
Looking Around the Mind's Eye: How Internal Deployments of Attention Can Affect Visual Search Performance

Maithilee Kunda

Julia Ting

School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30308 USA

MKUNDA@GATECH.EDU

JULIA.TING@GATECH.EDU

Abstract

We present a new computational model that explores the role of visual working memory during the process of search, using the Embedded Figures Test (EFT) as an example visual search task. Our model offers new results to show how variations in attention to an internally stored target template can have a substantial effect on measures of search performance. Furthermore, these variations do not necessarily relate to quantitative differences in memory capacity, but in fact can arise through minor changes in sequences of attentional selection over the target template, without changing overall system storage capacities or search processes at all. We discuss the implications of our results for developing new computational models of visual attention and search as well as for the interpretation of human cognition through EFT observations.

1. Introduction

Visual search is a fundamental process of intelligence, guiding how agents (biological or otherwise) sample information from their visual environment, solve problems, and achieve goals. As AI systems are deployed in more and more complex, real-world settings, from sensor networks designed for surveillance to robots operating in naturalistic environments, understanding the detailed information-processing mechanisms that underlie visual search will be an important step towards developing robust and efficient search capabilities for these systems.

In studies of human cognition, search is often considered within the broader context of visual attention, in order to understand the spatial and temporal characteristics of overt or covert (i.e. with or without externally observable gaze shifts) deployments of attention across a search space in response to bottom-up (stimulus-driven) and top-down (goal-driven) factors.

A less widely studied aspect of visual search is how memory interacts with attention to produce observed search behaviors (Hutchinson & Turk-Browne, 2012). There are many roles that memory plays in visual search, including spatial memory of previous search patterns (Peterson, Kramer, Wang, Irwin, & McCarley, 2001) and memory-based cuing and priming of certain stimuli in the visual field (Desimone, 1996). In this paper, we focus on how variations in the **stored** memory representations of target items, particularly sequential, internal access to these representations, can affect overall search performance.

In general, this is a very complex problem, as stored memory representations of search target templates can come in drastically different flavors, depending on the search task. Simple visual

searches might use an iconic visual representation of the target, as in template-based searching, but one can imagine more complex search tasks in which the target is represented using a combination of visual, semantic, phonological, and other types of information, such as looking for “something to eat” in the wilderness, or looking for “something that rhymes with cat” in a picture book.

In simple visual searches, the idea that working memory (WM) can store and use an iconic visual template of the target item is borne out by evidence that visual information in WM is organized at the level of objects and not at the level of individual features (Luck & Vogel, 1997). Furthermore, individuals with high WM capacity for visual information do better on visual search than individuals with lower WM capacity (Reijnen, Hoffmann, & Wolfe, 2014).

There have been several previous computational models of visual search that operate at various levels of abstraction, including:

- Models that specify the use of iconic visual representations to guide visual search of objects on table (Rao, Zelinsky, Hayhoe, & Ballard, 2002);
- Models that account for overt visual search including saccade history (Zaharescu, Rothenstein, & Tsotsos, 2005);
- Models that implement a biased competition process with competition occurring between spatial, featural, and object information (Lanyon & Denham, 2004);
- Models that account for top-down and bottom-up search processes (Aziz & Mertsching, 2008);
- Models that plan saccades to maximize information gain (Pomplun, Reingold, & Shen, 2003);
- Models that utilize two stages of processing to initially guide and then refine visual search (Wolfe, 1994).

By and large, these models do not explicitly account for variations in the storage of or access to the target template that is used during search.

A notable exception is a computational model that investigates mechanisms for target acquisition and similarity comparisons (Zelinsky, 2008). This paper presents the Target Acquisition Model (TAM), which includes an explicit process of representing the search target image as an array of filter responses, intended to simulate the retinal processing of visual information. The search display is then represented in the same format, and search responses are modeled by correlating the feature vector associated with the target to the feature vector associated with the search display. While the TAM model does explicitly account for creation of the target template in memory, it does *not* address how the agent accesses this template during search; the TAM model assumes that the entirety of the target template is available at the same time, at each step during the search process.

In contrast, we explore what happens when the target template is itself subject to attentional deployments. Attention to different parts of visual working memory is often discussed in the literature as a change in the content of memory (Makovski, Sussman, & Jiang, 2008). However, it cannot be the case that unattended items simply disappear from visual working memory altogether, as people do have the capacity to sequentially remember and recall individual items from a single presentation of a set of many items.

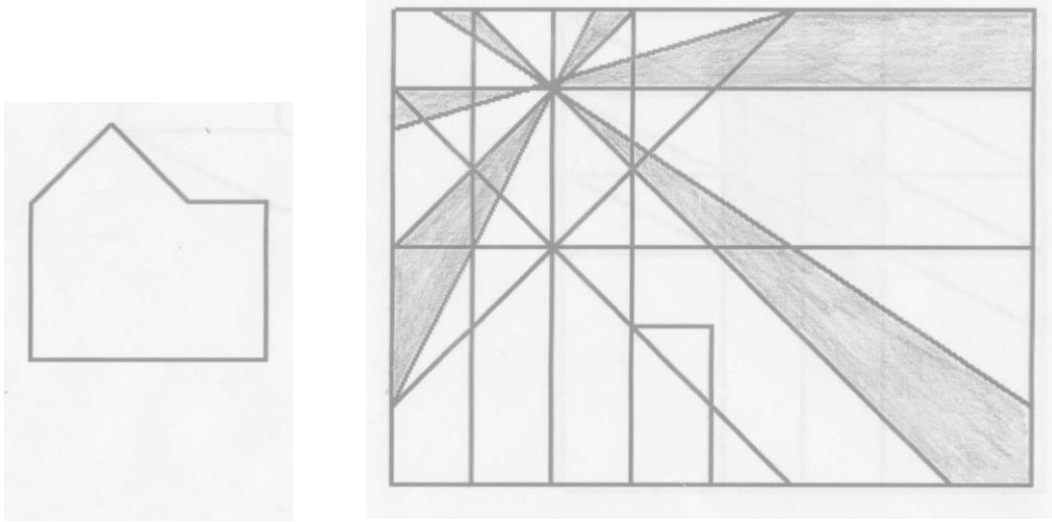


Figure 1. Example of problem similar to those found on the Embedded Figures Test (EFT). (Actual problems are not shown, in order to protect the security of the test.) The figure on the left is known as the “simple form” and must be located in the “test form” on the right. The simple form is present in the test form in its given configuration; rotating, scaling, or otherwise transforming the simple form to find it in the test form are not allowed. Notably, while taking the test, the simple form and the test form are never simultaneously visible; the test-taker *must* store a representation of the simple form in memory before attempting to search for it in the test form.

