

Short-term visual memory for location in depth: A U-shaped function of time

Adam Reeves¹ · Quan Lei¹

Published online: 6 July 2017
© The Psychonomic Society, Inc. 2017

Abstract Short-term visual memory was studied by displaying arrays of four or five numerals, each numeral in its own depth plane, followed after various delays by an arrow cue shown in one of the depth planes. Subjects reported the numeral at the depth cued by the arrow. Accuracy fell with increasing cue delay for the first 500 ms or so, and then recovered almost fully. This dipping pattern contrasts with the usual iconic decay observed for memory traces. The dip occurred with or without a verbal or color–shape retention load on working memory. In contrast, accuracy did not change with delay when a tonal cue replaced the arrow cue. We hypothesized that information concerning the depths of the numerals decays over time in sensory memory, but that cued recall is aided later on by transfer to a visual memory specialized for depth. This transfer is sufficiently rapid with a tonal cue to compensate for the sensory decay, but it is slowed by the need to tag the arrow cue’s depth relative to the depths of the numerals, exposing a dip when sensation has decayed and transfer is not yet complete. A model with a fixed rate of sensory decay and varied transfer rates across individuals captures the dip as well as the cue modality effect.

Keywords Visual working memory · Short-term memory · Binocular vision · Stereopsis

The time course for visual memory for depth is currently unknown. We studied this topic using the “partial-report” method of Sperling (1960, 1969), in which subjects report a

randomly cued row from briefly presented rectangular arrays of items such as letters. Partial report facilitates the study of visual memory because it evades the capacity limit of four items found with full report. Since the entire array must be attended in order to maximize performance, the “information available” can be taken as the report accuracy times the number of rows in the array (Sperling, 1960), which, for practiced subjects, comes to 11 letters with a flat 12-letter array. The available information decays over time in visual storage, as revealed by a steady decrease in accuracy to the level of full report as the cue is progressively delayed (Sperling, 1960, Fig. 7). Consistent with the characterization of visual storage as “iconic” (Neisser, 1967/2014)—that is, like a flat picture—adding clearly visible depth to Sperling’s letter arrays in tilted, convex, or concave profiles has almost no effect on partial report at any cue delay tested up to 1 s (Reeves & Lei, 2014). A flat icon and concurrent depth perception may seem contradictory, but one possibility is that the depth profile, once encoded, is rapidly transferred out of the icon to be retained in a distinct visual store; Xu and Nakayama (2007) found that adding depth to letter arrays slightly improved item reports at a retention interval of 2 s, supporting the idea that at least some depth information is retained. The purpose of the present study was to obtain data on the time course of information concerning location in depth, in order to explore such possibilities.

Stereoscopic disparity was used to create depth because it can be manipulated independently of the spatial configuration of the array, unlike depth cues such as interposition, perspective, and height in the visual field. Disparity signals depth even at the brief exposures needed to study sensory memory, since its integration time is about 100 ms (Harwerth, Fredenburg, & Smith, 2003). Because item recalls are used as probes in the partial-report paradigm, the loss of item depth over time has to be distinguished from the loss of items over time. Items may be lost because they have decayed away (Sperling, 1960) or have

✉ Adam Reeves
reeves@neu.edu

¹ Department of Psychology, Northeastern University, Boston, MA, USA

been misplaced (Mewhort, Campbell, Marchetti, & Campbell, 1981) or erased (Tijus & Reeves, 2004). We limited recalls to *one item per trial* so as to minimize these sources of item error and, ideally, to reveal any pattern of errors in the retention of depth. Here and throughout the article, we use the term “depth” as a shorthand for the locations of discrete items (numerals) in depth, given that we did not study the retention of scenes or of information varying continuously in depth. We present the basic depth retention curves in Experiment 1. A working memory (WM) load was imposed in Experiment 2 in order to test whether WM contributed to retention. A tonal cue was used in Experiment 3 to test the generality of the findings across cue modalities. Experiment 4 was a control experiment in which subjects counted backward by threes to eliminate verbal rehearsal, while retaining item depths. Experiment 5 demonstrated classic iconic decay with our equipment using flat Sperling-type displays. To account for the results, the General Discussion provides a model of visual memory for depth that assumes both exponential iconic decay and a posticonic integrating stage.

General method

Subjects A total of 44 Northeastern University undergraduates, 16 in each of Experiments 1 and 2 and 12 in Experiment 4, were run for just one session, and so were designated “naive.” A further five undergraduates were run for three sessions in Experiment 2; these were designated “experts.” Three more such experts were run in Experiments 3 and 5. All subjects had at least 20/20 visual acuity in both eyes, reported seeing distinct depth planes in a Julesz random-dot stereogram, had normal color vision as measured by the Ishihara plates in Experiment 2, and reported normal hearing in Experiment 3. They received credit for running whether or not they completed their experiment, although none in fact dropped out. The research protocol was approved by the institutional review board of Northeastern University.

Apparatus Subjects positioned their heads on a chinrest placed 61 cm from a ViewSonic Professional Series P220-f CRT monitor, 41 cm wide by 30 cm high. A speculum, formed from black cardboard and placed between the screen and the bridge of the nose, divided the screen vertically in half, such that the viewing area was 19 cm wide for each eye. Chair height was adjusted so the subject could place her or his chin comfortably on the chinrest, with eye height about 10 cm from the top of the screen. The head was held by pads at each temple and by a forehead rest. The left and right eye images were 10.5 cm apart, center to center—that is, 9.8° when viewed from 61 cm away. The vergence needed to obtain fusion was aided by a nominal 20-diopter crown glass prism placed base-out in front of the left eye. Positive lenses (+1.5 diopters) before each eye placed the image at optical infinity

and therefore broke the normal correlation between accommodative blur and distance. These lenses could be adjusted laterally by millimeters as needed to ensure perfect fusion. Thus, disparity remained the only depth cue. The head could be removed and replaced without disturbing the optics.

Custom code was written in MATLAB 6.5 to create stimuli using a Cambridge Research System 5.1 driver, specialized for a Dell Optiplex PC and P1130 ViewSonic 100-Hz raster-based monitor. The experimenter selected the disparities, number of stimuli, and display times with a software menu. In the “movie” mode used to control the display sequences, which overruled interrupts from the Windows XP operating system, the stimulus duration was within 1 ms of that programmed, as per the manufacturer’s specifications and checked with an oscilloscope and photo-detector. A 40-W tungsten bulb dimly illuminated the back wall of the experimental room.

Stimuli A vertical array of numerals, one per row, was shown in every experiment; Fig. 1a portrays the depth profile, and Fig. 1b the spatial profile. Numerals were presented on a midgray, 20 cd/m^2 field, 7 cm wide by 14.5 cm high (Fig. 1b). The numerals, fixation mark, and cue were white (116 cd/m^2). The rest of the screen, and the entire screen between trials, was a dark gray (8 cd/m^2). The fixation mark, a white asterisk, was presented 0.5 cm above the top of the field at the start of each trial. When presented, numerals descended from a starting position 2.5 cm below the asterisk, with each numeral occupying a different and distinct depth plane (Fig. 1a). The numerals were 9 mm high and 4–8 mm wide, subtending 50 by 22 to 44 arcmin, and were displayed in Ariel font (sans serif), with one 0.455-mm screen pixel subtending 2.56 arcmin at the eye. Numerals were spaced 2.7 cm apart center to center, and thus were separated by gaps of 18 mm, far enough apart to appear distinct and, when presented in depth, to escape crowding (Toet & Levi, 1992). The cue and numeral durations were always equal.

The array of four numerals appeared to cascade down and toward the subject, as programmed and as is illustrated in Fig. 1a in the “fixed-order” condition, in which numerals were shown in their natural order; we established this effect by asking each subject to point out the depths of the numerals with a ruler as the numeral array was presented continuously. For a numeral seen X cm from the monitor screen, the retinal disparity was $\delta = 3,438PX/(D^2 + DX)$ in arcminutes, where P is the interpupillary distance and $D = 61$ cm, the eye–screen distance. The disparities (δ) were 0, 12, 34, 55, and 77 arcmin in Numeral Depth Planes 1–5, corresponding to $X = 0, 2.4, 7.3, 13,$ and 20 cm for a subject with $P = 6.4$ cm. These disparities differed enough that the depth order was clear, but they were not so large as to defeat fusion, since the upper limit for reliable localization in depth is 240 arcmin disparity (Blakemore, 1970). Thus, accurate localization in depth was expected, and indeed, all the subjects could

