

Psychological Science

<http://pss.sagepub.com/>

Saved by a Log: How Do Humans Perform Hybrid Visual and Memory Search?

Jeremy M. Wolfe

Psychological Science 2012 23: 698 originally published online 22 May 2012

DOI: 10.1177/0956797612443968

The online version of this article can be found at:

<http://pss.sagepub.com/content/23/7/698>

Published by:



<http://www.sagepublications.com>

On behalf of:



[Association for Psychological Science](http://www.sagepublications.com)

Additional services and information for *Psychological Science* can be found at:

Email Alerts: <http://pss.sagepub.com/cgi/alerts>

Subscriptions: <http://pss.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - Jul 18, 2012

[OnlineFirst Version of Record](#) - May 22, 2012

[What is This?](#)

Saved by a Log: How Do Humans Perform Hybrid Visual and Memory Search?

Jeremy M. Wolfe

Brigham & Women's Hospital, Cambridge, Massachusetts, and Departments of Ophthalmology and Radiology, Harvard Medical School

Psychological Science
 23(7) 698–703
 © The Author(s) 2012
 Reprints and permission:
sagepub.com/journalsPermissions.nav
 DOI: 10.1177/0956797612443968
<http://pss.sagepub.com>



Abstract

Could you find 1 of your 1,000 Facebook friends in a crowd of 100? Even at a rate of 25 ms per comparison, determining that no friends were in the crowd would take more than 40 min if memory and visual search interacted linearly. In the experiment reported here, observers memorized pictures of 1 to 100 targets and then searched for any of these targets in visual displays of 1 to 16 objects. Response times varied linearly with visual set size but logarithmically with memory set size. Data from memory set sizes of 1 through 16 accurately predicted response times for different observers holding 100 objects in memory. The results would be consistent with a binary coding of visual objects in memory and are relevant to applied searches in which experts look for any of many items of interest (e.g., a radiologist running through a mental checklist of what might be wrong in a car-crash victim or an airport screener looking for any of a list of prohibited items in a carry-on bag).

Keywords

attention, memory, visual memory, visual attention, visual search

Received 9/12/11; Revision accepted 11/14/11

How do we search for one of many different target items in a visual world filled with many other objects? Although there are large literatures on visual search (Wolfe, 2010) and memory search (Van Zandt & Townsend, 1993), there has been 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 or digits (or both) in relatively small visual displays (Briggs & Blaha, 1969; Burrows & Murdock, 1969; Nickerson, 1966; Schneider & Shiffrin, 1977). For visual and memory set sizes in the range of 1 to 4 characters, search times are roughly linear with the product of the visual and memory set sizes (Schneider & Shiffrin, 1977). More recent work has painted a complicated picture of the effects of working memory loads on visual search (Balani, Soto, & Humphreys, 2010; Downing & Dodds, 2004; Han & Kim, 2004; Olivers, Peters, Houtkamp, & Roelfsema, 2011; Woodman, Vogel, & Luck, 2001), but as a general rule, these are studies in which working memory was pitted against search.

Researchers do not know how memory and vision work together when stimuli are drawn from the wide set of visual objects beyond alphanumeric characters, nor do they know what happens when the number 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 eight 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. The experiments reported here show that the “solution” to this problem is logarithmic search through the memory set. Earlier examples of reaction times (RTs) increasing with the log of the number of alternatives can be found in Burrows and Okada (1975) or in work on the Hick-Hyman law in motor response tasks (Hick, 1952; Hyman, 1953). Some work in the 1960s and 1970s 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.

Corresponding Author:

Jeremy M. Wolfe, Visual Attention Lab, Department of Surgery, Brigham & Women's Hospital, 64 Sidney St., Suite 170, Cambridge, MA 02139-4170
 E-mail: wolfe@search.bwh.harvard.edu

Experiment 1

Method

In Experiment 1, 10 observers 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. Every observer was tested on all five memory set sizes over five blocks of trials. All stimuli were photographs of isolated objects (provided by Talia Konkle). Stimuli were presented and responses collected on Macintosh computers running MATLAB and the Psychophysics Toolbox (Brainard, 1997). All observers gave informed consent and were paid \$10 per hour for their time.