We hypothesize that attention can be deployed internally to different spatial locations within visual working memory. This hypothesis is analogous to current theories of how attention can be deployed covertly (i.e. without associated eye movements) to different locations of the visual field. Our hypothesis is also consistent with findings from mental imagery that visual working memory does contain similar (and functionally useful) neural representations of visual information that are in many ways comparable what is received during perception (Kosslyn et al., 1999; Kosslyn, Thompson, Kim, & Alpert, 1995; Slotnick, Thompson, & Kosslyn, 2005; Stokes, Thompson, Cusack, & Duncan, 2009).

In this paper, we present a new computational model of visual search that explores this relationship between visual working memory, internally deployed attention, and search performance, by looking at how variations in sequences of attention to the target template can affect overall search efficiency. The task domain that we study is the Embedded Figures Test (EFT), a widely used neuropsychological assessment in which participants must search for a simple geometric figure within a larger, more complex figure.

We first describe the EFT in more detail, including key findings from human cognition that motivate our work. Next, we describe our computational model and present detailed results from the example problem shown in Fig. 1. Then, we present experimental results from running different configurations of the model on the actual EFT. Finally, we close with a discussion of our results and their implications for understanding the mechanisms that contribute to visual search and how these mechanisms can be leveraged in the design of intelligent systems.

2. The Embedded Figures Test (EFT)

Fig. 1 shows an example of an EFT-type problem. The Embedded Figures Test (EFT) was originally designed by Herman Witkin as a measure of “field-independence,” which refers to how well an individual can differentiate an individual stimulus from background elements or patterns (Witkin, 1950). Faster search times on the EFT are assumed to indicate greater field-independence.

Witkin describes the original EFT as a test that is given to a subject by an examiner (Witkin, Oltman, Raskin, & Karp, 1971). For each test item, the examiner first presents the “test form,” which is the complex figure to be searched, i.e. the search environment, and then presents a “simple form,” which is the item to be found, i.e. the search target. Then, the test form is presented to the subject once more, who then begins the actual search. Witkin specifies that the simple form and the test form should never be presented to the subject at the same time, but that the subject can ask to refer back to the simple form as needed. This setup effectively *requires* the subject to store the simple form in memory, before searching for it in the test form. Performance is measuring according to the time taken by the subject to complete each item.

There are several variants of the EFT that are currently in use (Ludwig & Lachnit, 2004); the most widely used seem to be the Group Embedded Figures Test (GEFT) and the Children’s Embedded Figures Test (CEFT). The GEFT uses some of the same items as the EFT along with some new items and was designed to be administered in a group setting (Witkin et al., 1971). The GEFT uses a paper-and-pencil format, and subjects are given fixed time limits to complete three different sets of items. The GEFT is scored according to the number of items correctly completed within the time limits. The GEFT uses a clever design in which each page presents a complex form together with a letter indicating which simple form is to be found. The set of simple forms, along with labels, is printed on the back of the test booklet. This design enforces Witkin’s specifications that 1) the test form is seen prior to the simple form for each item, and 2) the test form and simple form are never simultaneously visible to the subject.

The CEFT was designed to be an easier, more engaging test than the original EFT for use with young children. The CEFT introduces concrete shapes (e.g. houses, tents, strollers) as both the simple forms and test forms (Goodenough & Eagle, 1963). The CEFT is administered in a manner similar to the EFT, with a single examiner and a single subject (Witkin et al., 1971). Scores are recorded as the number of items correctly solved, though many research studies using the CEFT also record the time to completion as a variable of interest.

In this paper, we focus on the GEFT for our computational experiments. However, our observations do apply across all of the EFT variants, and so we use the more general abbreviation “EFT” to refer to this task domain throughout the remainder of this paper.

Studies of the EFT in typically developing individuals have found that EFT performance is related to performance on other, similar “disembedding”-type tasks (Ghent, 1956). In addition, certain manipulations in administration formats, i.e. group vs. individual administration, differences in the coloration of test items, and memory requirements imposed by the task administration format can affect performance in significant ways (Jackson, Messick, & Myers, 1964). Interestingly, there have been substantive sex differences observed for EFT performance, though practice appears to reduce or remove these differences (Goldstein & Chance, 1965). Many cultural differences in EFT performance have been observed as well (Kühnen et al., 2001).

Over the last few decades, many studies have found interesting patterns of differences in EFT performance between typically developing individuals and individuals diagnosed with autism spectrum disorders (ASD). These studies generally find that individuals diagnosed with ASD show superior performance on the EFT, in line with performance on other visual search tasks (Jarrold, Gilchrist, & Bender, 2005), either in the form of improved accuracy or shorter reaction times, or both (Shah & Frith, 1983). Similar patterns have also been found in individuals diagnosed with Asperger syndrome (Jolliffe & Baron-Cohen, 1997), and these differences appear to be related to differences in brain activity (Ring et al., 1999) and eye fixation durations (Keehn et al., 2009). There have also been observations of interactions between performance and cultural differences in these populations (Koh & Milne, 2012).