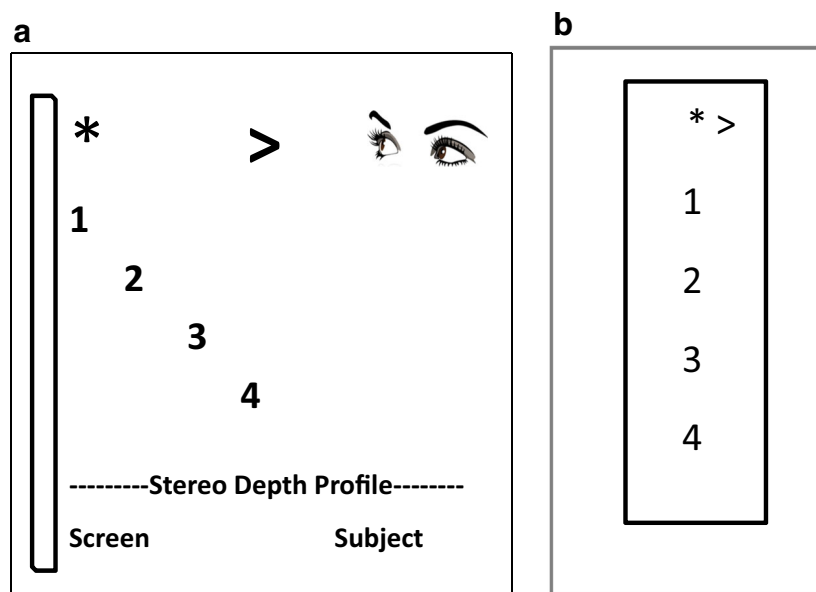


Fig. 1 (a) Depth profile. The fixation (*) and arrow cue (>) appeared near the top of the monitor screen. An array of four numerals descended below the top, one numeral per row. Depth, as determined by disparity, is portrayed horizontally, from the screen at left (disparity 0) toward the subject. The fixation had disparity 0, and the arrow the disparity of one

of the numerals. In the trial illustrated, larger numerals were both lower down the screen and closer to the subject, and the cue had the depth of numeral 4. (b) Spatial profile. Numerals and the arrow cue were presented in white on the gray field; the screen was dark.

point out the depth of each numeral and of the arrow cue. During the experiment, we emphasized that any loss of depth should be reported immediately, because this could indicate poor accommodation, poor lens positioning, loss of fusion, or incorrect vergence, but because the system was optically stable, such reports were infrequent.

Procedure Subjects were asked not to rehearse the array of numerals verbally, but rather to retain a visual impression of the array and thereby report the cued numeral once it had been presented. Those who found this task unnatural were permitted to name the numerals at first but were instructed to refrain from doing so after practice. The experiments were restricted to reporting from four or five rows of single items in order to avoid chance responding, since some individuals found it hard to report items from more rows than this.¹ Trials began with a

flat, 2-s array of four small crosses (or five, for the experts), one in each of the screen positions to be occupied by the numerals, to ensure vergence on the monitor screen and guide attention to the upcoming spatial positions of the numerals. On every trial, the arrow cue was presented at the same disparity as the randomly selected numeral that it cued (Fig. 1a). The cue delay was defined as the interstimulus interval (ISI) from the offset of the array of numerals to the onset of the arrow, and the stimulus onset asynchrony (SOA) was defined as the ISI plus the numeral duration. The arrow was presented either simultaneously with the numerals (SOA = 0) or after them (SOA > 0); negative SOAs were not used. Each subject was asked to report the numeral cued by the arrow on every trial, by guessing if necessary.

Numerals were presented in one of three conditions. In the *fixed order* used in practice, the array was identical from trial to trial, with the numeral 1 at the top and farthest away in depth, and larger numerals appearing progressively lower and closer to the subject, as in Fig. 1a. In the *random order* used for the naive subjects in the experimental trials, the top numeral was always farthest away and the bottom numeral closest, but the numeral order was chosen at random on each trial. In the *fully random order* used for the experts after the first session, both numeral order and depth order were randomized independently, so that, for example, the randomly chosen top numeral might be at any one of the depths. Because the items were known in advance (1 . . . 4 or 1 . . . 5) and never varied, loss of item information was unlikely to affect the experimental outcome.

¹ One expert subject, an undergraduate (D.L.), had previously learned to see and distinguish up to eight transparent RDS depth planes in the same apparatus, showing excellent stereopsis. However, even after some weeks of practice he never learned to report single items from more than four rows in depth. Using the numeral 8 as a cue in place of the arrow, to ensure that all the stimuli were from the same category, did not help D.L., nor did providing small dots before each trial at each possible depth plane as guides. D.L. reported that when more than four numerals were presented, they seemed to float in depth, disassociated from the depth of the arrow. In contrast, the best expert subject, author A.R., age 65, could report the depths of single items in eight rows after minimal practice and did not experience disassociation. Such large individual differences are likely not due to optics (given the presentation at optical infinity) or eye position, memory capacity, or overall perceptual or cognitive ability, but may reflect, as D.L. suggested, variations in the ability to associate the perceived depth of a cue to the memory of the depths of the numerals.

Since the identities and spatial dispositions of the numerals and arrow on the screen were the same from trial to trial (Fig. 1b), the subject needed to encode and retain only the stimulus depths in order to report which numeral was in the depth plane indicated by the arrow. Reports were made on a keypad that could be operated by feel, so that the subject did not have to look away from the screen. Subjects placed their first, second, and third fingers over the “1,” “2,” and “3” keys, and reached forward to access higher numbers (“4” or “5”) as needed. The fourth (little) finger was used to hit Enter, which caused the answer to be recorded. This was easy because the keypad was familiar and had tangible bumps surrounding the central key. Because numbers above 4 (for naive subjects) or 5 (for experts) were never recorded, we inferred that keystroke errors were not made.

Experiment 1

In this experiment we investigated the short-term retention of depth order by naive subjects. Their accuracy for reporting cued numerals presented at different depths was measured as a function of the delay time from the numeral array to the arrow cue. The *random-order* condition described in the General Method was used, in which the top numeral was always farthest away and the bottom one closest, but the numeral order was chosen at random on each trial so that the subject could not anticipate which numeral would be at which depth.

Method

Subjects A total of 16 naive subjects participated, each for a single session of somewhat under 1 h. In pilot work, we found that naive subjects could report a single numeral in the cued depth plane on most trials when the numeral array comprised four or five numerals, but not more than this, so a display of four numerals was chosen for all. Two potential subjects who only managed three numerals were replaced.

Procedure In each block of 24 trials, the delay (the ISI) from the offset of the numeral array to the onset of the arrow cue was increased progressively. This “ascending” order of ISIs had been used to trace the decay of visual storage by Sperling (1960), since the cue was maximally useful at simultaneity and subjects would continue to use it even as accuracy fell off; given the opposite, “descending” order, this might not be the case.

The stimulus duration (i.e., the durations of both the numeral array and the cue, which were always the same) was initially 800 ms for all subjects. The stimulus duration was then reduced to 200 ms, but if performance then

fell below 40% at simultaneity, those subjects were returned to 800 ms in order to avoid chance responding at longer ISIs. Thus, the subjects run at 800 ms were likely less able than those run at 200 ms. Whereas 200 ms is too brief for eye movements or vergence movements to occur during stimulus presentation (Haddad & Steinman, 1973), this is not so at 800 ms. Lacking direct evidence that the 800-ms subjects obeyed the instructions to keep their eyes fixed, their data are not definitive. Data collection was continued until there were eight subjects in each group. The SOAs for the four delays were 0, 200, 700, and 1,700 ms for the 200-ms group, and 0, 800, 1,300, and 2,300 ms for the 800-ms group. The delays were run in ascending order four times in each block, so that there were 24 trials per delay per block. Trials took from 2 to 3 s to run, depending on the subject’s work rate. After about 20 min of practice in the *fixed order*, these naive subjects completed either two or three blocks of experimental trials, totaling 48 or 72 trials per delay. The *random order* (of numerals) was employed, not the more difficult *fully random order* (of both numerals and depths) we used with the experts.

