Published in final edited form as:

Psychol Sci. 2012 July 1; 23(7): 698–703. doi:10.1177/0956797612443968.

Saved by a log: How do humans perform hybrid visual and memory search?

Jeremy M Wolfe

Brigham & Women's Hospital and Harvard Medical School

Abstract

Could you find one of your 1000 Facebook friends in a crowd of 100? Even at rates of 20–40 comparisons/sec, determining that no friends were in the crowd would take ~40–50 minutes if memory and visual search interacted linearly. Our observers memorized pictures of 1–100 targets and then searched for any target in visual displays of 1–16 objects. Response times varied linearly with visual set size but 1–16 accurately predicted response times for different observers holding 100 objects in memory. The results could support a binary representation of visual objects in memory and are relevant to applied searches in which experts look for any of many items of interest (e.g. medical or airport screening tasks).

How do we search for one of many different target items in a visual world filled with many other objects? While there are large literatures on visual search (Wolfe, 2010) and memory search (Van Zandt & Townsend, 1993), there is much less work examining the interaction of these two search processes. The bulk of the work that does exist deals with search for members of small memory sets of letters and/or digits in relatively small visual displays (Schneider & Shiffrin, 1977) (Nickerson, 1966) (Burrows & Murdock, 1969) (Briggs & Blaha, 1969). For visual and memory set sizes in the range of 1–4 characters, search times are roughly linear with the product of visual and memory set sizes (Schneider & Shiffrin, 1977). More recent work paints a complicated picture of the effects of working memory loads on visual search (Balani, Soto, & Humphreys, 2010; Downing & Dodds, 2004; Han & Kim, 2004; Woodman, Vogel, & Luck, 2001) (Olivers, Peters, Houtkamp, & Roelfsema, 2011) but as a general rule these are studies in which working memory is pitted against search. We do not know how memory and vision work together when stimuli are drawn from the wide set of visual objects beyond alphanumeric characters nor what happens when the numbers of task relevant items held in memory gets well beyond the limits of working memory. For example, how do we search a crowded pantry for the 8 ingredients needed for dinner? How does an intelligence analyst search satellite imagery for any of many targets of interest? At the usual rates proposed for visual search (Wolfe, 1998) and memory search (Sternberg, 1966), a linear dependence on the product of visual and memory set sizes would become prohibitive once those set sizes become large. Here we show that the “solution” to this problem is logarithmic search through the memory set (For earlier examples of reaction times (RTs) increasing with the log of the number of alternatives see, for example, Burrows (D. Burrows & Okada, 1975) or the Hick-Hyman law (Hick, 1952; Hyman, 1953) in motor response tasks). Some work in the `60s and `70s addressed combined visual and memory search with as many as 8 well-learned alphanumeric stimuli as the memory set (Neisser, 1974). No one has addressed the problem of visual search for arbitrary objects when the number of possible targets is far outside the limits of working memory.

Experiment 1

Methods

In Experiment One, ten observers (Os) searched visual displays of 1, 2, 4, 8, or 16 photographs of objects for any of 1, 2, 4, 8 or 16 items held in memory. In a block of trials, the Os first memorized a set of 1–16 items (Fig 1a) to an average accuracy of 97%. In order to proceed, the procedure required Os to be above 80% correct on two successive tests of their recognition memory for the memory set. In practice, memory for visual objects is much better than memory for letters and digits (Brady, Konkle, Alvarez, & Oliva 2008; Konkle, Brady, Alvarez, & Oliva, 2008), so recognizing up to 16 distinctive objects as members of a memory set is quite trivial; hence the 97% accuracy. Memorization and testing of that memory took no more than five minutes with longer time required for the larger memory set sizes. The minimum number of memory test blocks was two. In practice, the average number of blocks required rose from 2.7 at memory set size 1 to 3.4 at memory set size 16. The modal number of required blocks was two at all memory set sizes.

With the memory set encoded with good accuracy, Os proceeded to 500 trials of visual search through random arrays of objects (Fig 1b). One and only one of the remembered targets was present on 50% of trials. No targets were present on the other 50%. Trials were randomly divided among the 5 visual set sizes (1, 2, 4, 8, 16). Os responded with a target-present, target-absent key press under instructions to be as quick and accurate as possible.