The EFT shares similarities with some other cognitive tests in this regard; the EFT, a test called block design, and another test called the Raven’s Progressive Matrices (RPM) are all often found to represent “peaks” of ability among individuals diagnosed with ASD (Dawson, Soulières, Gernsbacher, & Mottron, 2007; Shah & Frith, 1993). Like the EFT, the block design task, in which colored blocks must be put together to match a given pattern, and the RPM, in which an incomplete matrix of figures must be completed with the correct missing figure, are visually presented tasks that involve visual reasoning, as opposed to linguistic or semantic reasoning (Kunda, McGreggor, & Goel, 2013). All three tests are widely used as cognitive assessments in clinical and scientific settings.

In addition to perceptual, motor, and other reasoning components, we believe visual memory plays an important role across all of these tasks. Understanding the specific mechanisms at play in each would greatly improve the usefulness of these tasks as cognitive assessments as well as our general understanding of visual cognitive processes in humans and in artificial agents, with special relevance to populations who have experience atypical cognitive development.

In previous work, we examined the nature of problem solving using purely visual representations on the RPM (Kunda et al., 2013). Here, we focus on the EFT to more closely examine the interplay between visual memory and visual search. Ultimately, we aim to develop integrated models that combine perception, memory, and reasoning across all of these tasks.

3. Our Model

Our model of the EFT is different from many previous computational models of visual search and attention because it does not focus on notions of saliency as the primary variables of interest. Instead, our model focuses on the role of visual working memory as a key part of the overall search task. In particular, our model contains parameters that produce changes in how internally-directed attention is deployed sequentially to the stored search template, in order to experimentally examine how these variations affect overall search performance.

Interestingly, in Witkin’s original EFT paper, qualitative observations about people’s search strategies indicate that individuals pick a “complex” part of the simple form to anchor their search at various points in the test form, and then try to trace the outline in the complex figure (Witkin, 1950). It is this observation about human behavior that motivates the design of our model.

In particular, the model takes as input PNG image files of the simple form and test form for each EFT problem, scanned directly from a paper copy of the test. The model reads each input as

a grayscale image and represent it as a two dimensional array of binary true/false pixels, thresholded using single static threshold values chosen manually for each problem. Values less than the specified threshold, or grayscale value, are considered to be black (or true), and values greater than the specified threshold are considered white (or false).

To compute image similarity, the program implements template-based matching. In particular, for two images A and B, the program applies A as a convolution filter to B; similarity at each desired image offset is computed using a modified Jaccard coefficient, with the number of pixels in the intersection of the simple and test forms divided by the total number of pixels in the simple form. Thus, maximum image similarity between images A and B is computed using the Jaccard coefficient over binary pixel counts.

While solving a problem, the model stores each simple form as an ordered collection of visual “features,” which are defined manually according to points of interest in the simple form (line 4 in Fig. 3). Fig. 2 (top) illustrates how the example simple form shown in Fig. 1 has been broken into a small collection of features. Then, the model uses these features to define a first-stage saliency map according to all the locations in the test form where each of these features is observed, i.e. locations that have high similarity values (lines 7-9 in Fig. 3). Fig. 2 (bottom) shows an example of a partial saliency map computed for the example problem shown in Fig. 1, using the topmost corner of the simple form as the current feature of interest.

The model then performs a random walk, without replacement, of these high-saliency points in the test form (line 13 in Fig. 3). At each location, or “**fixation**,” the model computes its template-based similarity matching between each feature in the simple form and the corresponding location in the test form to determine whether a match has been found (lines 16-20 in Fig. 3). This term comes from studies of human attention and eye movements, in which a fixation refers to a short time interval during which gaze is directed to a specific location in the environment. Whenever our model makes a comparison, it is simulating a fixation in the sense that the comparison takes place over a short time period in which a perceptual operation over a small, localized region of the visual environment must be completed before the model can move onto the next step in its search process.

From each location, the model searches within a 20x20 pixel window of x-y alignments in order to determine whether a match exists, as determined by a thresholded similarity value (lines 39-42 in Fig. 3). Once a match is found, the model ceases its search and goes onto the next feature, or if all features have been found, the item has been solved (lines 24-25 in Fig. 3). The model is able to search using any ordered collection of features that describe the simple form, and the model can also treat any given feature as the “**anchor feature**” that defines the initial saliency map.

We also implemented one additional variation in the model, which is that instead of storing the simple form as a sequence of individual features during the similarity matching, the model can store the simple form as a single search target and perform the similarity matching in one step (lines 27-35 in Fig. 3). We refer to the first variation as performing “**PieceWise**” comparison of the simple form, going feature by feature, and the second variation as performing “**Comprehensive**” comparisons of the entire simple form at once.

