with the appropriate set of lag values. Each Pearson *r* was then Fisher-transformed to minimize skewness, and the overall mean was calculated and inverse-Fisher-transformed back onto the *r* scale. On P2, the average *r* between mean

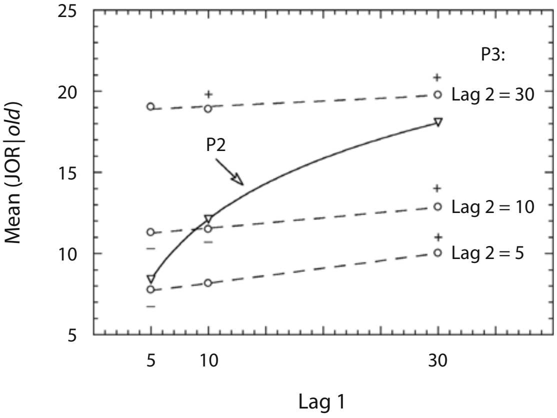

Figure 3. Mean judgment of recency (JOR) from Experiment 1. P2 and P3 are the item's second and third presentations. To find the nonrepetition control for a given P3 value, scan laterally to the point on the P2 curve that lies at about the same level. P3 means that are significantly higher or lower than their P2 controls are marked by "+" and "−", respectively.

JOR and Lag 1 was .982. On P3, the average r was .947 with Lag 2 and .153 with Lag 1. Even this last value was reliably different from 0 (each of the 23 subjects yielded a positive correlation). The slope relating JOR on P3 to Lag 1 tended to decrease with each step increase in Lag 2 (.092, .064, and .033, respectively), but the difference between the largest and smallest of these values fell just short of significance [$t(23) = 2.068, p = .051$].

Overall mean JOR on P3 was not significantly different from that on P2, which, taken by itself, would suggest an absence of any repetition effects, but this null comparison masks several reliable differences. P3 performances at each value of Lag 2 can be compared with P2 performances at the equivalent value of Lag 1, as the nonrepetition baseline. Of the nine different P3 conditions, seven had mean JORs significantly different from those of P2 ($p < .05$ by t test). Figure 3 shows that three different P3 means were significantly shorter (−) and four were significantly longer (+) than their respective P2 nonrepetition controls. With a single exception, mean JOR on P3 was shorter than on P2 when Lag 1 was short, and it was always longer than on P2 when Lag 1 was long.

Discussion

The observed effects of Lag 1 on P3 judgments are regular, but do not square with any extant hypothesis regarding repetition and JOR. At a strictly qualitative level, a P3–JOR increase with Lag 1 is consistent with both the cumulative-strength hypothesis (because trace strengths should sum) and the independent-trace hypothesis (because Lag 1 affects the overlap between P1 and P2 time-tag distributions). The result is similar in direction to the massed- versus spaced-repetition difference reported by Peterson et al. (1969), but it shows that the effect is not specific to massed repetition, because there were no massed (Lag 1 = 0) items in this experiment.

The truly puzzling aspect of the results is that, when Lag 1 was long, repeated items were judged significantly less recent than their nonrepeated controls. Neither the cumulative-strength nor the independent-trace hypothesis predicts this result. One might be tempted to suppose that on some P3 trials, the subjects had forgotten P2 and retrieved only P1 (which has an especially long recency of Lag 1+Lag 2). But to explain the obtained pattern in this way, one would have to assume—implausibly—that the forgetting of P2 depends on Lag 1 and not on Lag 2. A more promising approach is to assume that the feeling of recency that the subjects experienced on P2 is revived on P3, where it influences the new JOR. This idea is developed further in the General Discussion section.

EXPERIMENT 2

According to the independent-trace model of Flexser and Bower (1974), information on the recency of P1 should survive the encoding of P2. By contrast, the cumulative-strength explanation of JOR assumes that strengths attributable to P1 and P2 merge, which should destroy evidence of their individual recencies. The aim of Experiment 2 was to determine the extent to which subjects can remember two different recencies of an item by applying to JOR the independent-manipulation logic that was used with position judgments by Hintzman and Block (1971). In the case of repeated items, as is illustrated in Figure 2, the lag between P1 and the test on P3 (R first) was manipulated separately from the lag between P2 and the test on P3 (R last). On P3, some subjects judged the recency of the first occurrence (JOR first), and others judged the recency of the last occurrence (JOR last). As in Experiment 1, a JOR was requested only if the subject first responded that the test item was old.

Method

Subjects and Materials. Twenty-eight University of Oregon students served for course credit. On the hunch that the JOR-first task would be more difficult and the data less reliable, the subjects were assigned to the two instruction conditions unequally—17 to the JOR-first condition and 11 to the JOR-last condition. The master list of nouns was based on that of Experiment 1 but was expanded to 458 words.

Design and Procedure. The design required separate control over the spacings between P1 and P3 (R first) and between P2 and P3 (R last). Spacings were chosen so that both R first and R last were within the allowed JOR range of 5–30 items. To minimize the extent to which repeated words from the different lag conditions would fall into a repeating sequence, prime numbers (5, 11, 17, and 29) were used as the lag values.

As in Experiment 1, the subjects had to indicate on every trial whether the test word was old or new. Unlike in Experiment 1, however, there was no JOR test on a word until its final presentation. Experimental words were randomly assigned to nine different conditions. Four conditions had the JOR test on P2, at each of the lags: 5, 11, 17, and 29. The other five conditions had the JOR test on P3. The respective R first–R last combinations for these conditions were 11–5, 17–5, 29–5, 17–11, and 29–11.

To ensure a uniform distribution throughout the list, conditions were selected from the set of nine without replacement until the set was exhausted, and the process was then repeated. Stepping through positions $i = 1–600$, a word was selected from the master list at random, and P1 of the word was assigned to position i if that location was unoccupied. If the other positions (or position) dictated by the selected lag condition were free, repetitions of the word were assigned also to those positions. If the designated positions were not free, the value of i was incremented by 1 (the word was not repeated) and a new word was chosen—a process that iterated until the selected lag condition could be implemented. Cycling through the nine conditions in this way ensured that they occurred equally often.

On each trial, the subjects first decided whether the word was new or old, as in Experiment 1. A response of *old* was followed by a JOR test only if the item really was old (P2 or P3) and this was the item's final presentation. To motivate good recognition performance, the computer issued a beep for each false alarm on P1 and for each miss on P2.