Results

Figure 1c shows target-present response times (RTs) as functions of the visual set size for each of the five memory set sizes. It is clear that the effects of visual set size are quite linear, as is typical in search for one object among many (Vickery, King, & Jiang, 2005). Larger memory set sizes produced progressively steeper, but still linear visual set size functions. Target absent data are comparable with slopes about 2.5X the target present slopes. In contrast, as shown in Fig 1d, RTs are decidedly not linear as a function of memory set size (Note: The data in 1c and 1d are the same RTs plotted either as a function of the visual set size or the memory set size). As shown in Figures 1e–f, RTs appear to be a linear function of the logarithm of the memory set. Regression coefficients can be calculated for $RT \times \text{Set size}$ and $RT \times \text{Log}_2(\text{set size})$ functions. The fits are better for the $\log_2$ function ($t(9) > 2.3$, $p < 0.05$ for visual set sizes of 1, 4, 8, & 16; $T(9) = 2.2$, $p = 0.058$ for visual set size 2).

Experiment 2: Memorizing 100 objects

The ability to remember large numbers of objects (Brady, et al., 2008; Konkle, et al., 2008) allows us to push this task far beyond anything attempted with alphanumeric characters. In Experiment 2, a new group of 10 observers committed 100 objects to memory. As in Experiment One, they were required to pass two consecutive memory tests at no less than 80% correct. This portion of the experiment took approximately 10–15 minutes. One observer needed three blocks to meet the memory criterion. All others succeeded in the minimum of two. The mean accuracy on the last memory block was 93%. After reaching criterion, observers proceeded to perform 300 visual searches for any of the 100 objects in displays of 1, 2, 4, 8, or 16 items. Over 2000 objects were used as distractors. These were sampled without replacement.

First of all, the task was surprisingly easy. Miss error rates rose from 2% at visual set size 1 to 19% at set size 16. False alarms rose from 1% to 7%. D' falls from 4.4 to 2.4. With a memory set size of 100, RT remained a linear function of visual set size, albeit with very steep slopes of 139 msec per visual item for target present trials; 314 msec/item for target absent.

Because we are measuring just one memory set size in Experiment 2, it is not possible to describe the shape of the RT $\times$ memory set size function within the experiment. However, we can use the data from Experiment 1 to predict the RT for a memory set size of 100. The average RTs for Experiment 2 are shown as the solid data points at memory set size 100 in Figures 1e&f. The linear regression of RT on $\log_2(\text{memory set size})$ is shown for the data of Experiment One as solid lines. The extrapolation to set size 100 is shown as the dashed line. It is clear that the logarithmic prediction is a remarkably good fit to the data. The average errors in prediction for target present trials are -52, 1, 60, 68, & -74 msec for visual set sizes 1,2,4,8, & 16 respectively (within 1–5% of the average RT). As a somewhat ludicrous comparison, the data of Experiment Two fall short of the predictions of a linear model by -1156, -1814, -2505, -4162 & -6844 msec (about 200% of the average RT). As can be seen in Fig 1f, the predictions of the logarithmic model are comparably good for the absent trials.

Experiment 3: Localizing the target

Returning to Experiment 1, one could argue that the departure from linearity represents a speed-accuracy tradeoff (Shiffrin, 1988). Errors rose with both visual and memory set sizes (Both $F(1,9) > 9.4$, both $p < 0.0001$). Errors rose more markedly with memory set, reaching a level of miss errors of 17% when visual and memory set sizes were both 16. False alarm rates were low (<3% in all cells). Larger errors could reflect earlier termination of search. This could depress RTs at the larger set sizes, producing a less interesting compression of the RT $\times$ set size function. As a test of this hypothesis, we repeated Experiment 1 using a localization response, rather than a present/absent keypress. Ten new Os were tested with memory set sizes of 1,2, 4, 8, & 16 and visual set sizes of 2, 4, 8, & 16. Again, Os needed to pass the memory test twice as in Experiment One. The average number of blocks required rose from 2.2 at memory set size 1 to 4.2 at memory set size 16. The modal number of required blocks was two at all memory set sizes. Final accuracy was over 97%. This portion of the experiment took about 5 minutes (somewhat longer for the observer who required 12 tries to get 16 items into memory!) After passing the memory test, observers were tested for 300 trials with one target present on each trial. Response was a mouseclick on the target. This reduced error rates to 2% or less in all cells except for an 8% error rate when visual and memory set sizes were 16.