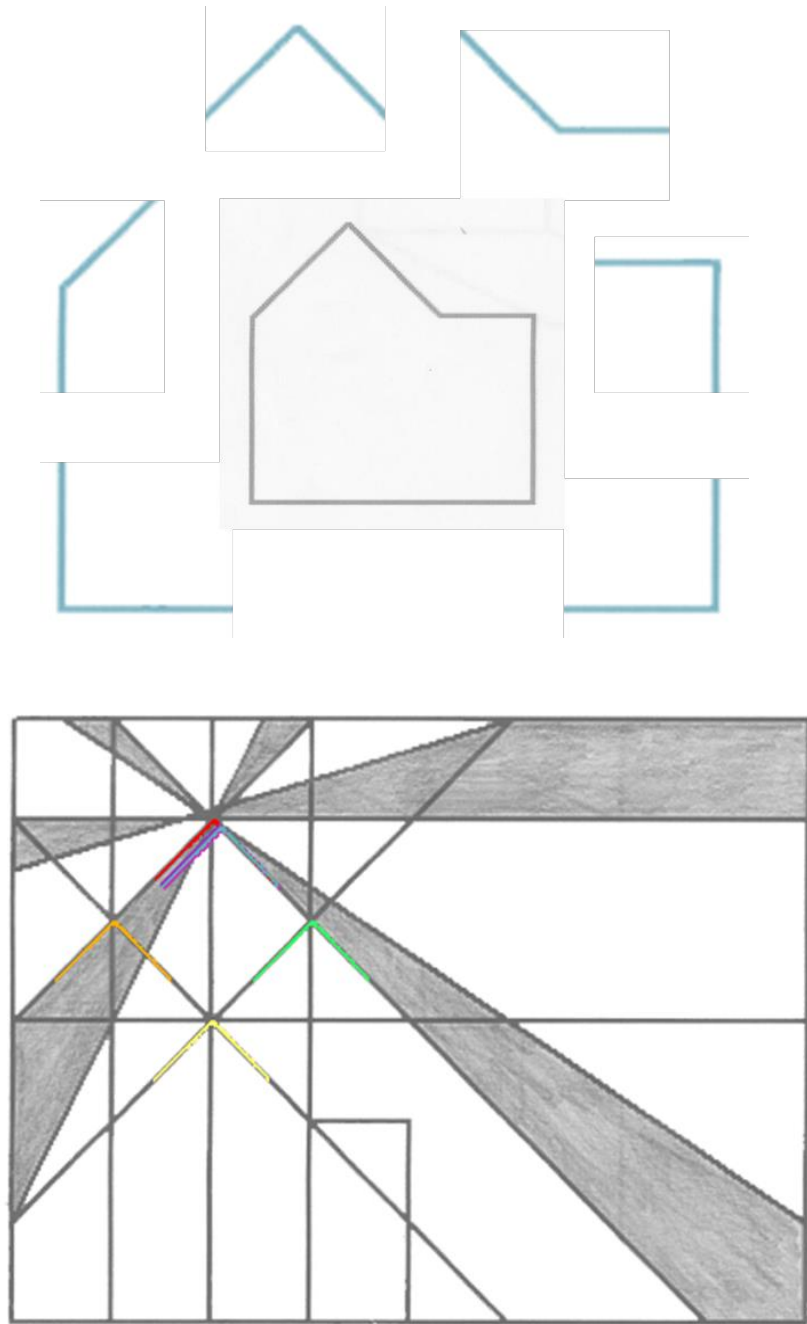


Figure 2. Top: Example simple form from Fig. 1, carved up into six visual features of interest with which to search test form. Bottom: Example of partial saliency map created for problem shown in Fig. 1. Anchor feature is defined to be the topmost corner in the simple form, and five high-similarity matching locations for this anchor feature in the test form are shown. (This figure is best viewed in color.)

```

1  solveItem(Image SimpleForm, Image TestForm, int anchor)
2      // first divide SimpleForm into features  $f_1, f_2, \dots, f_n$ 
3      // for our experiments, this step was performed manually for each EFT item
4       $f_1, f_2, \dots, f_n$  = ordered list of subimages that together comprise SimpleForm
5
6      // choose  $f_{\text{anchor}}$  as anchor feature and compute saliency map over TestForm
7      for all  $(x, y)$  in TestForm
8           $\text{saliency}_{(x,y)} = \frac{\sum_{i,j} (\text{TestForm}_{i+x,j+y} \cap f_{\text{anchor}_{i,j}})}{\sum_{i,j} (f_{\text{anchor}_{i,j}})}$ 
9          if  $\text{saliency}_{(x,y)} > \text{threshold}$ 
10             add  $(x, y)$  to list of high-saliency points HS
11
12     // do random walk search
13     numFixations = 0
14     while HS is not empty
15          $(x_{\text{anchor}}, y_{\text{anchor}})$  = point randomly selected (and removed) from HS
16         // PieceWise strategy: Comparisons are made using each feature
17         if PieceWise strategy is selected
18             for  $i$  in 1 to  $n$ , starting with anchor
19                 Target =  $f_i$ 
20                  $(x_{\text{search}}, y_{\text{search}})$  =  $(x_{\text{anchor}}, y_{\text{anchor}})$  + offset of  $f_i$  in SimpleForm
21                 bool success = doComparison(Target, TestForm,  $(x_{\text{search}}, y_{\text{search}})$ )
22                 numFixations++
23                 if not success
24                     break
25             end of for loop
26             if success // search has succeeded
27                 return numFixations
28         end of PieceWise strategy
29
30     // Comprehensive strategy: Comparisons are made using the entire SimpleForm
31     if Comprehensive strategy is selected
32         Target = SimpleForm
33          $(x_{\text{search}}, y_{\text{search}})$  =  $(x_{\text{anchor}}, y_{\text{anchor}})$ 
34         bool success = doComparison(Target, TestForm,  $(x_{\text{search}}, y_{\text{search}})$ )
35         numFixations++
36         if success // search has succeeded
37             return numFixations
38     end of Comprehensive strategy
39     end of while loop
40     return null // search has failed
41 end of function solveItem
42
43 doComparison(Image Target, Image TestForm, int  $(x_{\text{search}}, y_{\text{search}})$ )
44      $\text{similarity} = \frac{\sum_{i,j} (\text{TestForm}_{i+x,j+y} \cap \text{Target}_{i,j})}{\sum_{i,j} (\text{Target}_{i,j})}$ 
45     return ( $\text{similarity} > \text{threshold}$ )
46 end of function doComparison

```

Figure 3. Pseudocode for EFT computational model, with both PieceWise and Comprehensive strategies.