The instructions informed the subjects that some words would be repeated in the list and that the research was concerned with the effects of repetition on the ability to remember a word's recency. The only difference between the JOR-first and JOR-last instructions was in the words *first* and *last*. The crucial passage said, "We want you to always judge the recency of the word's first (last) occurrence. If you saw the word once before, you should judge the recency of its only previous occurrence. But if you saw the word twice before, you need to ignore its last (first) occurrence and judge how long ago you saw it first (last). Don't worry if you find this difficult. What we want to learn from your judgments is how difficult it actually is." The importance of judging only the first or the last occurrence was reiterated just before onset of the list.

Results

Data from Serial Positions 1–31 were treated as practice and discarded, resulting in about 20 trials per condition per subject. One subject in the JOR-first instruction condition, who produced judgments that were negatively correlated with the target recency, qualified as a statistical outlier. That subject's data were dropped, leaving 16 in the JOR-first condition and 11 in the JOR-last condition. To avoid clutter, only findings of direct interest are reported in this section. Overall means, along with those from Experiments 3 and 4, can be found in Appendix A.

Recognition. There were no reliable differences in recognition performance between the two instruction conditions. Combining both data sets, *old* responses were made 5.2% of the time on P1, 90.4% on P2, and 99.2% on P3. As happened in Experiment 1, recognition performance on P2 decreased with lag, from 93.9% at lag = 5 to 86.4% at lag = 29 [$t(25) = 3.01, p = .006$]. On P3, recognition was effectively at ceiling.

JOR. On P2, mean JOR first (17.1) and JOR last (17.6) did not differ reliably [$t(25) = 0.85$], which is consistent with the view that both groups of subjects were judging the same (single) recency, as they were instructed to do. On P3, mean JOR first (18.3) was significantly greater than mean JOR last (12.9) [$t(25) = 6.82, p < .001$], as would be expected if the two groups were judging different recencies; however, this difference could reflect a response bias based on remembered frequency. That is, a subject might reason that, if a word occurred twice, R first must be relatively long or R last must be relatively short and adjust the numerical judgment accordingly. To determine whether subjects are differentially sensitive to R first and R last, one must examine how the judgments relate to manipulation of the two recencies, a question addressed below.

In Experiment 1, JORs to repeated items on P3 tended to be shorter than to P2 controls items when Lag 1 was short, and this result replicated in Experiment 2. There were two combinations of R first and R last where JOR last was reliably shorter on P3 than on the P2 control. These were the combination 11–5 (mean JOR last = 10.0 vs. 12.5 for the lag = 5 control) and the combination 17–11 (mean JOR last = 15.2 vs. 16.5 for the lag = 11 control) [$t(10) = 2.43$ and 3.29, respectively, both $p < .05$]. In both of these cases, the P1–P2 spacing was six items. There were no conditions in this experiment like the ones in Experiment 1 that produced reliably longer JORs on P3 than on P2, because the lag manipulations precluded large values of Lag 1 (i.e., large differences between R first and R last; see Figure 2).

To examine differential sensitivity to R first and R last, the 11–5 combination was dropped from analysis, leaving a simple orthogonal manipulation of the two recencies (R first = 17 vs. 29, crossed with R last = 11 vs. 5). In order to evaluate the dependencies, each subject's four JOR means were separately correlated with the values of R first and R last, and the correlations were Fisher transformed to minimize skewness. Figure 4 shows the means and 95% confidence intervals, transformed back onto the Pearson *r* scale. Squaring a given value of *r* estimates the proportion of variance in JOR first or JOR last that is explained by the R-first or R-last manipulation. Data presentations such as Figure 4 will be called *influence graphs* in this article.

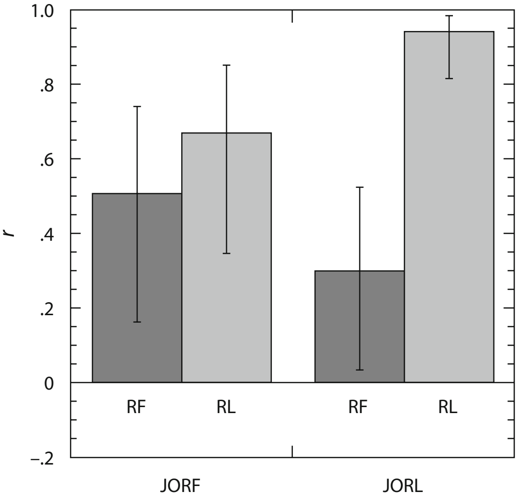

Figure 4. Influence graph for P3 of Experiment 2 (see text for explanation). JORF and JORL are judgments of R first (RF) and judgments of R last (RL), respectively. Error bars show 95% confidence intervals.

It is clear from Figure 4 that the two types of judgment were differentially affected by the two recencies. However, the difference lies primarily in the JOR-last condition, which was strongly affected by R last, and only weakly by R first. By contrast, JOR first was influenced to about the same degree by both recencies. This result suggests that subjects may be able to effectively ignore early presentations when judging more recent presentations, but not vice versa. The outcome is difficult to interpret, however, because there is no way to infer, on a given trial, whether the subject was remembering just one or both presentations. The average effect of R last on JOR first, for example, could arise from trials on which P2 was remembered but P1 was not. Experiment 3 explored this issue further.

EXPERIMENT 3

This experiment exactly replicated Experiment 2, with one exception: Instead of judging whether each word was old or new, the subjects indicated how many times it had been seen previously.

Method

Thirty-four University of Oregon students served for course credit: 17 in the JOR-first instruction condition and 17 in the JOR-last condition. The design and procedure were the same as those in Experiment 2, except that the subjects judged the number of times that they had previously seen the word (JOF = 0, 1, or 2), instead of giving a recognition response. The instructions were modified as appropriate for this change. JOFs were made on the "V," "N," and "<" keys of the keyboard, which had been labeled with the digits 0, 1, and 2 for this purpose.