Results

Report accuracy formed a U-shaped “dip” function of cue delay for 14 of the 16 naive subjects, and no subject showed classic iconic decay. To characterize the dips, we defined the dip size as the minimum of the accuracies at SOA = 0 and at the longest SOA minus the lowest accuracy at all SOAs. The mean dip sizes were 7% for the 200-ms subjects [$t(7) = 4.46, p < .05$] and 11% for the 800-ms subjects [$t(7) = 2.43, p < .05$]. The dip SOA was defined as the SOA at which a subject’s accuracy was lowest (this being the longest SOA for the two subjects who performed without dips). The mean dip SOA for the 200-ms subjects was 340 ms, and that for the 800-ms subjects was 1,090 ms—a difference of 750 ms, partly accounted for by the 600-ms difference in display duration, and partly by the group difference. The dip times and sizes were rather scattered, but subjects who performed better tended to have later dips, with the correlation between dip SOA and mean performance in d' units being significant ($r = +.54$), [$t(14) = 2.26, p < .05$]. The subjects varied considerably in overall accuracy. Figure 2 presents mean accuracies versus SOAs for the four best and four worst naive subjects run with 200-ms displays (top) and for the four best and four worst run with 800-ms displays (bottom), with subjects being separated in this way to show that dips occurred among both the best and the worst subjects. The bars in this and subsequent figures show ± 1 standard error computed conservatively—that is, without removing the subjects’ grand means. Dips still

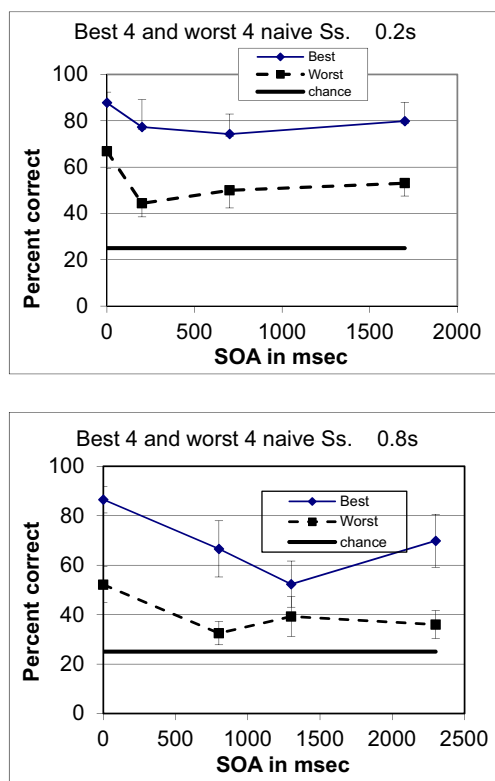


Fig. 2 Mean accuracy for the four best naive subjects (upper curves) and the four worst ones (lower curves) in Experiment 1, plotted against stimulus onset asynchrony (SOA). Top panel: Stimulus duration 200 ms. Bottom panel: Stimulus duration 800 ms. Bars show ± 1 standard error across subjects. The data reveal unexpected dips that became the focus of the rest of the study. The horizontal line in each panel depicts chance (25%).

appear in the mean data in Fig. 2, even though the individual data (shown later, in Figs. 9 and 10) are smeared by averaging.

Discussion

The results of Experiment 1 were unexpectedly U-shaped, since a progressive fall-off in accuracy over time is typical of iconic memory. True, Sperling (1960) observed a dip at an SOA of 150 ms in one subject and attributed it to a suboptimal strategy at just that SOA—namely, guessing in advance which row might be cued rather than attending equally to all rows and waiting for the cue. However, our conditions favored equal attention to all rows, since the chance of guessing the correct row was low and only single numerals had to be reported. Assuming, therefore, that our dips reflect processing rather than poor strategy, one wonders whether they might have been due to crowding or masking at short SOAs, since flanking masks can disrupt any following cues (Wilschut, Theeuwes, & Olivers, 2013). However, the arrow was 1.9 deg from the closest numeral and was flanked on only one side, avoiding both metacontrast (Weisstein, 1971; Lefton, 1973) and crowding

(Bouma, 1972). Moreover, a post-hoc data analysis showed that the error rates for reporting the numeral closest to the cue differed from the error rates for reporting the numeral most distant from the cue unsystematically, on average by less than 4% at each SOA, including at the dip, and by 1.7% averaged over SOAs. We therefore explored other possibilities—namely, working memory (WM) in Experiment 2, sensory modality in Experiment 3, and verbal memory in Experiment 4.

Experiment 2

To determine whether the unexpected dips seen in Experiment 1 were related to retention in WM, we imposed a second task of concurrently retaining colored shapes. This task was adapted from Luck and Vogel (1997), who concluded that WM combines shape, color, and orientation to form “visual objects,” rather than retaining these features separately.² The task was tailored for each subject so that WM would be loaded by either four or five colored shapes, the capacity range indicated by Luck and Vogel for these types of stimuli (but not for all; see Alvarez & Cavanagh, 2004). In theory, if depth was encoded into the same WM as shape and color, the load would depress the depth recall rate; indeed, if the colored shapes fully occupied WM, recalls might then come from sensory memory alone and show the expected (iconic) drop-off with SOA. Alternatively, if the memory for depth is independent of that for color and shape, the dip should remain even if retention was reduced overall by load.

Method

Subjects A group of 24 new subjects, denoted “naive,” were drawn from the Northeastern subject pool. Each ran in a single session of 60 to 75 min. A further six subjects from the pool, denoted “experts,” were run for three sessions, the first being treated as practice and the second and third sessions being analyzed to determine whether the results seen with naive subjects survived experience. Naive subjects were run with 200-ms or 800-ms displays, as in Experiment 1. All the experts were run with 200-ms displays; two potential “experts” were dismissed when they did not reach criterion accuracy at SOA = 0 in the first hour.

Procedure The subjects in the *WM load* condition saw a sample array of either four shapes (square, circle, triangle, and rectangle) or five shapes (these four, plus a bar), if they were above 90% correct in reporting four shapes during practice. Each shape had a

² The concept of WM introduced by Baddeley and Hitch (1974) has come to include both a memory buffer and an executive function. Although these ultimately need to be distinguished, the methods used here did not do so, and we retain the portmanteau usage.

distinct color—red, yellow, blue, violet, or green—on each trial; the color–shape assignment was randomized on every trial. The shapes were each 1.8 to 2.6 cm wide and were randomly scattered over the gray field with at least 4 cm between them. The WM load imposed was to retain the sample color–shape array for comparison with a “test” array presented after the subject had reported the numeral in the cued depth plane. The sample and test arrays were both 2 s in duration. During presentation of the colored shapes, the fixation target was turned off and subjects could look where they wanted; its re-presentation after the sample array signaled subjects to refixate during the 1-s blank period between the sample and the numeral array, in preparation for the depth-retention task. The test array was the same as the sample on a random half of the trials; on the remaining trials, one randomly chosen shape changed color. Subjects keyed 1 for “same” and 2 for “change” at the end of the trial. Subjects were asked to respond to each task quickly, but not at the expense of accuracy. They were also told that their response times were being recorded. In the *no-load* condition, the sample and test arrays were turned off, and only the numeral report was made.

The depth task was the same as in Experiment 1, with the numeral–cue delays run in ascending order. (Note, however, that a pilot experiment with six additional subjects showed that reversing the order of the SOAs had almost no effect on the reports, as might be expected, since the presence of the dip obviated the advantage of using the ascending method that had been noted by Sperling, 1960.) After brief practice with the *fixed order*, the naive subjects were run in *random order* with four delays, whereas the experts were run in the *fully random order* with five delays (see the General Method). WM load and no-load trials were run on different naive subjects, for a total of four blocks of trials per subject, whereas for the experts WM load and no-load conditions were alternated across eight blocks of trials. Trials averaged 7 s each. Every delay was run 24 times in each block of trials. Thus, for each naive subject, run in one session, there were a grand total of 384 experimental trials (96 per delay), and for each expert, run in two experimental sessions after the practice session, there were a grand total of 960 trials (192 per delay). Sessions were a little longer than in Experiment 1, to accommodate the additional trial blocks.

Results and discussion

Mean accuracy is plotted as a function of numeral–cue asynchrony (i.e., SOA) in the top panel of Fig. 3 for the eight naive subjects in the WM load condition and the eight naive subjects in the no-load condition (shown by different lines and symbols). All these subjects were run with 200-ms displays. The figure shows mean accuracy over SOAs—namely, 75.8%, 58.0%, 61.2%, and 66.5% at SOAs of 0, 200, 700, and 1,700 ms—once more showing an overall dip despite smearing over subjects’ different dip times. Analyzing the individual data, the mean dip SOAs were 477 ms with WM load and 461 ms with no load,

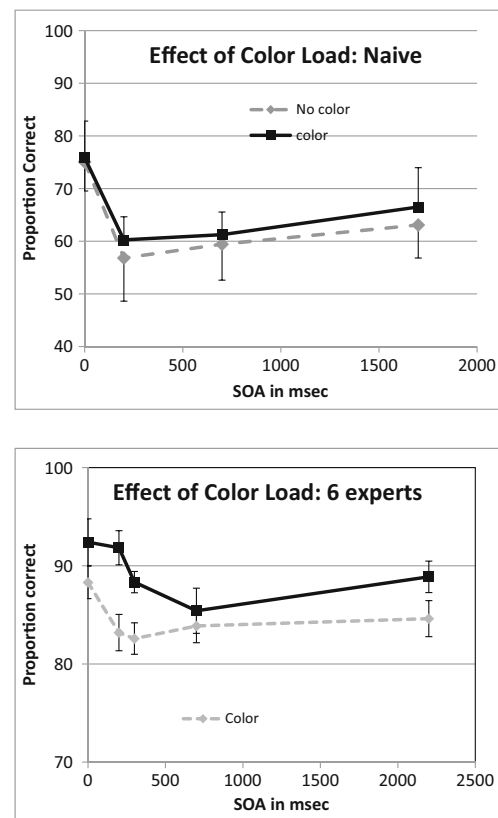


Fig. 3 Mean accuracies in Experiment 2 for 16 naive subjects (top panel) and six experts (bottom panel), plotted against numeral–cue SOA (in milliseconds). Each panel compares WM (color) load and no-load conditions. All displays were 200 ms in duration. Bars show ± 1 standard error across subjects.

and the mean dip sizes were 17.8% with WM load [$t(7) = 3.15$, $p < .05$], and 25.8% with no load [$t(7) = 2.58$, $p < .05$], showing similar dips with and without load.