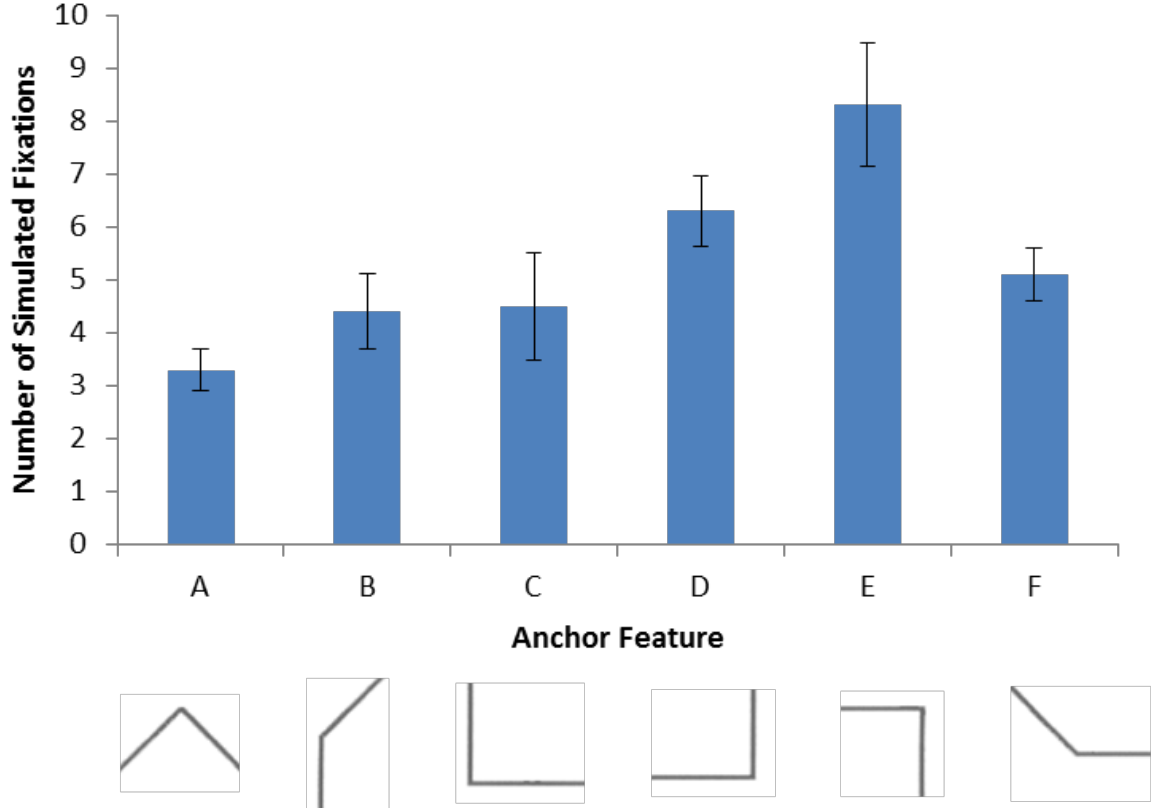


Figure 4. Number of simulated fixations made by the model to solve the example problem shown in Fig. 1, using the Comprehensive matching strategy, across different choices of anchor features. This graph shows mean number of fixations made across 10 runs, and error bars indicate the standard error.

4. Experimental Design and Results

We tested this model across variations in two independent variables: the feature selected to be the **anchor feature**, and also the matching strategy that is used (**Piecewise** vs. **Comprehensive**). The dependent variable is the number of comparisons that the model makes before finding a correct solution to the item. We label this dependent variable as the number of “**fixations**” made by the model.

Because the model performs a random selection as part of its search, results can be highly varied across multiple runs of the model. For the experiments presented in this paper, we ran the model on each EFT item 10 times, to get aggregate measures of performance. We also tested a few items at 100 runs; qualitative observations seemed to indicate that aggregate performance was not considerably different between 10 and 100 runs, and so we chose 10 runs for the full set of experiments. Here, we present detailed results from the example problem shown in Fig. 1 and Fig. 2, along with summary results for the actual items from the EFT.

Fig. 4 shows the mean (and standard error) for number of fixations made by the model in solving the example problem, for various choices of the anchor feature, using the **Comprehensive** matching strategy. The anchor feature corresponding to the topmost point of the simple form resulting in the fewest number of fixations, averaging about 3 fixations to find the solution, while the top right corner of the simple form resulting in the largest number of fixations, averaging about 8 to find the solution.

Fig. 5 shows the minimum, mean, and maximum number of fixations over all possible anchor features, averaged across 10 search trials, for each of the 18 problems on the EFT. Note that one problem, problem 4, was not able to be solved by the model due to misalignments in the original figure drawings presented in the EFT booklet.

Fig. 6 shows comparisons between the number of simulated fixations made using the comprehensive matching strategy and the piecewise matching strategy, averaged over 10 search trials, on the problem from Fig. 1. We also ran these two different strategies on the problems from the EFT and observed similar patterns of performance.

5. Discussion

The primary finding of interest from our experiments is that the choice of anchor feature can have a significant impact on the overall efficiency of the search. This is notable because the choice of anchor feature is certainly an element of search strategy that can vary from individual to individual, and in fact these differences do not require differences in memory capacity or even in the overall structure of the search process. Yet these individual differences, if they exist, can yield vastly different levels of performance.

As shown in Fig. 4, the difference in performance can vary by more than a factor of two, as we see moving from the most efficient anchor feature, feature A, to the least efficient, feature E. Fig. 5 shows that these patterns persist across the actual problems on the EFT, with very sizable differences for each problem between the least and most efficient anchor features. Fig. 6 shows differences that emerge between the **Comprehensive** and **PieceWise** matching strategies. As expected, the piecewise strategy involves many more fixations than the comprehensive strategy, as it is doing the final matching feature-by-feature instead of using the entire simple form at once. However, what is interesting about these results is that, again, the choice of anchor feature actually has a considerable impact on *how much* the piecewise strategy contributes to the overall slowing of the search.

For feature A, again the most efficient, performing a piecewise search reduces search duration by about a factor of three. For feature D, the least efficient according to this new metric, the search is slowed by about a factor of 9. Similar patterns were observed across the actual test items from the EFT, though detailed results are not shown here.