Results

To ensure the reliability of mean JOR conditionalized on JOF, it was required that the subjects correctly respond with JOF = 2 on at least 40% of the P3 trials. Five subjects—2 in the JOR-first and 3 in the JOR-last condition—failed to meet this criterion. One additional JOR-first subject gave JORs that were atypically low for that condition ($z = -2.42$). The data of these 6 subjects were excluded from further analysis, leaving a total of 28 subjects, 14 per instruction condition.

As a first check on the effectiveness of replication, the JOF = 1 and JOF = 2 results were combined and treated as *old* recognition judgments and the data were compared with those of Experiment 2. There were no reliable differences between the experiments, either in recognition rates or in recency judgments (see Appendix A). Thus, there is no indication that requiring a JOF instead of a recognition judgment affected the discrimination of new and old items. Because the data when collapsed over JOF merely replicated those of Experiment 2, and collapsing masks several theoretically important differences, the data are segregated by JOF in all subsequent analyses.

JOF. Combining both conditions, the respective JOF = 0, 1, and 2 percentages were 96.8%, 3.1%, and 0.1% on P1; 8.1%, 87.3%, and 4.6% on P2; and 0.4%, 41.7%, and 57.9% on P3. There were no statistically reliable differences in JOF between instruction conditions.

JOR. On P2, when the subjects correctly reported JOF = 1, mean JOR first (15.8) and JOR last (15.3) did not differ significantly [$t(13) = 0.68$]. When the subjects incorrectly gave JOF = 2, JOR first (23.6) was on average longer than JOR last (15.1). This difference is consistent with a response bias based on remembered frequency, but the data are too sparse to permit a meaningful statistical test. On P3, when the subjects incorrectly reported JOF = 1, mean JOR first (15.4) was again longer than mean JOR last (12.4) [$t(13) = 3.42, p = .002$]. When the subjects correctly gave JOF = 2, the difference between instruction conditions was much larger (21.1 vs. 10.9).

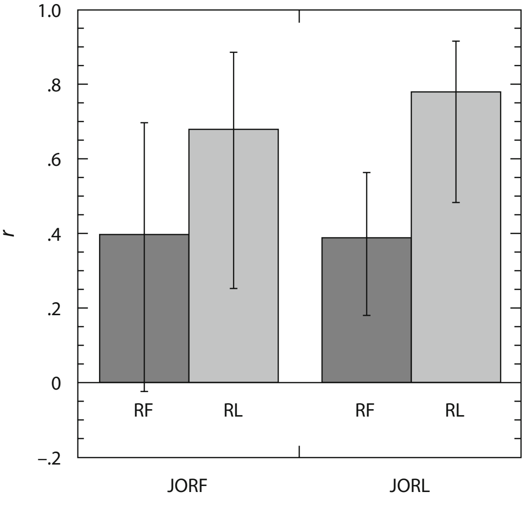

Figure 5. Influence graph for P3 of Experiment 3 on trials in which the subjects gave judgments of frequency (JOF) of 1. JORF and JORL are judgments of R first (RF) and judgments of R last (RL), respectively. Error bars show 95% confidence intervals.

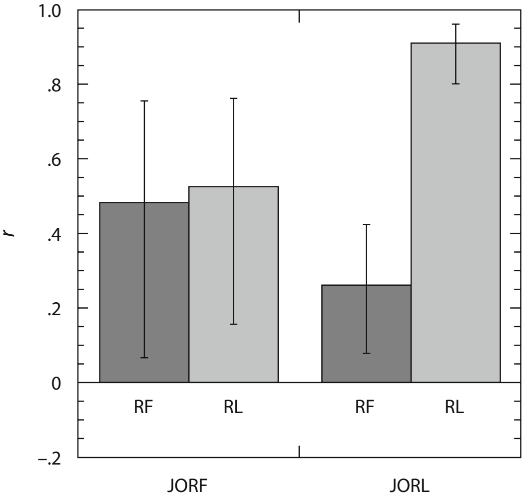

Figure 6. Influence graph for P3 of Experiment 3 on trials in which the subjects gave judgments of frequency (JOF) of 2. JORF and JORL are judgments of R first (RF) and judgments of R last (RL), respectively. Error bars show 95% confidence intervals.

The small but reliable JOR difference on P3 trials when the subjects incorrectly gave JOF = 1 suggests a tendency for the subjects in the JOR-first condition to retrieve only P1, and those in the JOR-last condition to retrieve only P2, which would be an interesting example of temporally biased retrieval. Frequency-based response bias cannot be ruled out as an explanation, however. It could be that, on some of these trials, the subjects felt uncertain and vacillated between JOF = 1 versus 2, eventually giving a biased JOR despite having settled on a JOF of 1. This consideration again suggests caution in interpreting overall differences between JOR-first and JOR-last means.

The influence graphs for P3, conditional on JOF = 1 and JOF = 2, are shown in Figures 5 and 6, respectively. It can be seen that when the subjects gave JOF = 1, there was little difference between the two recency judgments (Figure 5). P2 had a bigger influence than P1 regardless of the target recency. The fact that JOR first and JOR last were not differentially sensitive to the recencies of P1 and P2 suggests that the mean JOR difference that was found in this cell can be attributed to response bias.

When the subjects correctly gave JOF = 2, by contrast, there was a sizable difference in sensitivity to the two recencies (Figure 6). The unconditional data of Experiment 2 showed that JOR last was especially sensitive to P2, and we see here that this was primarily attributable to trials on which the subjects remembered there having been two presentations. Nevertheless, although the JOR-last subjects appear to have been able to ignore P1 when judging P2, the JOR-first subjects were influenced by P1 and P2 about equally. The result is an interaction, but not a crossover interaction like the one obtained by Hintzman and Block (1971) in the position-judgment task. Different materials and lags were used in Experiment 4, to see whether a crossover interaction could be obtained.

EXPERIMENT 4

The logic of this study was identical to that of Experiment 3, but a wider range of recencies was used and the stimuli were colored photographs, which generally surpass words in supporting accurate memory judgments. One hope was that the JOF = 2 percentage on P3 would be higher than it was in Experiment 2, giving more stable JORs means and a lower likelihood of item selection.

Method

Subjects were 39 students recruited as in the previous experiments. Twenty subjects were given the JOR-first instruction, and 19 were given the JOR-last instruction. The stimuli were 216 digitized color photographs taken from various electronic sources. They depicted a wide variety of scenes and objects and were chosen to be of low interitem similarity. On presentation, a typical picture occupied about 230 cm^2 of the computer screen.