Figure 2 shows the results of Experiment 3, here with both memory and visual set sizes plotted on linear x-axes. The functions relating RT to the visual set size remain linear. The functions relating RT to memory set size are clearly non-linear. Just as we predicted the set size 100 data with the data from set sizes 1–16, in this experiment, we can predict RTs for set size 16 from set sizes 1–8. Open circles show the logarithmic predictions. We cannot reject this model ($F(1,9) = 0.21$, $p = 0.65$, $p_{\text{es}} = 0.003$). Note that the actual values are slightly lower than the predicted values, perhaps reflecting a small speed-accuracy trade-off. In contrast, the asterisk symbols show the less successful predictions for a linear extrapolation to set size 16 from set sizes 1–8. This model can be rejected ($F(1,9) = 26.0$, $p = 0.0004$, $p_{\text{es}} = 0.34$).

General Discussion

Thus, visual searches for multiple targets held in memory are accomplished in a reasonable amount of time because memory search times increase with the $\log_2$ of the memory set size. The hypothetical search for any of 1000 friends in a picture of a 100 people would take over 40 minutes if memory search and visual search were both linear and each step in the search took 25 msec. With logarithmic memory search, that time drops to a more plausible 25 seconds. What does it mean for RTs to vary with $\log_2$ of the set size? This is the pattern that would be seen if half the items could be eliminated on the first step, another half on the next

step, and so on. This could occur if objects were represented in the equivalent of a binary code in memory. When the observer attended to a visual object, its first “bit” of information could be compared to the first bits of items in memory. On average, half the items would match and the other half could be eliminated. Another half could be rejected by examining the second bit and so on until the item was uniquely identified as belonging to the memory set of targets or rejected as a distractor. This is essentially the same as the information theoretic account traditionally offered for the Hick-Hyman law (Hick, 1952; Hyman, 1953). Of course, the code might not be binary. The underlying function might be $\log_3$ or $\log N$. Nor is it required that every item be coded with the same number of bits (e.g. Huffman coding (Huffman, 1952)), but this result illustrates a solution to the problem of joint visual and memory searches and raises interesting questions about the underlying neural representation of remembered items.

Real world searches are likely to be more complex and more efficient than those described here because objects are represented at more than one level (Rosch, 1973). Thus, a search for a robin, a wren, a blue jay, or a penguin in a set of objects is likely to be a search for members of the category, “bird” followed by a memory search for specific types of bird only through the subset of objects identified as birds. Any of these search tasks that feature more than one possible target will require careful coordination of visual and memory search because evidence suggests that, at any one moment, only one target “template” is active (Houtkamp & Roelfsema, 2009; Olivers, et al., 2011).

In sum, when you search that photo for any of your many Facebook friends, your attention will be guided to humans and away from other objects (Wolfe, 1994). Search time will be a linear function of the number of humans in the visual scene and a logarithmic function of the number of friends held in memory.

References

- Balani AB, Soto D, Humphreys GW. Working memory and target-related distractor effects on visual search. *Memory & Cognition*. 2010; 38(8):1058–1076. [PubMed: 21156870]
- Brady TF, Konkle T, Alvarez GA, Oliva A. Visual long-term memory has a massive storage capacity for object details. *Proc Natl Acad Sci U S A*. 2008; 105(38):14325–14329. [PubMed: 18787113]
- Briggs GE, Blaha J. Memory retrieval and central comparison times in information processing. *Journal of Experimental Psychology*. 1969; 70(3, pt1):395–402.
- Burrows D, Murdock BB Jr. Effects of extended practice on high-speed scanning. *Journal of Experimental Psychology*. 1969; 82(2):231–237. doi: 10.1037/h0028145.
- Burrows D, Okada R. Memory Retrieval from Long and Short Lists. *Science*. 1975; 188(4192):1031–1033. [PubMed: 17759685]
- Downing P, Dodds C. Competition in visual working memory for control of search. *Visual Cognition*. 2004; 11(6):689–703.
- Han S-H, Kim M-S. Visual search does not remain efficient when executive working memory is working. *Psychological Science*. 2004; 15(9):623–628. [PubMed: 15327634]
- Hick WE. On the rate of gain of information. *Quart. J. Experimental Psychology*. 1952; 4:11–26.
- Houtkamp R, Roelfsema PR. Matching of visual input to only one item at any one time. [Article]. *Psychological Research*. 2009; 73(3):317–326. doi: 10.1007/s00426-008-0157-3. [PubMed: 18665392]
- Huffman DA. A Method for the Construction of Minimum-Redundancy Codes. *Proceedings of the I.R.E.* 1952; September:1098–1102.
- Hyman R. Stimulus information as a determinant of reaction time. *J Exp Psychol*. 1953; 45(3):188–196. [PubMed: 13052851]
- Konkle T, Brady T, Alvarez G, Oliva A. Remembering thousands of objects with high fidelity. *Journal of Vision*. 2008; 8(6):694–694.