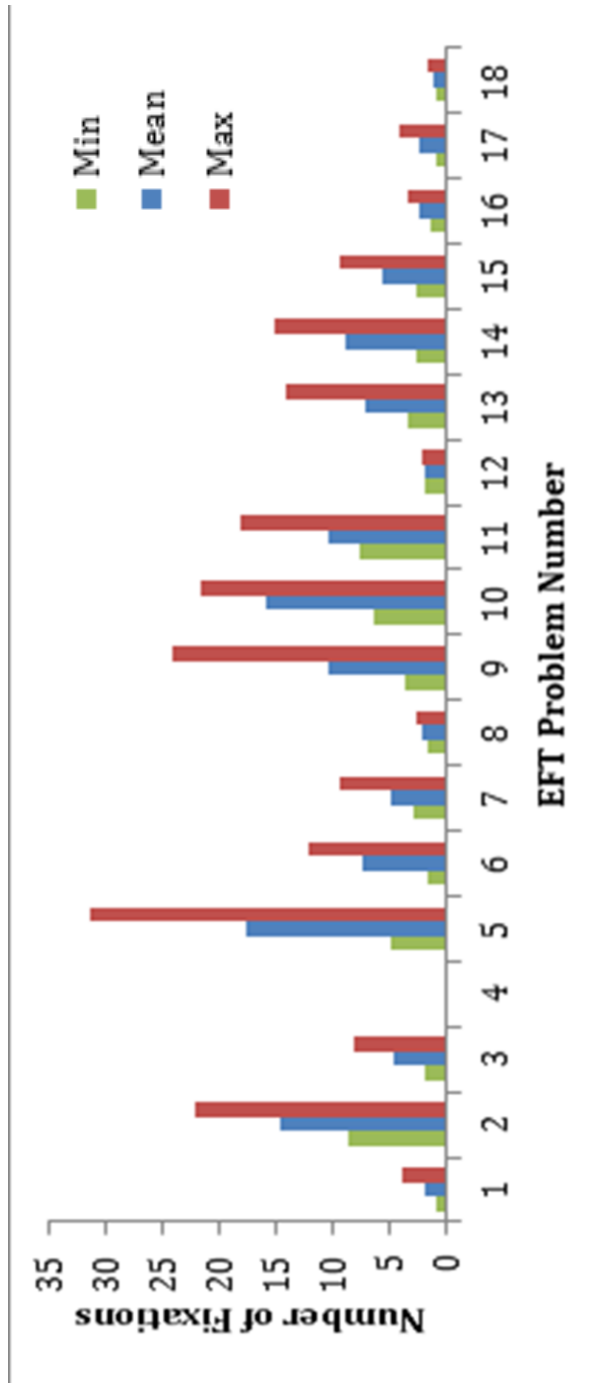


Figure 5. Number of simulated fixations (minimum, mean, and maximum across all choices of anchor feature) made by the model to solve 18 problems from the EFT, averaged over 10 search trials.

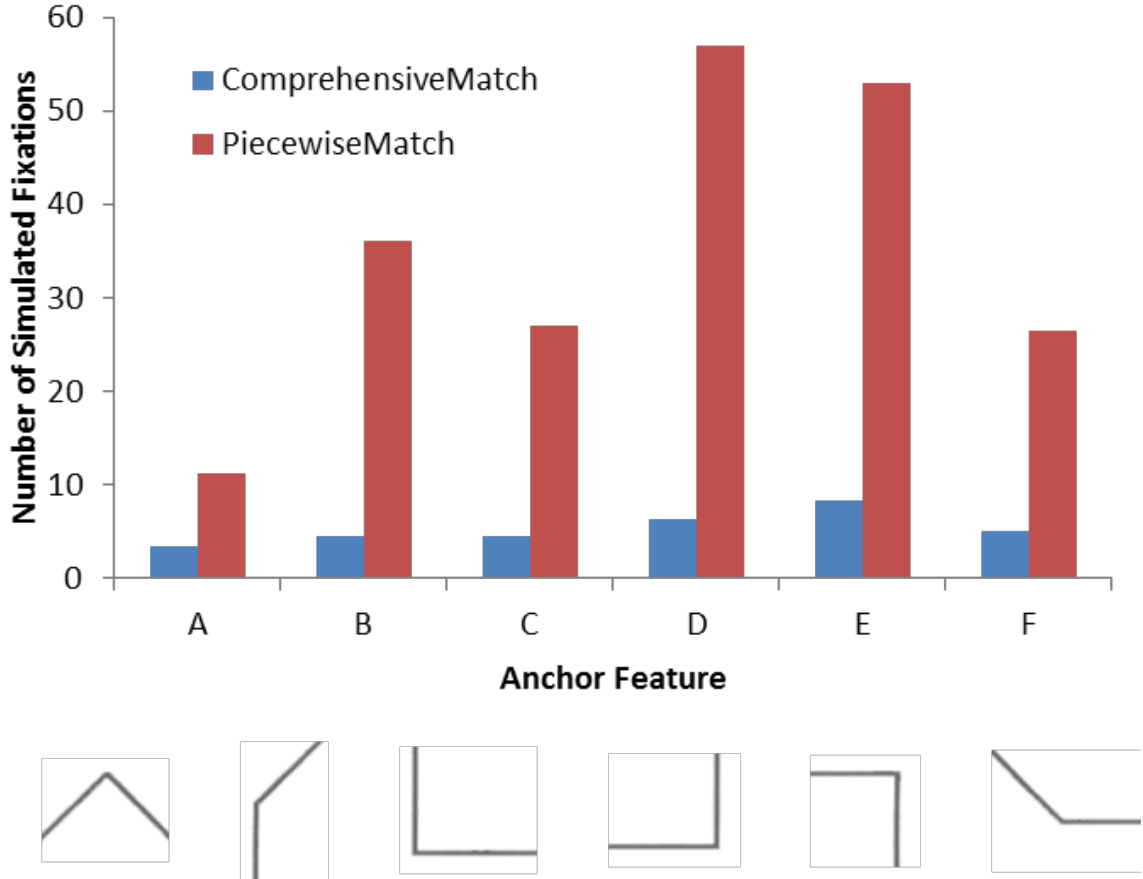


Figure 6. Example of partial saliency map created for problem shown in Fig. 1. Anchor feature is defined to be the topmost corner in the simple form, and five high-similarity matching locations for this anchor feature in the test form are shown.