The lags and list length also differed from those of Experiment 3. The lags for the items tested on P2 were: 5, 11, 18 and 35 items. Again, five combinations of R first and R last (11–5, 18–5, 35–5, 18–11, and 35–11) were tested on P3. To encompass the longer lags, the total list length was increased to 650. A final change from Experiment 3 was that JOF response latencies were recorded.

Results

Data from Positions 1–35 were first discarded, leaving about 22 trials per condition per subject. One JOR-last subject produced judgments on P3 that were wildly discrepant with those of the other subjects, correlating only .04 with the target recency. This subject's data were dropped from further analysis, leaving $N = 20$ in the JOR-first instruction condition and $N = 18$ in the JOR-last condition.

JOF. Frequency judgments were essentially the same in the JOR-first and JOR-last conditions and were highly accurate (see Appendix A). On P1, JOFs of 0, 1 and 2 were given 98.5%, 1.4%, and 0.1% of the time, respectively. On P2, these figures were 6.8%, 90.6%, and 2.6%, respectively; and on P3, they were 0.5%, 15.0%, and 84.5%, respectively. The low P3 error percentage precludes analysis of JOR conditional on JOF = 1, as was done for Experiment 3, but should ensure that item selection is not a problem.

JOF latencies also failed to differentiate the JOR-first and JOR-last instruction conditions. The mean JOF latencies were 1,965 and 1,717 msec on P2 and 1,380 and 1,280 msec on P3, respectively. These differences appear large, but within-groups variability was high, and neither difference was reliable [$t(36) < 1.20$ in both cases].

JOR. When a correct JOF = 1 was given on P2, mean JOR first was 16.0, and mean JOR last was 17.3, a difference that was not significant owing to high intersubject variability [$t(36) = 1.53$, $p > .10$]. As in Experiment 2, JOR first (22.4) tended to be longer than JOR last (18.2) when the subjects incorrectly gave JOF = 2, but this error was far too rare for the difference to be evaluated statistically.

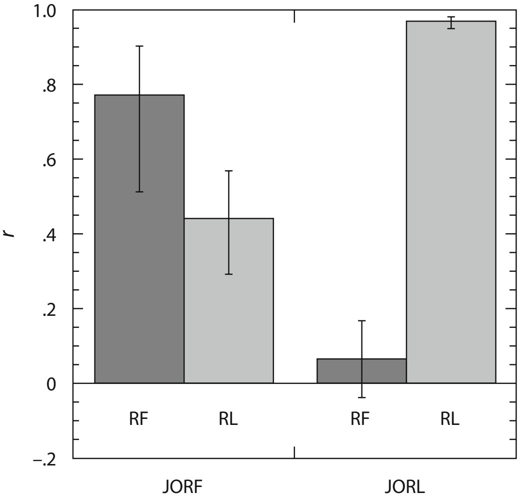

Figure 7. Influence graph for P3 of Experiment 4 on trials in which the subjects gave judgments of frequency (JOF) of 2. JORF and JORL are judgments of R first (RF) and judgments of R last (RL), respectively. Error bars show 95% confidence intervals.

On P3, when the subjects incorrectly reported JOF = 1, mean JOR first (16.0) was longer than mean JOR last (13.2). This difference cannot be evaluated statistically, but it is in the same direction as the significant difference found in Experiment 3. When the subjects correctly gave JOF = 2, mean JOR first (21.9) was again much longer than mean JOR last (11.0), also replicating the Experiment 2 outcome.

Figure 7 presents the influence graph for P3, conditional on a correct JOF = 2. As in Experiment 3, JOR last was focused more exclusively on the target recency than was JOR first, exhibiting a small and nonsignificant influence of P1 on the P2 judgment. JOR first was again influenced by P2, but not as heavily as by the target recency (P1). In contrast to Experiment 2, the graph displays a definite crossover interaction, showing that the subjects were able to judge the two recencies with a high degree of independence.

DISCUSSION OF EXPERIMENTS 2–4

The results of Experiments 2–4 suggest that, just as subjects can remember two serial positions of a repeated item in the position-judgment task, they can also remember two recencies of a repeated item in the running JOR task. Qualitatively, at least, this result appears consistent with the independent-trace theory that Flexser and Bower (1974) proposed to explain their recency- and distance-discrimination findings. The result appears inconsistent, however, with a cumulative-strength explanation of JOR. If strengths attributable to different presentations merge, subjects should not be able to apportion that strength to different presentations of the same item.

Before turning to a general theoretical discussion of the results, however, we need to confront a possible criticism of Experiments 2–4. To minimize confusion about the task, different groups of subjects were tested in the two instruction conditions, and this raises the possibility that the groups differed in their encoding strategies. Could the JOR-first and JOR-last subjects have prepared for their respective tests on P3 by encoding information differently on the two prior presentations?

Given the independent-groups design, there is no way to definitively rule this scenario out, but there are two strong arguments against it. First, it is hard to imagine what special encoding strategies the subjects could have deployed, given that they were under constant pressure to respond to each new instruction and each stimulus. The JOR-first subjects could not ignore P2, because they had to respond to it, with a recognition judgment or JOF and often with a JOR. Nor could the JOR-last subjects ignore P1, for similar reasons. In principle, because a JOR test came only on an item's final presentation, the absence of a JOR test on P2 could be a signal to forget P2 (for the JOR-first subjects) or to remember P2 and forget P1 (for the JOR-last subjects). But the subjects were not told that there would be only one JOR test per item, and it seems highly implausible that they could have discovered this regularity. Second, there were no reliable differences that would suggest that the subjects in the two conditions used different study strategies—not in the recognition rates, JOFs, JOF latencies, or mean JORs on P2. Moreover, in Experiment 3, when the subjects incorrectly reported that a repeated item had occurred only once, JOR first and JOR last were about equally sensitive to the two manipulated recencies (Figure 5). It therefore seems unlikely that the intended retrieval differences were contaminated by strategic differences in encoding.

GENERAL DISCUSSION

These experiments uncovered two new findings concerning the effects of repetition on JOR. First, although repetition can make an item seem more recent, it can also make an item seem less recent. Whether the repeated item is judged to be more recent or less recent than a nonrepeated item depends on the spacing of the repetitions (Experiment 1). Second, subjects can judge the recencies of two presentations of an item with some degree of independence, particularly when they correctly remember that the item occurred twice (Experiments 2–4). We discuss the second of these findings first.