As in Experiment 1, naive subjects unable to reach criterion with 200-ms displays were run with 800-ms displays, and these eight subjects were run only in the WM load condition. The mean accuracies with WM load for these subjects were 75.9%, 53.4%, 58.4%, and 64.9%, at the same SOAs as in Experiment 1 (namely, 0, 800, 1,300, and 2,300 ms). Their individual dip sizes and SOAs averaged 27.1% and 1,020 ms, similar to the no-load dips found in Experiment 1 (30% and 1,090 ms) for the 800-ms subjects. It seems that WM load had as little effect on the 800-ms as on the 200-ms subjects; dips always occurred, whether or not there was load. SOA was blocked over trials, so color–shape accuracy might have been sacrificed through a trade-off at later SOAs to keep numeral depth reports high; however, the correlations across SOAs between numeral accuracy and accuracy in reporting the color change averaged $r = .063$ (standard error .11), indicating that no such trade-off occurred, and suggesting instead that WM and memory for item depth are independent.

Mean accuracy is plotted in the bottom panel of Fig. 3 for the six experts, also with 200-ms displays, with the data

collected after an initial hour of practice. Each ran in both WM load and no-load conditions. Critically, we observed no substantial difference in the magnitude of the dip due to load; the overall effect of load, averaged over SOAs, was just 2% for the naive subjects and 4% for the experts. Interestingly, Fig. 3 suggests that load hastened the dip for the experts, but their individual data, plotted in Fig. 4 against ISI (i.e., SOA = 200 ms, the stimulus duration), show that any hastening effect was inconsistent. At any rate, apart from subject I.N. (lower right panel), in no other subject did the WM load generate the pure iconic decay

that would be expected if the colored shapes fully occupied WM for depth and forced recalls to draw on sensory memory alone. We conclude from these data that the cue–numerical interaction that creates the dips in depth recall occurs independently of the featural contents of WM.

Experiment 3: Tonal cue

Cue-type effects in visual short-term memory experiments are well documented. A change of cue that merely slows

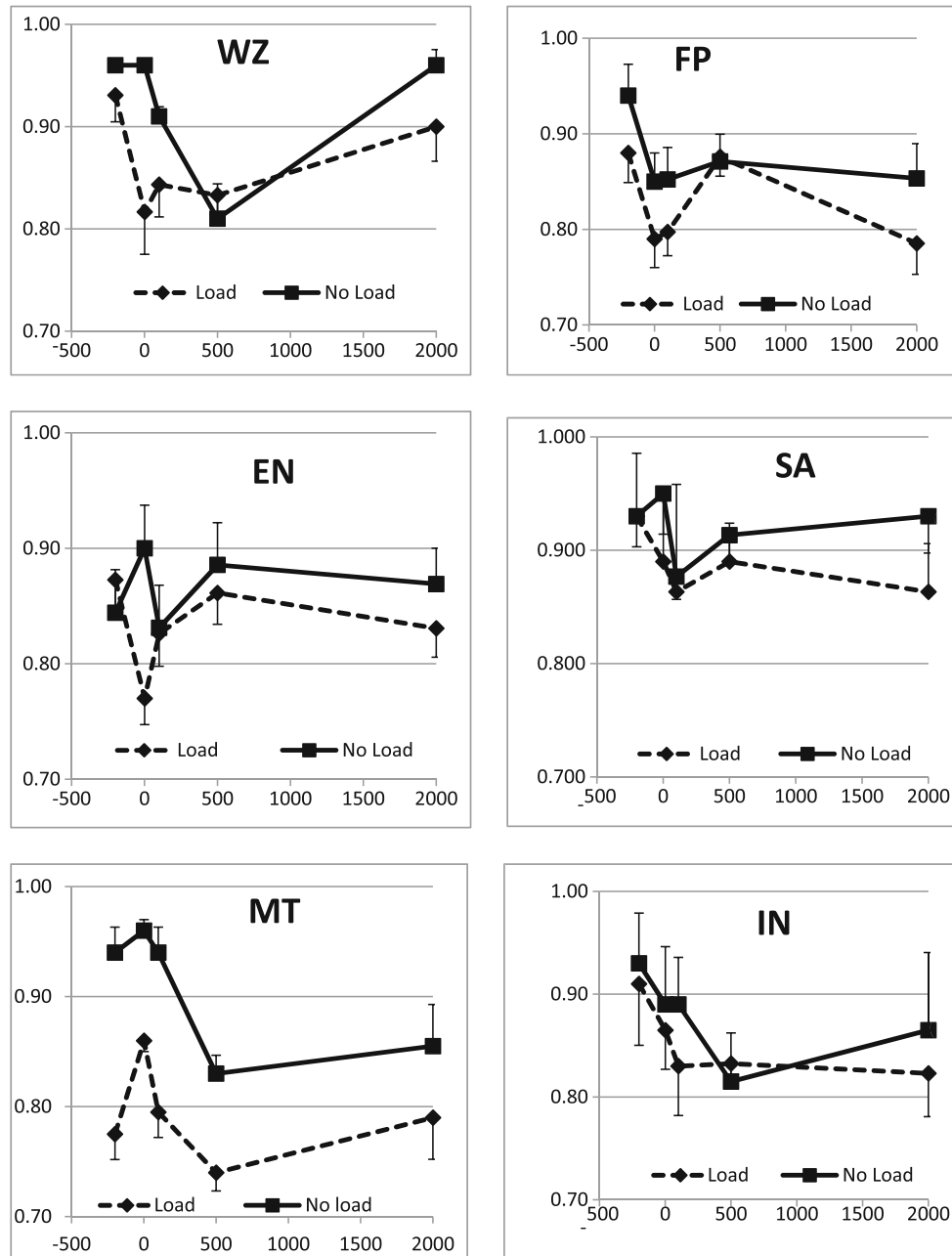


Fig. 4 Individual data for the six experts run in Experiment 2. Each panel compares accuracy in the WM load and no-load conditions across interstimulus intervals (ISIs). Load had inconsistent effects. Bars show ± 1 standard error within individual subjects.

identification will shift the decay curve laterally, but if the process required to identify the cue changes, the decay curve may also change in nature. For example, mislocations account for much of the data when a bar pointer is used to cue an individual item (Mewhort et al., 1981), but not when the cue is used to select an entire row (Sperling, 1960). In Experiment 3, our results with the arrow cue were compared to results with a *tonal* cue, to explore whether the dips survived this change of modality. If the dips reflect interactions in an abstract multimodal memory for layout, they should remain with auditory cuing, but if they reflect interactions within the visual system, they should disappear.

Method

Procedure Single-item rows and large stimulus spacings were again used, as in Experiments 1 and 2, to minimize spatial mislocations and other causes of item error. Three new experts first learned to associate a 2000-Hz tone with the farthest depth and progressively lower tones with closer depths, the pitch–depth associations being learned to 100% accuracy. Three conditions were run in separate blocks of 24 trials each—namely, tone alone (A), arrow alone (B), and both together (C). The tones and arrow were all 200 ms in duration. The condition order was ABC for one expert, CBA for the second, and CAB for the third. As before, accurate perception of the depths of the numerals and the arrow was checked initially. Five

cue delays were employed in the *fully random* condition (see the General Method).

Results and discussion

Mean accuracy is plotted against numeral–cue delay (ISI) in Fig. 5 for the three experts. The usual dip was obtained with the arrow cue alone (the bottom curve), but not when the tonal cue was presented without the arrow, in either the individual data or the group average (middle curve, dashed); there was no trace of a dip, or indeed, of any decay. When both the arrow and the tone provided cues (top curve), accuracy was higher than with either cue alone, but it was below the 96% predicted by independent processing of the two cues, suggesting either weak interference between the cues during each trial or an inability to attend to both of them on every trial. These data suggest that a *visual* interaction between the trace of the numerals and that of the arrow cue is required to obtain a dip, since the dip disappeared with a tonal cue.