There are two main implications of these results. First, computational models of search need to take visual memory for the target template into account, and not just in terms of memory capacity but also in terms of access, indexing, and attention to the target template *within* memory. Many computational theories and models of attention focus exclusively on how attention is deployed spatially and temporally across the search space, but our computational experiments have shown that, in fact, how attention is deployed within the target template can have an enormous impact on search performance. We have shown that this impact can be observed both in total search times, for simple searches, as well as in nonlinear effects on search times in more complex searches.

Second, in relation to studies of cognition and search using the EFT, we have identified a significant potential source of individual differences on the task that has not been previously discussed in the EFT literature: the choice of an anchor feature while solving each item. Our results serve as a proof by existence that this choice can affect search results. Our model does not intelligently make this choice; instead, our experiments exhaustively tested every one of a set of

features as the anchor feature, and we observed the effects of this variation on search performance. The next important research question, then, is how this choice of an anchor feature is made by human subjects, if this is indeed the strategy that humans use. An extension of this question becomes how artificial agents implementing this strategy should go about choosing such features during visual search tasks.

One possibility is that the anchor feature is chosen at random (as our model does). This seems unlikely, given our results about the importance of this choice and given the rarity of humans exhibiting truly random perceptual or cognitive behaviors. Given the wealth of literature on the combined effects of top-down and bottom-up influences on attention, and the fact that the choice of anchor feature is essentially a deployment of attention, it seems much more likely that some combination of perceptual and cognitive factors will drive a person's choice of anchor feature.

Another possibility is that the anchor feature is chosen based on properties of the simple form alone. However, there are two interesting observations coming from Witkin's design of the EFT that make this seem unlikely. On the EFT, the simple form changes from one item to the next, and Witkin found this to be an important aspect of the overall test administration (Witkin et al., 1971). He observed that if the simple form were kept the same for several items in a row, then the discriminability of the test seemed to be lessened, as all participants would begin to show increased field-independence, or the ability to easily find the simple form regardless of the complexity of the test form.

Witkin also specifies that the test form should be presented first, and that the subject should spend some time inspecting the test form before looking at the simple form. On both the original EFT and the CEFT, the subject is asked by the examiner to describe the test form out loud, to ensure that they are sufficiently attending to it: "During the initial 15-second exposure of each Complex Figure, the Subject should be asked to describe it in any way he pleases. The purpose of this procedure is to impress the organization of the Complex Figure upon the Subject" (Witkin et al., 1971, p. 17).

This suggests that there is some kind of mental set induced by looking at the test form that is necessary to EFT items functioning in the intended way. If the anchor feature were determined purely by the simple form, then this early presentation of the test form would make no difference to the anchor feature selection aspect of the search strategy.

The third possibility, and the most interesting one with respect to our results, is that perhaps the anchor feature in the simple form is chosen based on visual properties of the *test form*, after it has been initially seen and inspected by the subject. In fact, the efficiency of search using this type of strategy is directly a function of how frequently the anchor feature is found in the test form, not its relation to other features in the simple form. This is somewhat counter-intuitive; if one is searching for a target, it might seem sensible to pick the most distinctive part of the target to use as the anchor feature. However, under our model, the optimal strategy would be to pick the part of the target that is most distinctive in the search environment, even if it occurs multiple times in the target.

So what, then, is happening when the subject is gazing (for 15 seconds, which is a very long time) at the test form before looking at the simple form? In line with studies of visual priming, we conjecture that the subject might, in fact, be primed to attend to features of the simple form that were *most* frequent in the test form. If this occurs, it will actually result in the worst possible

choice of anchor feature! Subjects who exhibit the least amount of this kind of perceptual priming will thus gain an advantage on the task. This is entirely consistent with the general idea of the EFT being a test of “field-independence,” if we define field-independence as increased freedom from this perceptual priming effect. This idea is also consistent with Witkin’s observations that if the same simple form is used over and over, the ability of the EFT to discriminate field-independence is lessened.

6. Contributions

We have shown, using a new computational model of search on the Embedded Figures Test (EFT), that differences in internally deployed attention, i.e. sequential attention to the contents of visual working memory, can cause substantial differences in overall search performance. Our experiments directly support the existence of this relationship as a potential explanation for individual differences on the EFT.

Indirectly, our experiments have suggested to us an explanation of how the sequences of these internal attention deployments are selected. We propose that when a subject looks at the test form in an EFT item, and then afterwards are asked to look at the simple form, they exhibit a priming effect that causes them to attend to the part of the simple form that was most prevalent in the test form. The magnitude of this priming effect will have a direct influence on the quality of the anchor feature that the subject chooses; the larger the priming effect, the worse the anchor feature and overall search performance. Thus, our anchor-feature-based model, together with a priming mechanism that drives the selection of the anchor feature, represents one possible explanation of the origin of “field-independence” as a cognitive construct. To our knowledge, no other computational mechanism at this level of detail has been proposed as a possible explanation for the construct of “field-independence” in human cognition.

Our model of search on the EFT is admittedly simplified in many ways. The model does not account for traditional notions of visual salience, which have been extensively described in the attention literature and are generally assumed to be the primary drivers of visual search performance. However, our model does identify an additional mechanism that has been left out of much of the extant literature. Models that integrate the internal deployments of attention that we describe here together with other aspects of perception, such as salience and Gestalt perception, will undoubtedly be crucial to more fully understanding the computational and cognitive strategies used by intelligent agents on visual search tasks.

Acknowledgments

We would like to thank Sy Rashid for valuable contributions during the early stages of this research. Portions of this work were supported through funding from the Undergraduate Research Opportunities Program (UROP) at Georgia Tech and through the National Science Foundation (NSF) Expedition Award #1029679.