The independent-trace model of Flexser and Bower (1974) was intended to fit forced-choice data, and it would have to be elaborated in several ways to be applied quantitatively to numerical JORs. It is nevertheless clear that the results of Experiments 2–4 are in general agreement with the model. Comparisons of JOR first and JOR last—in mean judgments, dependencies on JOF, and differential sensitivities to R first and R last—are consistent with the theory at a qualitative level. That is, when the subjects said that they remembered two prior occurrences, the two kinds of JOR differed in ways consistent with the subjects having remembered both P1 and P2. These differences were either nonexistent or not as extreme when the subjects reported remembering only one occurrence. The cumulative-strength account of JOR provides no way to explain this pattern of findings.

The results of Experiment 1, however, are not entirely consistent with the independent-trace model. According to the model, a repeated item should seem more recent than a nonrepeated item of the same recency when the P1–P2 spacing is short, because a short lag increases the overlap of the time-tag distributions (see Figure 1). The experiment confirmed this prediction. But when the P1–P2 spacing was long, the effect of repetition did not simply disappear, as predicted—it reversed in direction. At all three values of Lag 2, when Lag 1 was at the longest value (30), mean JOR on P3 was significantly longer than it was on P2 at the equivalent recency. It is tempting to attribute these especially long JORs to the subjects' remembering P1 instead of P2, but this would require substantial forgetting of P2 after just five intervening items, an idea that is not supported by any other aspect of the data.

I propose that the entire pattern of data may be explained by taking a more complex view of repetition and memory. From this perspective, the cumulative-strength and the multiple-trace hypotheses suffer from complementary weaknesses. The weakness of the cumulative-strength hypothesis is that P2 does nothing other than contact and strengthen the trace that was laid down by P1. The weakness of the multiple-trace hypothesis is that it is irrelevant whether P2 makes contact with the trace of P1, because such contact has no further effect on memory. I will argue that neither hypothesis, in its simple form, is consistent with subjective experience.

It is commonly argued that memory judgments may be based not just on a one-dimensional process of familiarity, but also on a multidimensional, subjectively rich process of recollection. In particular, experimental subjects may decide that a word was presented in a recent study list because they believe that they recall specific thoughts that went through their minds when the word was originally studied (Strong, 1913; see also Gardiner, Ramponi, & Richardson-Klavehn, 1998). This idea—that the phenomenal qualities of an experience are encoded and later retrieved—is relevant to the problem of repetition, to the extent that repetitions of an item are differently experienced. An intuitively plausible hypothesis is that the second presentation (P2) of an item reminds the subject of the first presentation (P1), and that this experience of being reminded is itself encoded into memory, where it remains available for later recollection. In everyday terms, people not only notice repetitions, they can remember that they noticed the repetitions.

This reminding hypothesis was first advanced by Hintzman and Block (1973) to explain an unexpected experimental outcome: On a surprise test following presentation of a list, the subjects could judge the spacing between P1 and P2 of a word much more accurately than the spacing between two unrelated words. Hintzman and

Block (1973) proposed that at the time of P2, information on the recency of P1 is incidentally retrieved and encoded into memory, making it available for retrieval on the later spacing-judgment test. By this view, a judgment of spacing is essentially a delayed recency judgment. In confirmation of this account, if P2 is replaced with an associatively related word—which should also remind the subject of P1—the subjects can also judge their spacing (Hintzman, Summers, & Block, 1975).

Other researchers have explained memory for temporal order using the same reminding mechanism. Given the presentation of two words in a previous list, if the words are associatively related, subjects are better able to tell which one came first and which second than they can if the two words are unrelated (Tzeng & Cotton, 1980; Winograd & Selway, 1985). The idea is that being reminded of *doctor* by *nurse* is subjectively different from being reminded of *nurse* by *doctor*. The encoding of the reminding experience thus preserves information about the words' order (see also Friedman, 1991).

More recently, the reminding hypothesis has been revived to explain memory for presentation frequency (JOF). The idea is that iterative reminding and encoding across multiple study trials on an item yields a recursive representation in which early remindings are embedded in later remindings. On a JOF test, the rememberer can use the depth of recursion of such a representation to estimate the test item's presentation frequency (Hintzman, 2004). As discussed in that article, this *recursive-reminding hypothesis* seems to explain several otherwise-puzzling findings from the frequency-judgment literature, including the absence of intentional learning effects, and the fact that young children are about as accurate at JOF as are adults (e.g., Hasher & Chromiak, 1977). Because recursive reminding should occur automatically in anyone with a functioning episodic memory, it has been argued that such recursive representations could underlie the development of children's understanding of number (Hintzman, 2008).

The present proposal is that recursive reminding can be extended also to JOR experiments in which items are repeated. Consider first the present Experiments 2–4. As was suggested earlier, the experience of being reminded entails not just a reference to the past, but to some distance in the past; hence, the reminding that is encoded on P2 typically includes information on P1's recency. As earlier researchers argued, if this embedded recency information is retrieved on P3, it can subserve a delayed judgment of P1–P2 spacing (Hintzman & Block, 1973; Hintzman et al., 1975). If this account of spacing judgments is accepted, then at the time of P3, the subject must have access to two pieces of recency information: immediate recency (referring to P2) and embedded recency (referring to the recency of P1 at the time of P2). Given this necessity, the subjects in the JOR-first instruction condition could have estimated R first by piggybacking the item's embedded recency on its immediate recency. By contrast, a subject in the JOR-last condition would have tried to focus only on the immediate recency and to suppress information about the earlier lag.

As a general expression, assume that judgments are made by giving differential weight to the two sources of recency information:

$$\begin{aligned} \text{judgment} = {} & \alpha\,(\text{immediate recency}) \\ & + \beta\,(\text{embedded recency}), \end{aligned} \tag{1}$$

where $\alpha, \beta \geq 0$. On P2, only the immediate recency information exists, so β is effectively 0, for both JOR instruction conditions. On P3, the subjects have some control over the weights, which they can adjust with some flexibility, depending on the judgment task. Under JOR-last instructions, the subject tries to maximize α and minimize β. In a spacing judgment experiment (e.g., Hintzman & Block, 1973; Hintzman et al., 1975), the subject tries to do the opposite—that is, to minimize α and maximize β. (The degree to which the subjects were able to minimize α cannot be determined from existing data, because immediate recency was not controlled in the spacing-judgment experiments.) Under JOR-first instructions, the subject tries to set both α and β to intermediate values, effectively summing the estimates of Lag 2 and Lag 1. The reason that it is especially difficult to judge R first independently of R last may be that it is hard to know what values of α and β will strike the right balance.