In each memory block, the observers attempted to memorize a set of 1 to 16 items (Fig. 1a) and were then given a recognition test before proceeding to the visual search for these targets. In order to proceed to the visual search trials, observers had to score 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, observers attained an average accuracy of 97%. Memorization and testing of that memory took no more than 5 min, with longer times required for the larger memory set sizes. The criterion meant that the minimum number of memory blocks for a given memory set was 2. In practice, the average number of blocks required rose from 2.7 at the memory set size of 1 to 3.4 at the memory set size of 16. The modal number of required blocks was 2 at all memory set sizes.

With the memory set encoded with good accuracy, observers 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 the trials. No targets were present on the other 50%. Trials were randomly divided among the five visual set sizes (1, 2, 4, 8, and 16). Observers indicated by key press whether the target was present or absent, under instructions to be as quick and accurate as possible. They repeated the same process for each of the five memory set sizes.

Results

Figure 1c shows mean RT on target-present trials as a function of visual set size, for each of the five memory set sizes. It is clear that the effects of visual set size were 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. Data for the target-absent trials were comparable, with slopes about 2.5 times the target-present slopes. In contrast, as shown in Figure 1d, mean RTs were decidedly not linear as a function of memory set size. (Note that the same RTs are plotted in Figs. 1c and 1d, in one case as a function of visual set size and in the other case as a function of memory set size.) As shown in Figures 1e and 1f, mean RTs on both target-present and target-absent trials

appeared to be a linear function of the logarithm of memory set size. Regression coefficients were calculated for $RT \times \text{Memory Set Size}$ and $RT \times \text{Log}_2(\text{memory set size})$ functions; the fits were significantly better for the log_2 function for visual set sizes of 1, 4, 8, and 16, $t_s(9) > 2.3$, $ps < .05$; for the visual set size of 2, the log_2 function was a marginally better fit, $t(9) = 2.2$, $p = .058$.

Experiment 2: Memorizing 100 Objects

People's ability to remember large numbers of objects (Brady et al., 2008; Konkle et al., 2008) allows one 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 1, they were required to pass two consecutive memory tests at greater than 80% correct. This portion of the experiment took approximately 10 to 15 min. 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. More than 2,000 objects, sampled without replacement, were used as distractors. As in Experiment 1, a target was present on 50% of the trials.

The task was surprisingly easy. Miss error rates rose from 2% at the visual set size of 1 to 19% at the visual set size of 16. False alarms rose from 1% to 7%. The d' statistic fell 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 ms per visual item for target-present trials and 314 ms per item for target-absent trials.

Because just one memory set size was tested in Experiment 2, it is not possible to describe the shape of the $RT \times \text{Memory Set Size}$ function within this experiment. However, the data from Experiment 1 can be used to predict the RT for a memory set size of 100. Figures 1e and 1f show the average RTs for Experiment 2 (solid data points at the memory set size of 100), along with the linear regression of RT on $\text{log}_2(\text{memory set size})$ for the data of Experiment 1 (solid lines) and the extrapolation of the Experiment 1 data to the memory set size of 100 (dashed lines). It is clear that the logarithmic prediction is a remarkably good fit to the data. The average errors in prediction for target-present trials were -52 , 1 , 60 , 68 , and -74 ms for the visual set sizes of 1, 2, 4, 8, and 16, respectively (within 1%–5% of the average RT; see Fig. 1e). By comparison, the data from Experiment 2 fell short of the predictions of a linear model by $-1,156$, $-1,814$, $-2,505$, $-4,162$, and $-6,844$ ms, respectively. The linear prediction was about 200% of the actual average RT. The predictions of the logarithmic model were comparably good for the target-absent trials (see Fig. 1f).