Experiment 4: Backward counting

One account of the dips found in Experiments 1 and 2 is that subjects verbally recoded the displays and rehearsed this information during the trial. Such dips could have emerged had subjects used verbal memory to supplement

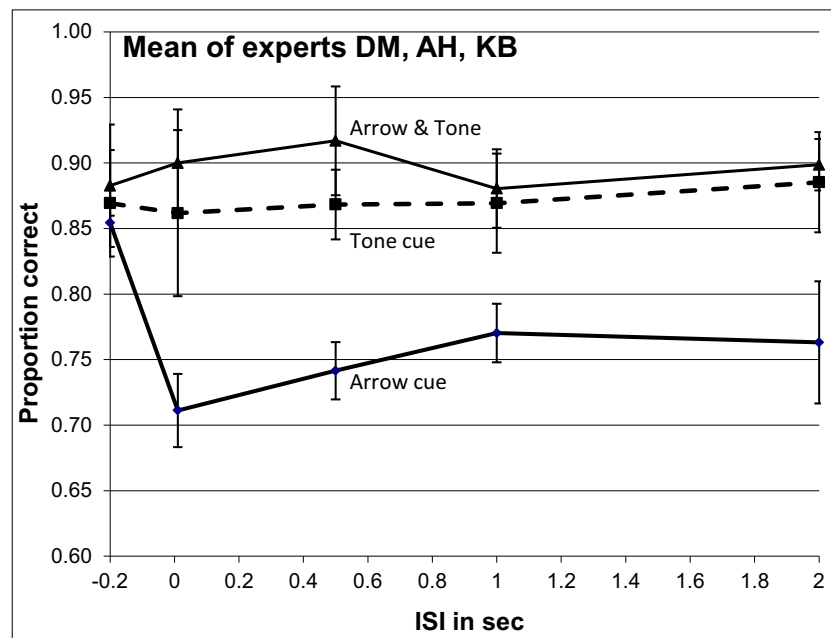


Fig. 5 Mean accuracies for the three experts in Experiment 3 plotted against ISI in seconds, with an arrow cue alone (bottom curve), with a tonal cue alone (middle curve), and with both cues presented together (top

curve). The dips revealed with the arrow cue were absent whenever the tonal cue was presented. Bars show ± 1 standard error across subjects.

recall at later SOAs after the icon had faded. We therefore ran Experiment 4, which repeated Experiment 1 while requiring backward counting to suppress verbalization. We adopted the method of Roediger, Knight, and Kantowitz (1977), who found that backward counting by threes “as quickly as possible” from a random start point between 501 and 699 increased errors to report whether a probe word had or had not appeared in a list of three words from 1% to 19%, and increased RTs from 649 to 1,636 ms, indicating a fairly severe load. In Experiment 4, a random number between 501 and 699 was read out 3 s before every trial, giving time for subjects to start counting down by threes, out loud, before presentation of the depth array; they continued counting down until after the arrow had been presented. Counting errors were ignored, since the aim was simply to prevent verbal rehearsal. Subjects averaged close to 1 s per spoken numeral and reported between four and six numerals on each trial. We anticipated that backward counting would have no effect on the dips, since the subjects in Experiments 1 and 2 had reported not verbalizing, but we did expect that the effort involved in counting would reduce accuracy overall. Eight naive subjects were run with a display duration of 200 ms, to ensure that vergence would not change during stimulation, and three further subjects whose numeral recalls were near chance were excused rather than being run at 800 ms. Otherwise, the procedure of Experiment 1 was employed, with subjects being run in the random-order condition.

Results and discussion

Figure 6 plots mean accuracies against SOA for all eight subjects in Experiment 4 with solid squares and black lines, along with mean accuracies for the eight subjects run with 200-ms displays in Experiment 1, plotted with black diamonds and a dotted line for comparison. Lighter symbols and dashed lines show mean accuracy for the best four and worst four subjects in Experiment 4. Clearly the added task of backward counting interfered overall, since the mean accuracy in Experiment 4 was below that in Experiment 1. Indeed, seven of the eight subjects in Experiment 4 reported that the counting task was distracting and sometimes led them to make mistakes. However, backward counting did not eliminate the dip, for either the better or the worse subjects, in accord with our expectations. We conclude that the dip and subsequent rise in accuracy at longer SOAs in the earlier experiments was not caused by retrieval from verbal memory.

Experiment 5: Iconic decay

The same three experts from Experiment 3 were also run with flat, Sperling-type arrays, as in Reeves and Lei (2014). The arrays comprised four rows of three letters each. The row to be reported was chosen at random on each trial and was cued by an arrow appearing to the side. All stimuli were portrayed with zero disparity. White letters were used with the same character sizes and luminances as the numerals in Experiments 1–4; the gray

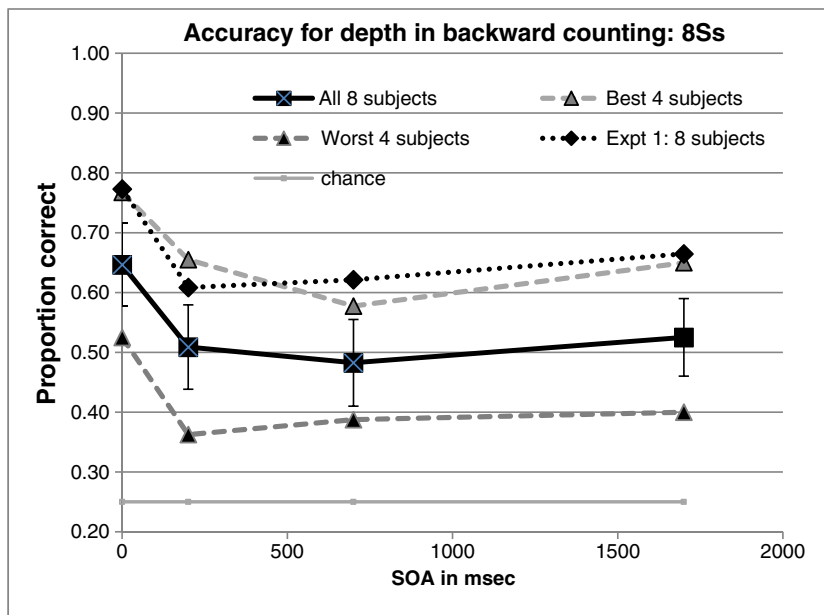


Fig. 6 Mean accuracies for reporting the cue numeral versus SOA, for subjects with 200-ms displays in Experiment 4 (heavy black lines connecting solid squares) and, for comparison, mean accuracies in Experiment 1 (dotted lines connecting black diamonds). Light dashed

lines connecting triangles show the means of the best and the worst four subjects in Experiment 4. The thin line at .25 indicates chance. Bars show $\pm 1 SE$, taken across subjects.

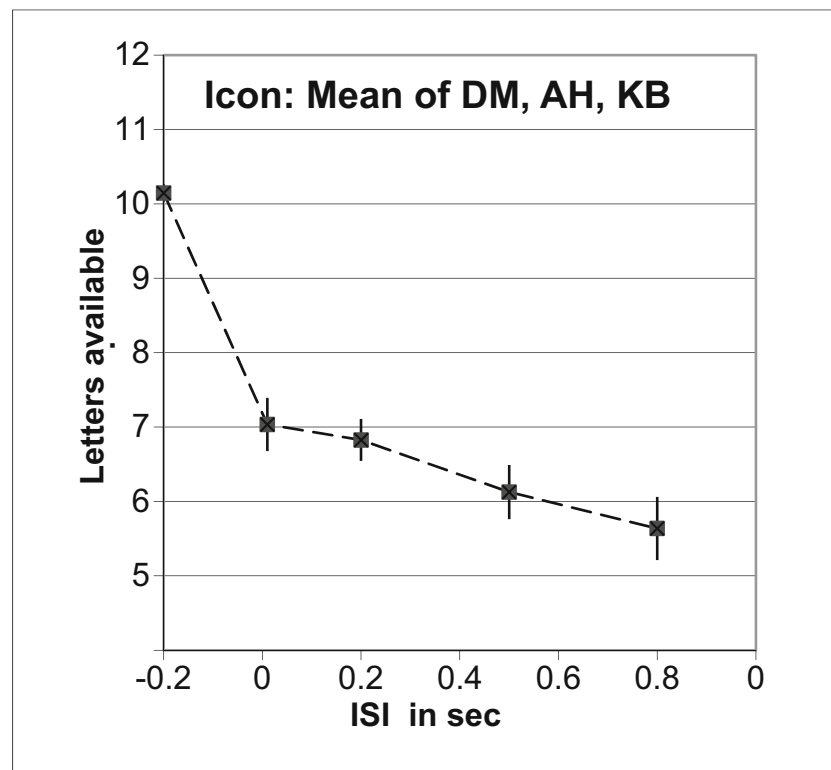


Fig. 7 Mean iconic decay for the three experts run in Experiment 5, plotted against ISI. Bars show ± 1 standard error for the most variable subject.

field was widened to accommodate the wider rows, but otherwise was unchanged. Sperling's (1960) ascending method was used, with the cue delay beginning at simultaneity and increasing progressively over trial blocks. On each trial, subjects called out the names of the three cued letters, and these were keyed into the computer by an experimenter who could not see the display.