Appendix B shows how the average correlations shown in the influence graphs (Figures 4–7) can be reparameterized in terms of the weights α and β in Equation 1. What is important is the ratio of the two weights. For example, the correlations in Figure 7 imply that β was 63.6% of the value of α in case of JOR first, but only 6.7% as large as α in the case of JOR last. This shows how the subjects could have judged the recencies of the first and last presentations of a word by adjusting the relative values of α and β.

The same scheme can be applied to the mean JORs of Experiment 1 (Figure 3), providing that we make two additional assumptions. The first is that it is not possible when making a JOR to completely suppress the influence of embedded recency (i.e., to set β exactly to 0). There is abundant evidence that a single prior processing episode can affect performance on a wide variety of tasks—for example, as bias in the spelling of homophones (Jacoby & Witherspoon, 1982), as voice mimicry in shadowing (Goldinger, 1998), and as bias in gymnastics judging (Ste-Marie & Lee, 1991). There is no reason to suppose that JOR would be immune to such prior-processing influences. Thus, it is plausible to assume that retrieved information regarding Lag 1 will creep in and contaminate attempts to estimate Lag 2, which was the task charged to the Experiment 1 subjects.

The second assumption is that JOR is like other psychophysical judgments, in that an individual sensation is evaluated relative to a standard established by previous trials in the experimental context (e.g., Helson, 1964; Parducci, 1965). Thus, a subject will give a long JOR to the extent that the retrieved memory seems less recent than is typical for the experiment and will give a short JOR to the extent that it seems more recent than is typical for the experiment. The comparison with the contextual standard holds for embedded recency, as well as for immediate recency.

Taking the two assumptions together, on P3, a longer-than-average Lag 1 will tend to increase the JOR, and a shorter-than-average Lag 1 will tend to decrease the JOR, relative to the P2 control condition, which has no embedded recency. Qualitatively, this is a fairly good description of the data in Figure 3. That is, P3 judgments overshot the P2 controls when Lag 1 was long and undershot the P2 controls when Lag 1 was short. Appendix B shows that a one-parameter model based on the foregoing assumptions does a decent job of fitting the Experiment 1 data. The best fit is obtained when β is 15.4% of the value of α.

CONCLUDING COMMENTS

The present results are generally consistent with previous findings regarding repetition and JOR, but they pose challenges for extant theories. Contrary to the cumulative-strength account of JOR, information on the recency of an item's initial presentation survives a subsequent presentation of the item. Contrary to both the cumulative-strength and independent-trace accounts, repetition can make an item seem either more recent or less recent than a nonrepeated item. The findings are more favorable to an independent-trace account of JOR than to one based on cumulative strength, but neither hypothesis does an entirely satisfactory job of explaining the data.

I have proposed that these findings and those from several other memory-judgment experiments might be understood by focusing on the subjective experience of reminding and the idea that this experience is encoded in memory for later recollection. The reminding experience typically includes a sense of the remembered event's recency. Recency information that is embedded in a reminding, and revived by a later cue, may intrude into the task of judging the recency of the last presentation, making it seem either more recent or less recent than a nonrepeated control item (Experiment 1). Alternatively, a subject may focus on embedded recency to formulate a judgment of spacing (Hintzman & Block, 1973; Hintzman et al., 1975) or combine embedded recency with the recency of the preceding trial to judge the age of an earlier occurrence (Experiments 2–4). It may be worth adding that if both recencies are available—as is assumed—they might be compared in order to obtain evidence on the regularity of repetition intervals.

A point in favor of this hypothesis is that it is applicable also to other memory-judgment tasks. In emphasizing the retrieval of encoded remindings, the account is consistent with efforts to explain memory for order (Tzeng & Cotton, 1980; Winograd & Selway, 1985) and memory for frequency (Hintzman, 2004).

The recursive-reminding hypothesis does not qualify as a process model or memory theory, because it leaves open some basic questions. Crucially, it does not specify whether or how the trace of P1 is affected by reminding. One possibility is that remindings are appended to the trace of P1 as an elaborative cognitive context. Another possibility—consistent with the multiple-trace hypothesis—is that the trace of P2, with its embedded reminding, is completely separate from the trace of P1. Alternatively, the trace of P2 could reference the trace of P1 with a pointer, otherwise leaving it unaltered. On either of the last two views, the traces of P1 and P2 would be separate, but not interchangeable or independent, in the sense intended by Flexser and Bower (1974). A final possibility, suggested by animal research on the *reconsolidation* of retrieved memories (e.g., Abraham, 2006; Przybyslawski & Sara, 1997; Sara, 2000), is that the trace of P1 is erased by the act of retrieval and replaced by a trace representing the experience of P2, which includes the reminding. I know of no evidence that would favor one of these mechanisms over the others.

A potential advantage of the recursive reminding hypothesis is that it might be incorporated into theories that have been developed primarily for recognition memory, allowing their extension to other memory-judgment tasks. A minimal requirement is that the theory accommodate recollection. The crucial elements are an ability to maintain a complex mental state; to encode a representation of that state in memory; and to form a new, compounded mental state upon later retrieval of the representation. Such models might help focus memory researchers' attention on the fundamental but neglected question of how we use episodic memory to perceive the temporal patterns of recurrence that characterize the environment.

AUTHOR NOTE

Correspondence concerning this article should be addressed to D. L. Hintzman, Department of Psychology, University of Oregon, Eugene, OR 97403 (e-mail: hintzman@uoregon.edu).

REFERENCES

Abraham, W. C. (2006). Memory maintenance: The changing nature of neural mechanisms. *Current Directions in Psychological Science*, **15**, 5-8.

Bower, G. H. (1967). A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1, pp. 229-325). New York: Academic Press.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, **107**, 127-181.