Experiment 3: Localizing the Target

One could argue that the departure from linearity in Experiment 1 reflected a speed-accuracy trade-off (Shiffrin, 1988).

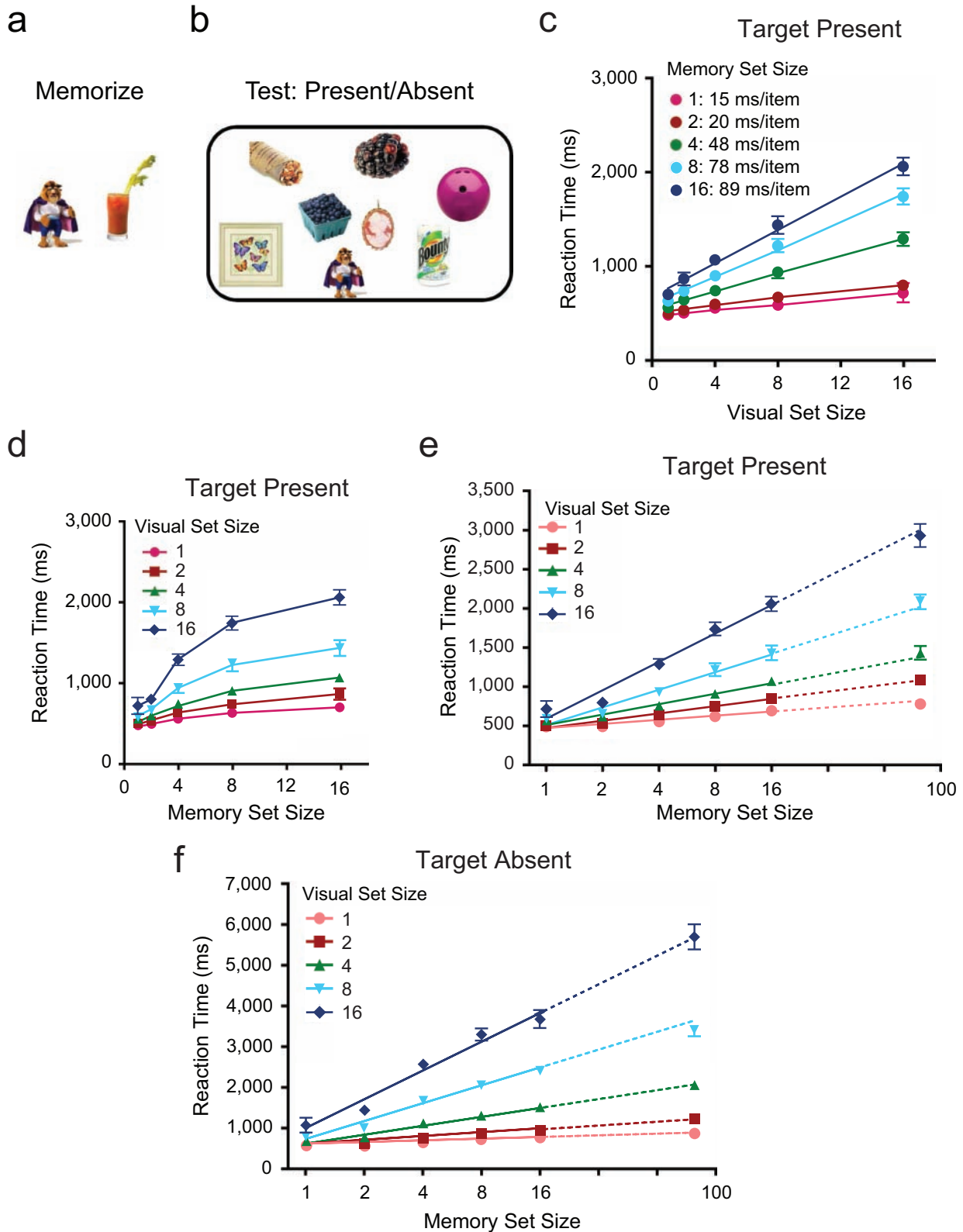


Fig. 1. Example memory set of two items (a), example visual search display with a set size of 8 (b), and experimental results (c–f) for Experiments 1 and 2. Observers memorized a set of 1 to 16 objects in Experiment 1 and 100 objects in Experiment 2 (a). After they were tested to confirm that they retained these objects in memory, they performed a visual search task in which they indicated whether a display of 1 to 16 objects contained any memorized target (b). The first graph (c) shows reaction time (RT) on target-present trials in Experiment 1 as a function of visual set size for each of the five memory set sizes; the key lists the slope of the RT × Visual Set Size function for each memory set size. The second graph (d) presents the same data, but in this case, the RTs are plotted as a function of memory set size for each of the five visual set sizes. The final two graphs incorporate the data from Experiment 2, extending the results to a memory set size of 100; RT is plotted as a function of memory set size, separately for each visual set size, on (e) target-present trials and (f) target-absent trials. Note the logarithmic scale of the x-axes. Solid lines are the best-fit regression lines for the RT × log₂(memory set size) function for the Experiment 1 data. Dashed lines are the extrapolation of those lines to the memory set size of 100. The solid data points at the set size of 100 show the results for the observers in Experiment 2, a group of 100 participants who did not participate in Experiment 1. In all the graphs, solid symbols indicate the averages across 10 observers, and error bars represent ±1 SEM.

Error rates rose with both visual and memory set sizes, both $F_s(1, 9) > 9.4$, both $p_s < .0001$, although they rose more markedly with memory set size. The maximum level of miss errors was 17% and occurred when the visual and memory set sizes were both 16. False alarm rates were low ($< 3\%$ in all cells). Larger error rates could reflect earlier termination of search, and earlier termination of