Results are plotted in terms of Sperling's "information available" (the mean number of letters reported times the

number of rows) versus the cue delay (ISI). The means for the three experts are shown in Fig. 7, along with standard error bars for the most variable subject. Iconic decay was found for each expert: That is, as the ISI in seconds, denoted t , increased, accuracy decreased. The best-fitting exponential, $0.87e^{-0.00325t}$, decayed to 50% in 213 ms, a value used in fitting the model (below). Our finding of decay with flat displays, as expected and as had been found by Reeves and Lei (2014), makes it

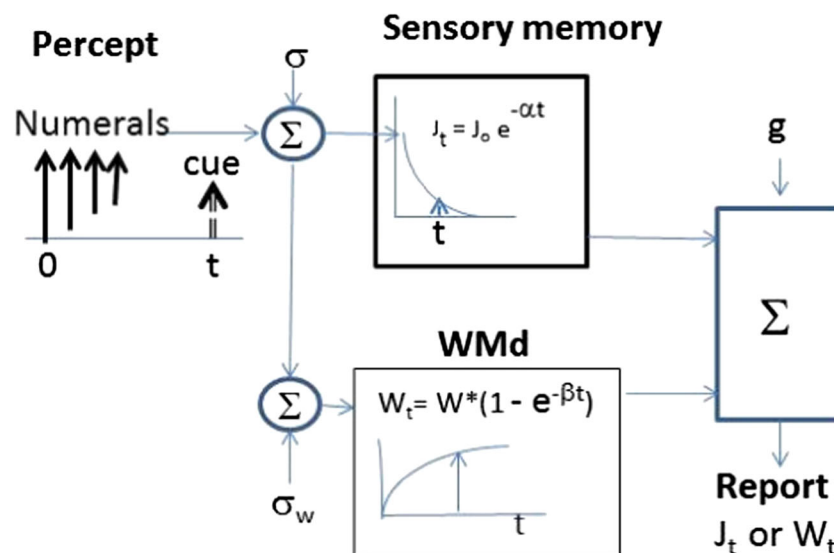


Fig. 8 Block diagram of our model. Percepts are stored in sensory memory, J_t , where they decay, and in working memory (VMd), W_t , where they accumulate; reports can come from either store (but not from both) or be random guesses (g).

Table 1 Model fits

	RMS (α/β free)	α (Exp. 5)	RMS	α (best)	RMS
Exp. 1	.044	.00325	.109	.0110	.054
Exp. 2	.030	.00325	.040	.0040	.040

highly unlikely that the dips found in Experiments 1–3 are artifacts of our equipment, displays, or subjects.

General discussion

Transfer model

The dips can be explained by assuming that subjects initially report from a decaying sensory store, but after some

period a depthful visual memory representation (VMd) is formed, and subjects can also report from this store. VMd is assumed to be visual because neither color–shape load nor backward counting by threes had much effect on the shape or timing of the dip. How could such visual memories help the subject report the numeral in the cued depth plane? Although VMd in principle may retain exact metric depth information, here we need postulate only that discrete “tags” are retained for each numeral in each depth plane, and that reports are accurate on trials in which the tag for the perceived depth plane of the cue is the same as that for the recalled depth plane of the cued numeral; errors arise when the numeral and cue tags fail to correspond due to noise. A model of this sort is illustrated in the block diagram in Fig. 8. Denoting by P_t the predicted accuracy for reporting a cued item at cue delay t , where $t = 0$ at simultaneity,

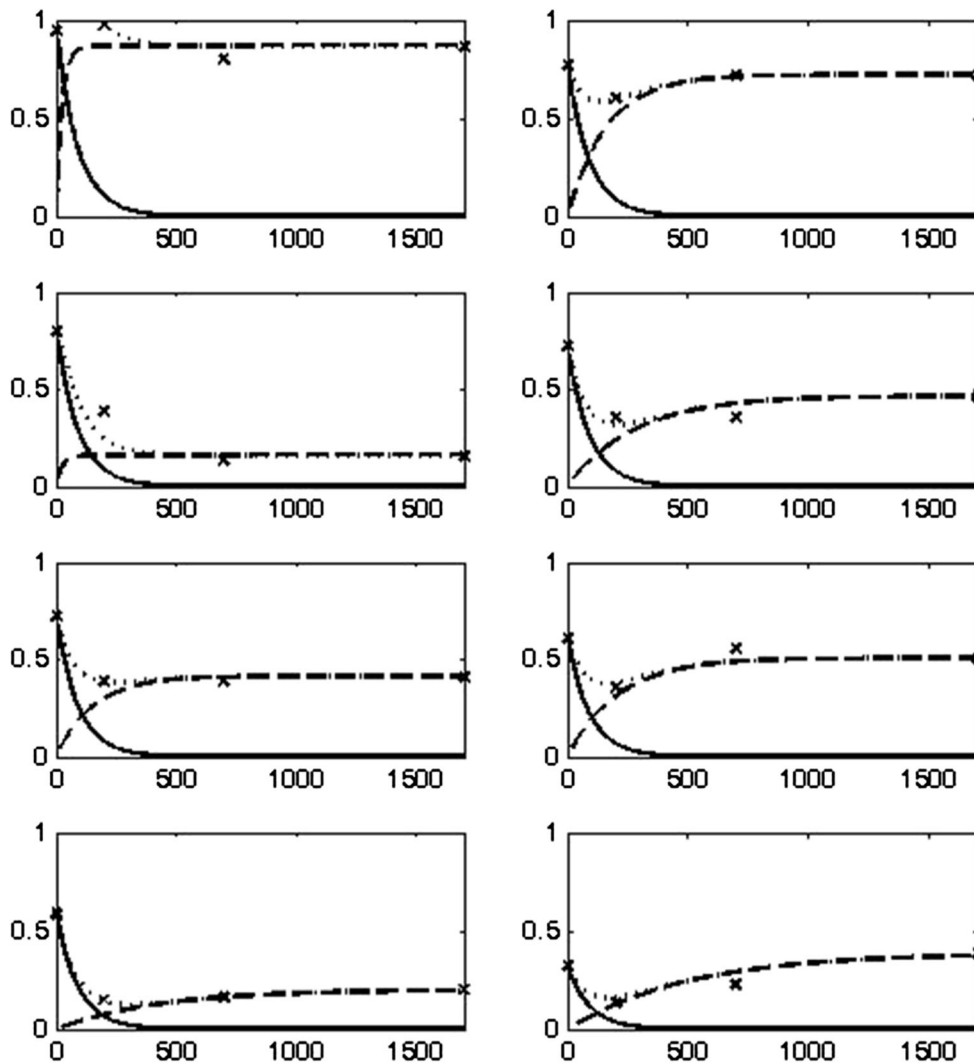


Fig. 9 Model fits for the eight naive subjects run with 200-ms displays in Experiment 1, plotted against the ISI in milliseconds from numeral array offset (time 0).

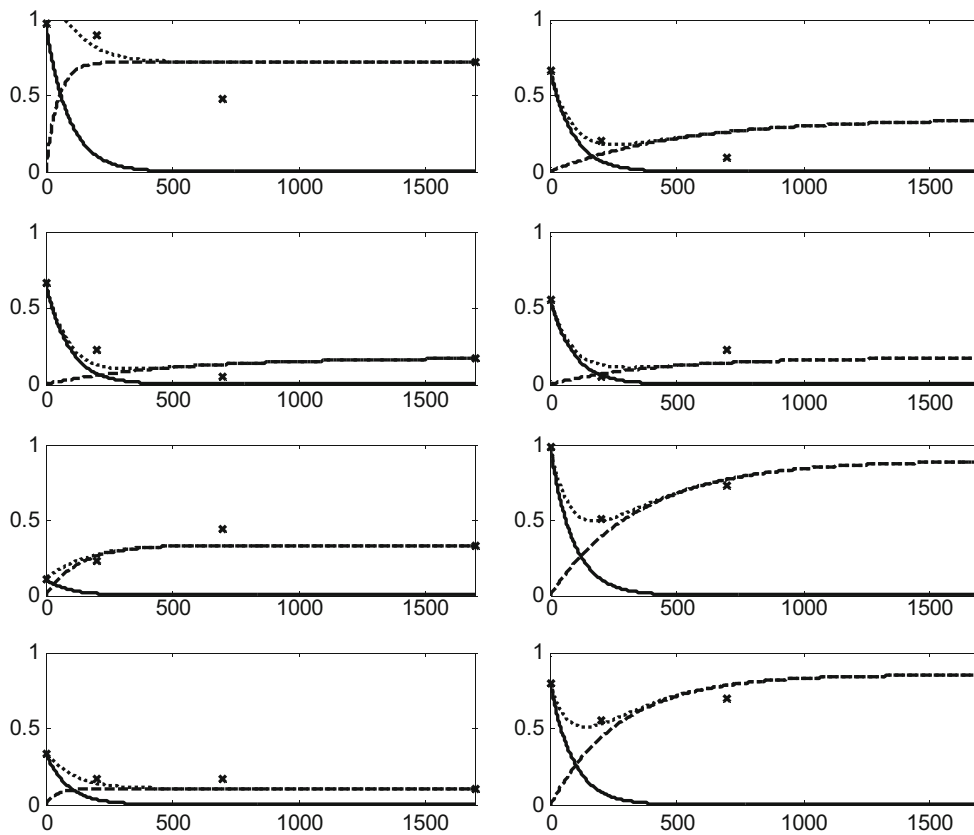


Fig. 10 Model fits for the eight naive subjects run with 800-ms displays in Experiment 1, plotted as in Fig. 9.

$$P_t = J_t + W_t + g \tag{1}$$

Here, J_t and W_t denote the proportions of correct numeral–cue tags in the sensory store and in VMd, respectively, at time t , and g denotes successful guesses. The proportions J_t , W_t , and g add because we assume that tags can be picked out from either store (but not both) and that numerals are guessed when both memory stores fail. The raw proportions, P , were corrected for such guessing by calculating $(P - 1/n)/(1 - 1/n)$ for $n = 4$ or 5 items (as in Loftus, Duncan, & Gehrig, 1992), with $g = 1/n$, and the model was fit to the corrected data.