- Neisser U. Practiced card sorting for multiple targets. *Mem Cognit.* 1974; 2(4):781–785.
- Nickerson RS. Response times with a memory-dependent decision task. *Journal of Experimental Psychology.* 1966; 72(5):761–769. doi: 10.1037/h0023788. [PubMed: 5972018]
- Olivers CN, Peters J, Houtkamp R, Roelfsema PR. Different states in visual working memory: when it guides attention and when it does not. *Trends Cogn Sci.* 2011; 15(7):327–334. doi: S1364-6613(11)00085-4 [pii] 10.1016/j.tics.2011.05.004. [PubMed: 21665518]
- Rosch EH. Natural categories. *Cognitive Psychology.* 1973; 4:328–350.
- Schneider W, Shiffrin RM. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol. Rev.* 1977; 84:1–66.
- Shiffrin, RM. Attention. In: Atkinson, R.; Herrnstein, RJ.; Lindzey, G.; Luce, RD., editors. *Steven's Handbook of Experimental Psychology.* 2nd ed.. Vol. Vol. 2. John Wiley & Sons; New York: 1988. p. 739-812.
- Sternberg S. High-speed scanning in human memory. *Science.* 1966; 153:652–654. [PubMed: 5939936]
- Van Zandt T, Townsend JT. Self-terminating versus exhaustive processes in rapid visual and memory search: An evaluative review. *Perception and Psychophysics.* 1993; 53(5):563–580. [PubMed: 8332425]
- Vickery TJ, King L-W, Jiang Y. Setting up the target template in visual search. *J. of Vision.* 2005; 5(1):81–92.
- Wolfe JM. Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review.* 1994; 1(2):202–238. [PubMed: 24203471]
- Wolfe JM. What do 1,000,000 trials tell us about visual search? *Psychological Science.* 1998; 9(1):33–39.
- Wolfe JM. Visual search. *Current Biology.* 2010; 20(8):R346–R349. doi: DOI: 10.1016/j.cub.2010.02.016. [PubMed: 21749949]
- Woodman GF, Vogel EK, Luck SJ. Visual search remains efficient when visual working memory is full. *Psychological Science.* 2001; 12(3):219–224. [PubMed: 11437304]

Figure 1.

(a) Observers memorized a set of 1–16 (Exp 1) or 100 (Exp 2) objects and were tested to confirm that these were retained (which they were with >90% accuracy). (b) Os then performed visual search for any memorized target among 1–16 other objects. (c) RTs increased linearly as a function of the visual set size. The slope of the RT $\times$ visual set size function is steeper when larger numbers of targets are held in memory. (d) In contrast, RTs increased as a logarithmic function of memory set size. (e–f) Exp 2 extends the result to memory set size of 100. RT rises linearly with the log of the memory set size (note log x-axis). Solid lines are best-fit regression lines for RT $\times$ $\log_2$ (memory set size) for the Exp 1 data. Dashed lines are the extrapolation of those lines to a memory set of 100. Solid data points at set size 100 show the results for Exp 2 based on data from the entirely different set of observers in Exp 1. Clearly, the extrapolation of Exp 1 results predicts Exp 2 results. In all graphs, solid symbols are average data for 10 observers $\pm$ 1 s.e.m.

Figure 2.

Average RT data for Experiment 3 in which targets were present on each trial and observers localized the target with a mouseclick. (a) RT rises linearly with visual set size. (b) RT rises with the logarithm of memory set size. Open circles show the predicted RT for set size 16 based on a logarithmic relationship. Asterisk symbols show the unsuccessful linear predictions.