search could have depressed RTs at the larger set sizes, resulting in a compression of the $RT \times Set Size$ function. This hypothesis was tested by repeating Experiment 1 using a localization response, rather than a “present”/“absent” key-press response. Ten new observers were tested with memory set sizes of 1, 2, 4, 8, and 16 and visual set sizes of 2, 4, 8, and 16. As in Experiment 1, observers needed to pass the memory test for a given set size twice. The average number of blocks required rose from 2.2 at the memory set size of 1 to 4.2 at the memory set size of 16. The modal number of required blocks was 2 at all memory set sizes. Final accuracy was above 97%. This portion of the experiment took about 5 min (though somewhat longer for the observer who required 12 tries to get 16 items into memory!).

After reaching the memory criterion, observers were tested on 300 trials, with one target present on each trial. Response was a mouse click on the target. Use of a localization response reduced error rates to 2% or less in all cells, with the exception that the error rate was 8% when the visual and memory set sizes were both 16.

Figure 2 shows the average RTs as a function of visual set size (Fig. 2a) and memory set size (Fig. 2b), with set size in

each case plotted on a linear scale. The functions relating RT to visual set size remained linear. The functions relating RT to memory set size were clearly nonlinear. Just as the data for the memory set size of 100 (Experiment 2) were predicted with the data for set sizes of 1 through 16 (Experiment 1), RTs for the memory set size of 16 were predicted from RTs for the set sizes of 1 through 8 in this experiment. The predictions of a logarithmic model (open circles in Fig. 2b) were close to the observed data, and we cannot reject this model, $F(1, 9) = 0.21$, $p = .65$, $\eta_p^2 = .003$. Note that the actual values were slightly lower than the values predicted by the logarithmic model, a result that may reflect a small speed-accuracy trade-off. In contrast, a linear extrapolation (asterisks in Fig. 2b) was less successful in predicting results for the memory set size of 16, and this model can be rejected, $F(1, 9) = 26.0$, $p = .0004$, $\eta_p^2 = .34$.