For a sensory store that decays exponentially at rate α from an initial level J_0 ,

$$J_t = J_0 e^{-\alpha t} \tag{2}$$

At time $t = 0$, the cue is simultaneous with the target and no transfer to VMd has yet occurred, so $W_0 = 0$, and therefore P_0 , the probability of a correct response at SOA = 0 corrected for guessing, must equal J_0 . That $P_0 < 100\%$ is due to tagging noise, denoted σ in Fig. 8. We assume that σ decreases as the stimulus duration increases, but that once the stimulus turns off, the decay rate, α , is a property only of sensory memory and is independent of the stimulus duration: Loftus, Duncan,

Table 2 Model parameters in Experiment 1, for Subjects 1–8 with 200-ms displays and Subjects 9–16 with 800-ms displays

Exp. 1	(alpha = .0110)	
	Beta	RMS
200 ms		
1	.0418	.03356
2	.0062	.00471
3	.0418	.07034
4	.0036	.04025
5	.0066	.01163
6	.0048	.03680
7	.0024	.00309
8	.0020	.03562
800 ms		
9	.0216	.12043
10	.0020	.07996
11	.0020	.04170
12	.0024	.07353
13	.0074	.05504
14	.0036	.05298
15	.0222	.03319
16	.0036	.04546

and Gehrig (1992) have provided evidence for both assumptions.³

Transfer to VMd is accomplished by an incremental growth process, with rate β :

$$W_t = W'(1 - e^{-\beta t}). \tag{3}$$

At $t = 0$ transfer has not yet started, so $W_0 = 0$. Transfer accumulates toward an asymptote $W' = P_{1500}$ —that is, the report accuracy when the cue is so late ($t = 1,500$ ms) that the sensory information has entirely decayed. W' is less than 100% due to a second source of noise, σ_w in Fig. 8, that degrades VMd.

This model represents information as decaying in sensory memory while growing in VMd, as if both forms of information begin at time 0 but then pursue independent paths. Alternatively, perceptual input might be encoded first into sensory memory and then transferred: If so, the term $-\beta t$ in Eq. 3 becomes $-(\beta/\alpha)\log_e(J_t/J_0)$, generating a mathematically equivalent model.

Model constraints Since J_0 , W' , α , and β are the four parameters for each subject, equal to the number of data points, fitting the model required that some parameters be constrained. Parameter α was constrained by assuming the same sensory decay for everyone, and was either treated as a single free parameter or set to .00325 (i.e., decay to half strength in 213 ms), the decay found in Experiment 5 for the experts. Since $J_0 = P_0$ and $W' = P_{1500}$ were estimated directly from the individual data, only the growth rate (β) needed to be fit for each subject when α was constrained.

In Experiment 1, with β free, setting $\alpha = .00325$ led to poor least-squares fits to the data for many of the naive subjects, the average root-mean squared (RMS) error being 10.9% (Table 1). However, the best-fitting α for all naive subjects, $\alpha = .011$, reduced the RMS error to 5.4%, close to the 4.4% found with α and β both free to vary. Fits are therefore plotted in Fig. 9 for the 200-ms display subjects, and in Fig. 10 for the 800-ms subjects, with $\alpha = .011$ (see the individual parameters in Table 2). The time axis plots the ISIs, shifted so that time 0 is locked to stimulus offset. The smooth lines in each plot trace out the best-fitting curves for Eq. 1 (dipping), Eq. 2 (declining), and Eq. 3 (rising), calculated every 5 ms. The data, corrected for guessing and symbolized by x's, are close to the dipping curves representing P_t from Eq. 1. The two naive subjects whose data fit poorly with $\alpha = .011$ (see the

Table 3 Model parameters in Experiment 2

Exp. 2	(alpha = .00325)	
	Beta	RMS
Load		
WZ	.00220	.01696
IN	.00320	.01449
FP	.00260	.04351
EN	.00280	.04861
SA	.00320	.02274
MT	.00380	.05517
No Load		
WZ	.00240	.07806
IN	.00380	.00972
FP	.00280	.02526
EN	.00340	.03713
SA	.00300	.03269
MT	.00380	.04342

top panels of Fig. 10) were fit well with $\alpha = .00325$ (not shown).

Data (as always, corrected for guessing) for the six experts in Experiment 2 were fit well with β free and $\alpha = .00325$ (from Exp. 5; see Table 3). The results are shown in Fig. 11 without load and in Fig. 12 with load. Note that the y-axis is plotted from .4 to 1.0 to save space; below .4 the icon and VMd curves go smoothly to the x-axes as in Fig. 9. The average RMS error was 4.0%, close to the 3.0% found for the experts with both α and β free. Thus, the model can describe the data with only one free parameter per subject. The different α values for experts and naive subjects were not predicted, but it is plausible that training helped the experts use the decaying information in the sensory store to ever-greater effect, extending its apparent life. Indeed, unlike the experts, naive undergraduates in my laboratory courses often fail to use the cue with ISIs over 100 ms when shown Sperling-type flat letter arrays.

Critically, subjects' showing small dips can be accounted for by increasing the transfer rate (β) so that VMd quickly substitutes for the decaying icon (as in the top-left panel of Fig. 9), whereas subjects with large dips are accounted for by decreasing the transfer rate and so revealing the dip while VMd is still developing (as in the top-right panel Fig. 9). Thus, the model can handle both monotonic and U-shaped curves.⁴

³ A single exponential stage of iconic decay, as assumed here, fit their data with a root-mean squared error of only 6.6 ms. This was reduced significantly to 5.5 ms by incorporating an additional stage of exponential decay, a second-order effect that we will ignore.

⁴ Normally distributed errors in the assignments of tags to depths can also account for the iconic decay seen in Experiment 5, since the best-fit exponential agrees well ($r^2 = .98$), over the limited time domain tested, with a normal distribution whose σ increases linearly with time.

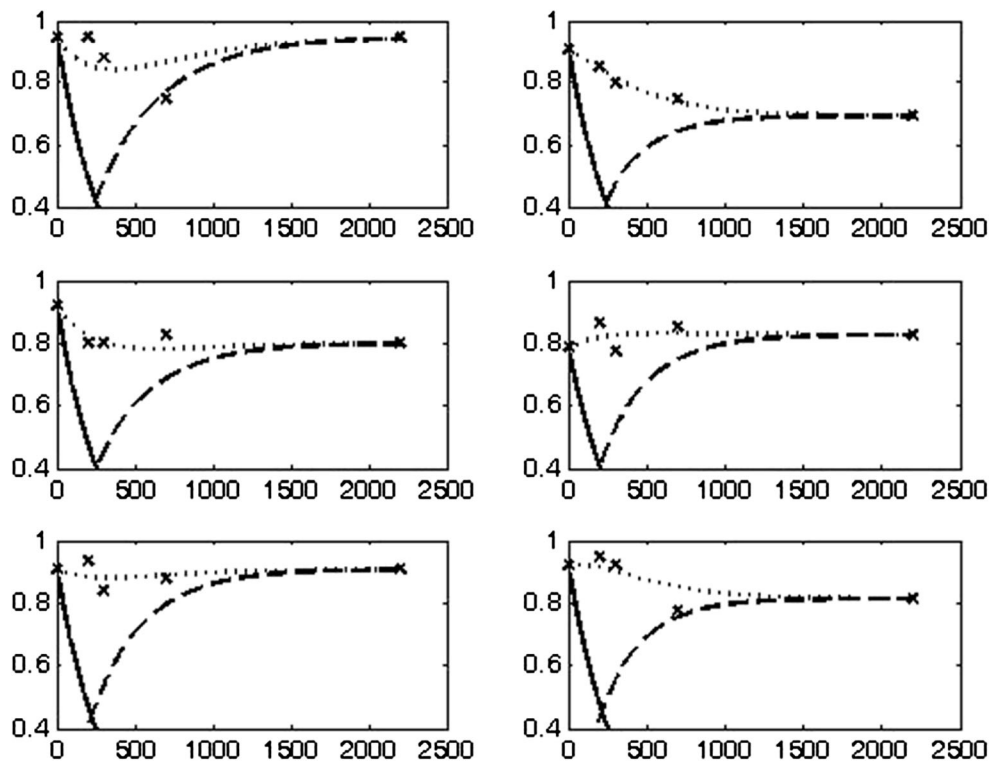


Fig. 11 Model fits for the six experts run with 200-ms displays with no load, plotted against the ISI from numeral array offset. Proportions correct are shown from .4 to 1.0 to save space, and so, unlike in Figs. 9 and 10, the entire curves are not shown.