References

- Aziz, M. Z., & Mertsching, B. (2008). Visual Search in Static and Dynamic Scenes Using Fine-Grain Top-Down Visual Attention. In A. Gasteratos, M. Vincze, & J. K. Tsotsos (Eds.), *Computer Vision Systems* (pp. 3–12). Springer Berlin Heidelberg.
- Dawson, M., Soulières, I., Gernsbacher, M. A., & Motttron, L. (2007). The Level and Nature of Autistic Intelligence. *Psychological Science*, 18(8), 657–662.
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences*, 93(24), 13494–13499.
- Ghent, L. (1956). Perception of Overlapping and Embedded Figures by Children of Different Ages. *The American Journal of Psychology*, 69(4), 575–587.
- Goldstein, A. G., & Chance, J. E. (1965). Effects of practice on sex-related differences in performance on Embedded Figures. *Psychonomic Science*, 3(1-12), 361–362.
- Goodenough, D. R., & Eagle, C. J. (1963). A Modification of the Embedded-Figures Test for Use with Young Children. *The Journal of Genetic Psychology*, 103(1), 67–74.
- Hutchinson, J. B., & Turk-Browne, N. B. (2012). Memory-guided attention: Control from multiple memory systems. *Trends in Cognitive Sciences*, 16(12), 576–579.
- Jackson, D. N., Messick, S., & Myers, C. T. (1964). Evaluation of group and individual forms of embedded-figures measures of field-independence. *Educational and Psychological Measurement*, 24(2), 177–191.
- Jarrold, C., Gilchrist, I. D., & Bender, A. (2005). Embedded figures detection in autism and typical development: preliminary evidence of a double dissociation in relationships with visual search. *Developmental Science*, 8(4), 344–351.
- Jolliffe, T., & Baron-Cohen, S. (1997). Are People with Autism and Asperger Syndrome Faster Than Normal on the Embedded Figures Test? *Journal of Child Psychology and Psychiatry*, 38(5), 527–534.
- Keehn, B., Brenner, L. A., Ramos, A. I., Lincoln, A. J., Marshall, S. P., & Müller, R.-A. (2009). Brief Report: Eye-Movement Patterns During an Embedded Figures Test in Children with ASD. *Journal of Autism and Developmental Disorders*, 39(2), 383–387.
- Koh, H. C., & Milne, E. (2012). Evidence for a Cultural Influence on Field-Independence in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 42(2), 181–190.
- Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., L. W., ... Alpert. (1999). The Role of Area 17 in Visual Imagery: Convergent Evidence from PET and rTMS. *Science*, 284(5411), 167–170.
- Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, 378(6556), 496–498.
- Kühnen, U., Hannover, B., Roeder, U., Shah, A. A., Schubert, B., Upmeyer, A., & Zakaria, S. (2001). Cross-Cultural Variations in Identifying Embedded Figures Comparisons from the United States, Germany, Russia, and Malaysia. *Journal of Cross-Cultural Psychology*, 32(3), 366–372.
- Kunda, M., McGregor, K., & Goel, A. K. (2013). A computational model for solving problems from the Raven's Progressive Matrices intelligence test using iconic visual representations. *Cognitive Systems Research*, 22–23, 47–66.
- Lanyon, L. J., & Denham, S. L. (2004). A model of active visual search with object-based attention guiding scan paths. *Neural Networks*, 17(5–6), 873–897.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281.

- Ludwig, I., & Lachnit, H. (2004). Effects of practice and transfer in the detection of embedded figures. *Psychological Research*, 68(4), 277–288.
- Makovski, T., Sussman, R., & Jiang, Y. V. (2008). Orienting attention in visual working memory reduces interference from memory probes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2), 369–380.
- Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual Search has Memory. *Psychological Science*, 12(4), 287–292.
- Pomplun, M., Reingold, E. M., & Shen, J. (2003). Area activation: a computational model of saccadic selectivity in visual search. *Cognitive Science*, 27(2), 299–312.
- Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, 42(11), 1447–1463.
- Reijnen, E., Hoffmann, J., & Wolfe, J. (2014). The role of working memory capacity in visual search and search of visual short term memory. *Journal of Vision*, 14(10), 1073–1073.
- Ring, H. A., Baron-Cohen, S., Wheelwright, S., Williams, S. C. R., Brammer, M., Andrew, C., & Bullmore, E. T. (1999). Cerebral correlates of preserved cognitive skills in autism A functional MRI study of Embedded Figures Task performance. *Brain*, 122(7), 1305–1315.
- Shah, A., & Frith, U. (1983). An Islet of Ability in Autistic Children: A Research Note. *Journal of Child Psychology and Psychiatry*, 24(4), 613–620.
- Shah, A., & Frith, U. (1993). Why do autistic individuals show superior performance on the block design task? *Journal of Child Psychology and Psychiatry*, 34(8), 1351–1364.
- Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex (New York, N.Y.: 1991)*, 15(10), 1570–1583.
- Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-Down Activation of Shape-Specific Population Codes in Visual Cortex during Mental Imagery. *The Journal of Neuroscience*, 29(5), 1565–1572.
- Witkin, H. A. (1950). Individual Differences in Ease of Perception of Embedded Figures. *Journal of Personality*, 19(1), 1.
- Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). *A Manual for the Embedded Figures Tests*. Consulting Psychologists Press.
- Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202–238.
- Zaharescu, A., Rothenstein, A. L., & Tsotsos, J. K. (2005). Towards a Biologically Plausible Active Visual Search Model. In L. Paletta, J. K. Tsotsos, E. Rome, & G. Humphreys (Eds.), *Attention and Performance in Computational Vision* (pp. 133–147). Springer Berlin Heidelberg.
- Zelinsky, G. J. (2008). A Theory of Eye Movements during Target Acquisition. *Psychological Review*, 115(4), 787–835.