General Discussion

Thus, visual searches for multiple targets held in memory are accomplished in a reasonable amount of time because memory search time increases with the \log_2 of the memory set size. A hypothetical search for any of 1,000 friends in a picture of a 100 people would take more than 40 min if memory search and visual search were both linear and each step in the search took 25 ms. With logarithmic memory search, that time drops to a more plausible 25 s. 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

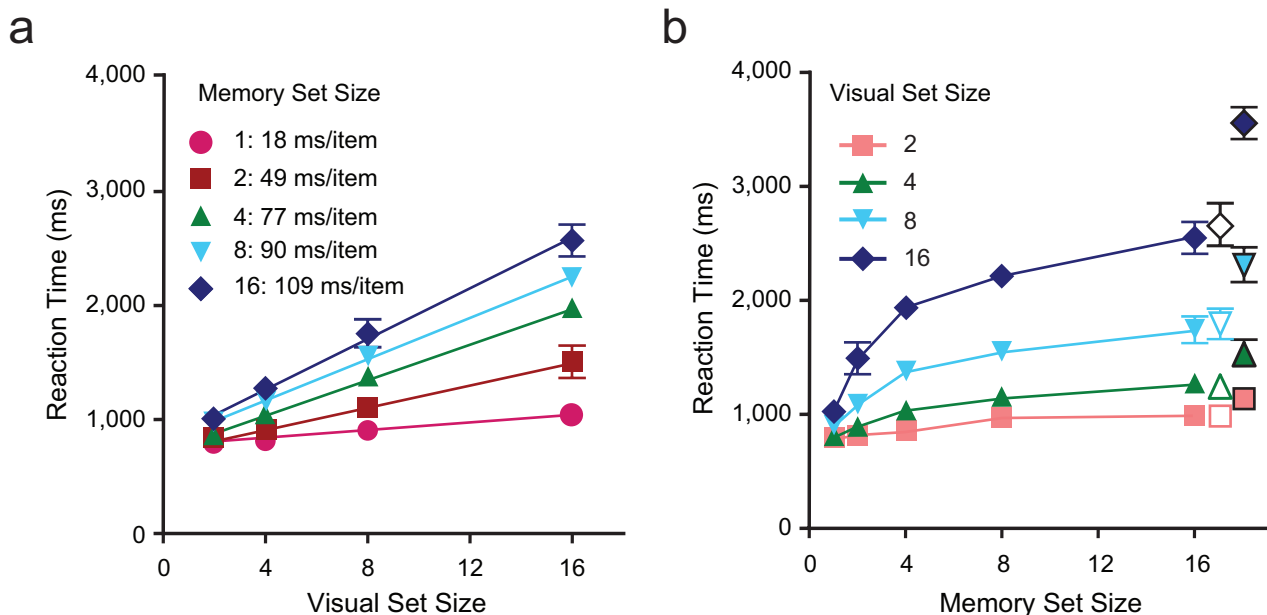


Fig. 2. Average reaction times (RTs) in Experiment 3, in which a target was present on each trial and observers localized the target with a mouse click. The graph in (a) shows RT as a function of visual set size, separately for each of the five memory set sizes; the key lists the slopes of the $RT \times Visual Set Size$ functions for the five memory set sizes. The graph in (b) shows RT as a function of memory set size, separately for each of the five visual set sizes. Open symbols show the predicted RTs for the memory set size of 16, extrapolated from memory set sizes 1 through 8, given a logarithmic relationship. Filled outline symbols show the much less successful predictions of a linear model. In both graphs, solid symbols indicate the averages across 10 observers, and error bars represent $\pm 1 SEM$.

occur if objects were represented in the equivalent of a binary code in memory. When an 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 one item was uniquely identified as belonging to the memory set of targets or all items were rejected as distractors. This account of logarithmic memory search 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 our 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.