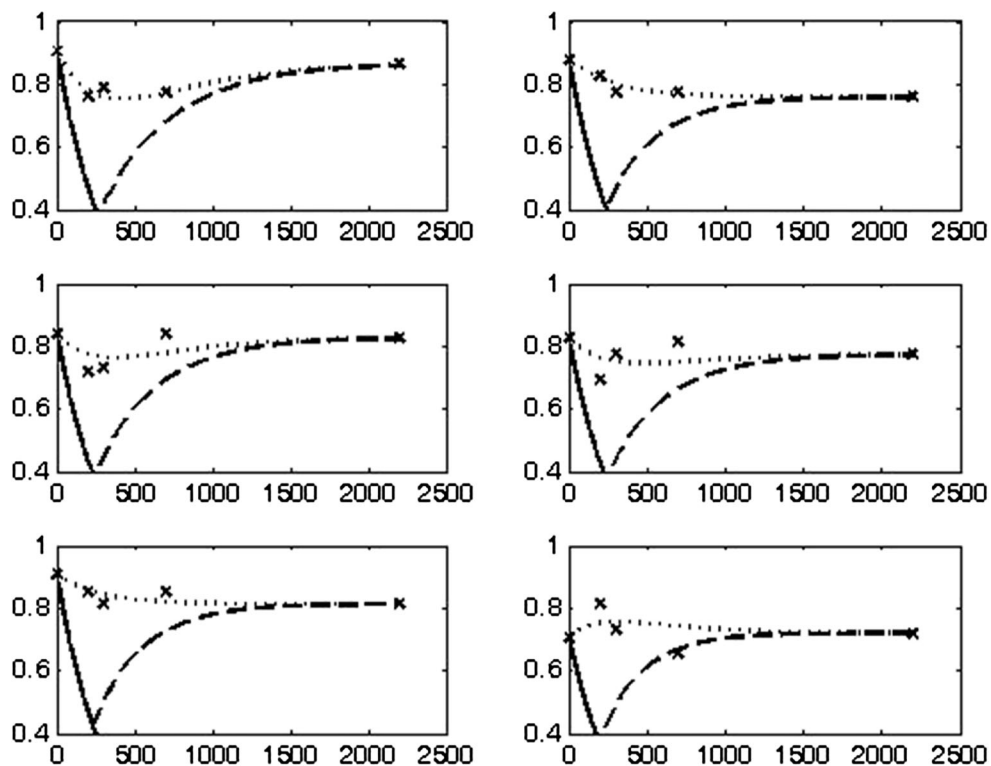


Fig. 12 Model fits for the six experts run with 200-ms displays with load, plotted as in Fig. 11.

Conclusions

The results can be explained if visual information is transferred to a visual memory for depth, termed VMd here, that is functionally distinct from the WM thought to retain object features (Luck & Vogel, 1997; Zhang & Luck, 2008). With a visual cue to location in depth, the arrow, transfer rates vary widely across individuals, so the capacity of VMd, as indicated by W' , varies from almost nothing up to a maximum determined by the quality of the initial encoding. For those individuals with sufficient W' , a slow transfer to VMd generated a dip about 0.5 s after presentation of the array. Using a tonal rather than a visual cue eliminated both decay and dip, suggesting that transfer to VMd can be rapid enough to compensate for sensory decay when the cue is in a different modality, possibly because the perceived depth of the cue did not have to be compared to the memory of the depths of the numerals when the tone was employed as a cue.

Some caveats follow: (1) Decay and transfer were not manipulated independently within subjects; the model fitting was done solely across subjects. (2) Our assumption that the second memory store is an integral of the input, though simple, may not generalize. (3) Assuming an architecture in which “where” information (layout) is retained in VMd separately from the “what” information (objects and features) in WM, prior to being coordinated by attention, is consistent with the theory of Grossberg, Mingolla, and Fazl (2009), but perhaps both are maintained in the same WM store, with only features having capacity limits; this would change the interpretation of the model, though not the equations. (4) The explanation of the cue modality effect in terms of speed of transfer is post hoc and needs to be tested in its own right. (5) Model testing was restricted to the present data—that is, recalls of the locations of numerals in depth—and recall of other forms of depth information, such as rich, continuously varying scene data, will need to be tested in future research.

Despite these caveats, given the magnitudes of the dips that we found, one wonders why dips in the short-term storage functions for other forms of visual information, such as features and objects, have not been noted before (except by Sperling, 1960, Fig. 6). One possibility is that “depth is special,” in that distinct stores, WM and VMd, exist, perhaps because saccades, which are typically <6 deg (Hwang, Wang, & Pomplun, 2011), alter the objects being viewed much more often than they alter the layout, so that only layout can benefit from a slow integration. If so, trans-saccadic integration (Wijdenes, Marshall, & Bays, 2015) should differ for objects and layouts. Alternatively, perhaps both WM and VMd operate in the manner modeled for VMd, in which case the model

structure implies that the transfer of features either must be very slow or must asymptote to W' at a low enough level that a dip is not normally seen.

References

- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by information load and by number of objects. *Psychological Science*, *15*, 106–111. doi:10.1111/j.0963-7214.2004.01502006.x
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.
- Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *Journal of Physiology*, 599–820.
- Bouma, H. (1972). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research*, *13*, 767–782.
- Grossberg, S., Mingolla, E., & Fazl, A. (2009). View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. *Cognitive Psychology*, *58*, 1–48.
- Haddad, G. M., & Steinman, R. M. (1973). The smallest voluntary saccade: Implications for fixation. *Vision Research*, *13*, 1075–1086.
- Harwerth, R. S., Fredenburg, P. M., & Smith, E. L., III. (2003). Temporal integration for stereoscopic vision. *Vision Research*, *43*, 505–518.
- Hwang, A. D., Wang, H.-C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, *51*, 1192–1205. doi:10.1016/j.visres.2011.03.010
- Lefton, L. A. (1973). Metacontrast: A review. *Psychonomic Society Monograph Supply*, *4*, 245–255.
- Loftus, G. R., Duncan, J., & Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 530–549. doi:10.1037/0096-1523.18.2.530
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. doi:10.1038/36846
- Mewhort, D. J., Campbell, A. J., Marchetti, F. M., & Campbell, J. I. (1981). Identification, localization, and “iconic memory”: An evaluation of the bar-probe task. *Memory & Cognition*, *9*, 50–67.
- Neisser, U. (2014). *Cognitive psychology*. Hove, UK: Psychology Press. (Original work published 1967)
- Reeves, A., & Lei, Q. (2014). Is Visual Short-term memory depthful? *Vision Research*, *96*, 102–112.
- Roediger, H. L., III, Knight, J. L., Jr., & Kantowitz, B. H. (1977). Inferring decay in short-term memory: The issue of capacity. *Memory & Cognition*, *5*, 167–176. doi:10.3758/BF03197359
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11, Whole No. 498), 1–29.
- Sperling, G. (1969). A model for visual information tasks. In R. N. Haber (Ed.), *Information-processing approaches to visual perception* (pp. 18–31). New York: Holt, Rinehart & Winston.
- Tijus, C. A., & Reeves, A. (2004). Rapid iconic erasure without masking. *Spatial Vision*, *17*, 483–495.
- Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, *32*, 1349–1357.
- Weisstein, N. (1971). W-shaped and U-shaped functions obtained for monoptic and dichoptic disk-disk masking. *Perception & Psychophysics*, *9*, 275–278.

- Wijdenes, L. O., Marshall, L., & Bays, P. M. (2015). Evidence for optimal integration of visual feature representations across saccades. *Journal of Neuroscience*, *35*, 10146–10153.
- Wilschut, A., Theeuwes, J., & Olivers, C. N. L. (2013). Early perceptual interactions shape the time course of cueing. *Acta Psychologica*, *144*, 40–50. doi:10.1016/j.actpsy.2013.04.020
- Xu, Y., & Nakayama, K. (2007). Visual short-term memory benefit for objects on different 3-D surfaces. *Journal of Experimental Psychology: General*, *136*, 653–662.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235. doi:10.1038/nature06860