Flexser, A. J., & Bower, G. H. (1974). How frequency affects recency judgments: A model for recency discrimination. *Journal of Experimental Psychology*, **103**, 706-716.

Friedman, W. J. (1991). The development of children's memory for the time of past events. *Child Development*, **62**, 139-155.

Galbraith, R. C. (1976). The effects of frequency and recency on judgments of frequency and recency. *American Journal of Psychology*, **89**, 515-526.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness & Cognition*, **7**, 1-26.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, **105**, 251-279.

Hasher, L., & Chromiak, W. (1977). The processing of frequency information: An automatic mechanism? *Journal of Verbal Learning & Verbal Behavior*, **16**, 173-184.

Helson, H. (1964). *Adaptation-level theory: An experimental and systemic approach to behavior.* New York: Harper & Row.

Hinrichs, J. V. (1970). A two-process memory-strength theory for judgment of recency. *Psychological Review*, **77**, 223-233.

Hintzman, D. L. (1969). Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Psychology*, **80**, 139-145.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551.

Hintzman, D. L. (2001). Judgments of frequency and recency: How they relate to reports of subjective awareness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1347-1358.

Hintzman, D. L. (2002). Context matching and judgments of recency. *Psychonomic Bulletin & Review*, **9**, 368-374.

Hintzman, D. L. (2003). Judgments of recency and their relation to recognition memory. *Memory & Cognition*, **31**, 26-34.

Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition*, **32**, 336-350.

Hintzman, D. L. (2005). Memory strength and recency judgments. *Psychonomic Bulletin & Review*, **12**, 858-864.

Hintzman, D. L. (2008). Recursive reminding and children's concepts of number. *Behavioral & Brain Sciences*, **31**, 656-657.

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, **88**, 297-306.

Hintzman, D. L., & Block, R. A. (1973). Memory for the spacing of repetitions. *Journal of Experimental Psychology*, **99**, 70-74.

Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Contextual associations and memory for serial position. *Journal of Experimental Psychology*, **97**, 220-229.

Hintzman, D. L., Summers, J. J., & Block, R. A. (1975). Spacing judgments as an index of study-phase retrieval. *Journal of Experimental Psychology: Human Learning & Memory*, **1**, 31-40.

Jacoby, L. L., & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology*, **36**, 300-324.

Lansdale, M., & Baguley, T. (2008). Dilution as a model of long-term forgetting. *Psychological Review*, **115**, 864-892.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, **95**, 492-527.

Malmberg, K. J., & Shiffrin, R. M. (2005). The "one shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 322-336.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective likelihood approach to the effects of experience in recognition memory. *Psychological Review*, **105**, 724-760.

Morton, J. (1968). Repeated items and decay in memory. *Psychonomic Science*, **10**, 219-220.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, **89**, 609-626.

Murdock, B. B., Jr., Smith, D., & Bai, J. (2001). Judgments of frequency and recency in a distributed memory model. *Journal of Mathematical Psychology*, **45**, 564-602.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, **72**, 407-418.

Peterson, L. R. (1967). Search and judgment in memory. In B. Kleinmuntz (Ed.), *Concepts and the structure of memory* (pp. 153-180). New York: Wiley.

Peterson, L. R., Johnson, S. T., & Coatney, R. (1969). The effect of repeated occurrences on judgments of recency. *Journal of Verbal Learning & Verbal Behavior*, **8**, 591-596.

Przybyslawski, J., & Sara, S. J. (1997). Reconsolidation of memory after its reactivation. *Behavioural Brain Research*, **84**, 241-246.

Sara, S. J. (2000). Retrieval and reconsolidation: Toward a neurobiology of remembering. *Learning & Memory*, **7**, 73-84.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, **4**, 145-166.

Ste-Marie, D. M., & Lee, T. D. (1991). Prior processing effects on gymnastic judging. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 126-136.

Strong, E. K. J. (1913). The effect of time-interval upon recognition memory. *Psychological Review*, **20**, 339-372.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's wordbook of 30,000 words.* New York: Teachers College, Columbia University.

Tzeng, O. J. L., & Cotton, B. (1980). A study-phase retrieval model of temporal coding. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 705-716.

Ward, J. (1893). Assimilation and association. *Mind*, **2**, 347-362.

Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology*, **9**, 418-455.

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition*, **2**, 775-780.

Winograd, E., & Selway, R. M. (1985). Reminding as a basis for temporal judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 262-271.

Yntema, D. B., & Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning & Verbal Behavior*, **2**, 65-74.

APPENDIX A
Means From Experiments 2–4

Table A1
Results From P2, Experiment 2

	JOR First		JOR Last	
R	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*
5	92.1	12.8	96.5	12.5
11	89.1	16.6	92.9	16.5
17	91.5	18.6	89.5	19.0
29	87.8	20.4	84.4	22.5
Mean	90.1	17.1	90.8	17.6

Note—R, recency or lag; JOR, judgment of recency; JOR First and JOR Last are the two instruction conditions.

Table A2
Results From P3, Experiment 2

		JOR First		JOR Last	
R First	R Last	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*
11	5	99.7	16.3	100.0	10.0
17	5	99.7	17.3	100.0	11.0
17	11	99.1	18.9	98.3	15.2
29	5	99.1	18.8	99.1	11.8
29	11	97.9	20.0	100.0	16.4
Mean		98.9	18.8	99.4	13.6

Note—R First, recency of the first presentation; R Last, recency of the last presentation. JOR First, judgment of R first; JOR Last, judgment of R last.

(*Continued on next page*)

APPENDIX A (Continued)

Table A3
Results From P2, Experiment 3

	JOF = 1				JOF = 2			
	JOR First		JOR Last		JOR First		JOR Last	
R	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*
5	86.7	11.5	89.6	10.7	4.9	20.0	5.5	11.4
11	87.7	15.4	91.6	14.7	4.2	24.3	4.9	14.1
17	86.6	16.7	90.0	17.1	4.1	23.6	5.2	16.6
29	84.2	19.7	85.4	18.6	4.1	26.4	3.1	18.4
Mean	86.3	15.8	89.1	15.3	4.3	23.6	4.7	15.1

Note—JOF, judgment of frequency; R, recency or lag; JOR First and JOR Last are the two instruction conditions.