Visual search and memory search interact in real-world search tasks. Consider the airport screener, who searches through the visual set of objects in a bag for any members of the remembered set of “threats.” Such real-world searches are likely to be more complex than the search tasks employed in the present experiments because objects are represented at more than one level (Rosch, 1973). Thus, a search for a robin, a wren, a blue jay, a sparrow, or a tiger might begin as a visual search for “birds” and “tigers”—a memory set size of 2. This might be followed by memory search to determine if a bird found in the display corresponds to one of the specific types of birds in the memory set. This memory set has a set size of 4, but is relevant only for birds. Search for threats might start with categories like “fluids” and “long metal objects,” with more detailed memory search performed only on items passing this first, categorical screen. Consider another interaction of memory and visual search. When a radiologist examines the whole-body scan of a patient who has been in a car crash, there is a very large set of remembered problems to search for. In this case, the memory set changes as a function of position in the image. For example, there is no point to searching for brain damage in the lower extremities. Again, the hybrid search will be a complex interplay of visual and memory search made more complicated if, as other evidence suggests, only one target “template” is active at any one moment (Houtkamp & Roelfsema, 2009; Olivers et al., 2011).

In sum, when you search a photo for any of your many Facebook friends, your attention will be guided to humans and away from other objects (Wolfe, 1994). Your 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.

Declaration of Conflicting Interests

The author declared that he had no conflicts of interest with respect to his authorship or the publication of this article.

References

- Balani, A. B., Soto, D., & Humphreys, G. W. (2010). Working memory and target-related distractor effects on visual search. *Memory & Cognition, 38*, 1058–1076.
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, USA, 105*, 14325–14329.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.
- Briggs, G. E., & Blaha, J. (1969). Memory retrieval and central comparison times in information processing. *Journal of Experimental Psychology, 79*(3, Pt. 1), 395–402.
- Burrows, D., & Murdock, B. B., Jr. (1969). Effects of extended practice on high-speed scanning. *Journal of Experimental Psychology, 82*(2), 231–237. doi:10.1037/h0028145
- Burrows, D., & Okada, R. (1975). Memory retrieval from long and short lists. *Science, 188*, 1031–1033.
- Downing, P., & Dodds, C. (2004). Competition in visual working memory for control of search. *Visual Cognition, 11*, 689–703.
- Han, S.-H., & Kim, M.-S. (2004). Visual search does not remain efficient when executive working memory is working. *Psychological Science, 15*, 623–628.
- Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4*, 11–26.
- Houtkamp, R., & Roelfsema, P. R. (2009). Matching of visual input to only one item at any one time. *Psychological Research, 73*, 317–326. doi:10.1007/s00426-008-0157-3
- Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. *Proceedings of the Institute of Radio Engineers, 40*, 1098–1102.
- Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 45*(3), 188–196.
- Konkle, T., Brady, T., Alvarez, G., & Oliva, A. (2008). Remembering thousands of objects with high fidelity. *Journal of Vision, 8*(6), Article 694. Retrieved from <http://journalofvision.org/8/6/694/>
- Neisser, U. (1974). Practiced card sorting for multiple targets. *Memory & Cognition, 2*, 781–785.
- Nickerson, R. S. (1966). Response times with a memory-dependent decision task. *Journal of Experimental Psychology, 72*(5), 761–769. doi:10.1037/h0023788
- Olivers, C. N., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences, 15*, 327–334. doi:S1364-6613(11)00085-4 [pii] 10.1016/j.tics.2011.05.004
- Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*, 328–350.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.
- Shiffrin, R. M. (1988). Attention. In R. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology* (2nd ed., Vol. 2, pp. 739–812). New York, NY: John Wiley.

- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*, 652–654.
- Van Zandt, T., & Townsend, J. T. (1993). Self-terminating versus exhaustive processes in rapid visual and memory search: An evaluative review. *Perception & Psychophysics*, *53*, 563–580.
- Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1), Article 8. Retrieved from <http://www.journalofvision.org/content/5/1/8>
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238.
- Wolfe, J. M. (1998). What do 1,000,000 trials tell us about visual search? *Psychological Science*, *9*, 33–39.
- Wolfe, J. M. (2010). Visual search. *Current Biology*, *20*, R346–R349. doi:10.1016/j.cub.2010.02.016
- Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science*, *12*, 219–224.