Table A4
Results From P3, Experiment 3

		JOF = 1				JOF = 2			
		JOR First		JOR Last		JOR First		JOR Last	
R First	R Last	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*
11	5	43.1	13.5	36.5	10.2	56.6	20.0	62.8	8.9
17	5	39.4	14.8	39.3	11.0	59.5	20.0	60.7	9.3
17	11	47.0	16.3	46.8	13.3	52.3	21.4	53.2	12.8
29	5	37.9	14.5	39.0	11.8	61.5	21.4	60.6	10.5
29	11	43.7	18.0	43.9	15.6	55.9	22.5	56.1	13.1
Mean		42.0	15.4	42.3	12.4	57.3	21.1	57.6	10.9

Note—JOF, judgment of frequency; R First, recency of the first presentation; R Last, recency of the last presentation; JOR First, judgment of R first; JOR Last, judgment of R last.

Table A5
Results From P2, Experiment 4

	JOF = 1				JOF = 2			
	JOR First		JOR Last		JOR First		JOR Last	
R	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*
5	90.7	10.5	93.1	10.9	6.4	17.8	2.3	14.0
11	91.7	14.4	92.3	16.1	5.2	22.4	4.4	16.5
18	90.8	17.0	90.4	19.1	7.7	24.2	7.0	20.6
35	88.3	22.2	87.4	23.6	10.1	25.0	9.1	21.9
Mean	90.4	16.0	90.8	17.4	7.4	22.4	5.7	18.2

Note—JOF, judgment of frequency; R, recency or lag; JOR First and JOR Last are the two instruction conditions.

Table A6
Results From P3, Experiment 4

		JOF = 1				JOF = 2			
		JOR First		JOR Last		JOR First		JOR Last	
R First	R Last	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*	Percentage of *Old* Responses	*M*
11	5	17.5	14.9	16.1	11.2	82.5	19.1	83.4	8.6
18	5	15.2	15.1	12.8	10.3	84.3	20.8	86.9	8.8
18	11	16.2	17.2	14.0	15.3	83.6	22.2	87.4	14.1
35	5	14.6	13.3	13.9	10.4	84.7	23.3	85.6	9.2
35	11	16.0	18.7	13.8	16.7	83.8	24.3	85.7	14.4
Mean		15.5	16.0	13.6	13.2	84.1	22.7	86.4	11.6

Note—JOF, judgment of frequency; R First, recency of the first presentation; R Last, recency of the last presentation; JOR First, judgment of R first; JOR Last, judgment of R last.

APPENDIX B
Quantitative Fits to the Data

Modeling the JOR Means of Experiment 1

Assume that experienced recency is a logarithmic function of the actual lag and that a JOR is formed by comparing the recency experienced on the present trial with the distribution of recencies experienced in the experimental context. This may be accomplished by computing the z score of the natural logarithm of the lag compared with those for the whole distribution (lags = 5, 10, and 30). Assume further that the JOR—also standardized to eliminate assumptions of scaling—is a weighted average of the immediate recency and the earlier recency embedded in the retrieved memory, if there is one. Thus,

$$z(\text{JOR}) = \alpha z[\ln(\text{immediate lag})] + \beta z[\ln(\text{earlier lag})], \tag{B1}$$

where α and β are between 1 and 0.

For simplicity, the parameter α can be arbitrarily set equal to 1. On P2, because an earlier lag does not exist, β is set to 0. On P3, β is free to vary, yielding a one-parameter model; the earlier lag is Lag 1, and the immediate lag is Lag 2 (see Figure 3). Excel Solver was used to minimize the squared prediction error, yielding a best fit when $\beta = .154$ ($r^2 = .986$). Figure B1 shows the fit of Equation B1 to the data. (A linear model without the logarithmic transformation fits only slightly worse, yielding $\beta = .155$ and $r^2 = .975$.)

An implicit assumption of this model is that retrieval and encoding were successful on P2, and that the P2 trace is always retrieved on P3; thus, there are no P3 trials where the JOR is based on the recency of P1. Modifying this assumption would, of course, add a parameter and thereby improve the fit, but there is not much room for improvement on the fit of this simple model.

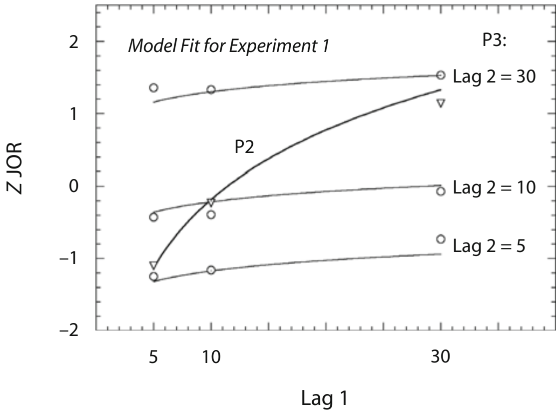

Figure B1. Fit of the one-parameter reminding model to the *Z*-transformed judgments of recency (JORs) of Experiment 1 (see Figure 3 for definitions).

Fitting the Influence Graphs of Experiments 2–4

Pearson r values, as plotted in Figures 4–7, express linear relationships, so the logarithmic transformation used above is unnecessary. By definition, immediate lag is R last, and embedded lag is the difference of R first – R last. Making these substitutions, the right side of the General Discussion section's Equation 1 can be rewritten as $(\alpha - \beta) * \text{R last} + \beta * \text{R first}$. If we denote the correlations of a judgment with R first and R last as r_F and r_L, respectively, then $\alpha - \beta$ must be proportional to r_L, and β must be proportional to r_F. Because our concern is with relative values of α and β, we can arbitrarily set $\alpha = 1$. Rearranging terms, we obtain $\beta = r_F/(r_L + r_F)$.

Taking Figure 7 as an example, the average correlations for JOR first are $r_F = .772$ and $r_L = .443$, which can be restated as $\alpha = 1, \beta = .635$. Likewise, the correlations for JOR last are $r_F = .067$ and $r_L = .969$, which can be restated as $\alpha = 1, \beta = .065$. Because α is fixed at 1, these are essentially one-parameter solutions. The data in Figures 4–6 can be reparameterized in the same manner.

(Manuscript received April 8, 2009;
accepted for publication April 20, 2009.)