

Towards Accurate and Practical Predictive Models of Active-Vision-Based Visual Search

David E. Kieras

Electrical Engineering & Computer Science
University of Michigan
Ann Arbor, MI 48109 USA
kieras@umich.edu

Anthony J. Hornof

Computer and Information Science
University of Oregon
Eugene, OR 97403 USA
hornof@cs.uoregon.edu

ABSTRACT

Being able to predict the performance of interface designs using models of human cognition and performance is a long-standing goal of HCI research. This paper presents recent advances in cognitive modeling which permit increasingly realistic and accurate predictions for visual human-computer interaction tasks such as icon search by incorporating an “active vision” approach which emphasizes eye movements to visual features based on the availability of features in relationship to the point of gaze. A high fidelity model of a classic visual search task demonstrates the value of incorporating visual acuity functions into models of visual performance. The features captured by the high-fidelity model are then used to formulate a model simple enough for practical use, which is then implemented in an easy-to-use GLEAN modeling tool. Easy-to-use predictive models for complex visual search tasks are thus feasible and should be further developed.

Author Keywords

Cognitive architecture; GOMS; human performance modeling; visual acuity; visual search.

ACM Classification Keywords

H.5.2. [Information Interfaces and Presentation]: User Interfaces - evaluation/methodology, screen design, theory and methods; H.1.2 [Models and Principles]: User/Machine Systems – Human factors; Human information processing.

INTRODUCTION

Visual search pervades everyday computer usage, such as the simple task of finding an icon on a smartphone. Icons that are always in the same location on the home screen and used often will be found quickly, but icons that are used occasionally or that change their position often as new applications are added will require visual search. Designing systems that support faster visual search will improve usability.

Figure 1 shows four pairs of icons from the Apple iPhone. Each pair shows two icons used for the same application, an



Figure 1. A set of icons from the iPhone interface.

older version (on the top or left of each pair) and a recent redesign (on the bottom or right of each pair). Each redesign does a better job of emphasizing the primary visual features of color, shape, and size (rather than detailed visual information) and is thus easier to distinguish than the older version in the near periphery. This can be seen by trying to describe each icon while holding your gaze fixed on the crosshairs. Vision psychology has long understood that, for objects that are away from the point-of-gaze, the primary visual features such as color, shape and size can be more easily perceived than the detailed features, and that primary features are thus very effective for guiding the eyes in visual search [30]. But not all visual designs utilize these human abilities.

Graphic designers discuss visual designs such as the icons in Figure 1 using terms such as “clutter” versus “simplicity” [22] perhaps without realizing that in part what they are discussing is which visual features can be perceived at a shorter *eccentricity* (angular distance from the point of gaze), which will in turn lead to more efficient eye movements. Interface design can benefit when designers, including anyone designing a web page, understand fundamental human performance abilities, but this is often not practical. For as long as human-computer interaction has been studied, researchers have been working to address this problem by developing theoretical models and design tools that simulate people using computer interfaces to do tasks, so that these simulations can stand in as real users early in the design cycle to help guide a designer to a usable interface that will fit well with

© 2014 Association for Computing Machinery. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of the United States government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada
Copyright © 2014 ACM 978-1-4503-2473-1/14/04...\$15.00.
<http://dx.doi.org/10.1145/2556288.2557324>

fundamental human perceptual, cognitive, memory, and motor abilities [4, 27].

Some cognitive modeling research focuses on developing the cognitive architectures (the software frameworks for building the models) and identifying appropriate parameters and settings for general classes of models, such as [7, 9, 20, 27, 28]. Other work aims more at making the modeling easier to use for everyday designers, such as GLEAN [21], Distract-R [25], and perhaps most notably the CogTool project [2, 12, 14, 27]. The work presented here contributes to both fronts, first introducing and validating an innovation that permits a cognitive architecture to more accurately simulate human vision, and then showing this innovation implemented and used within an easy-to-use modeling environment.

The modeling work presented here, partially reported previously in [16], goes beyond previous cognitive modeling of icon search (such as [7]) by incorporating a more advanced simulation of visual perception and ocular motor processing, and stands in contrast with mathematical modeling of human performance (such as [5]) in which empirically-derived relationships such as Fitts' Law are used to explain features of performance data without reference to the underlying cognitive architecture mechanisms in the human user; rather, architecture-based approaches seek to develop and apply a reusable unified theory in human-computer interaction [4, 20].

This paper presents cognitive models of a visual search task akin to icon search. First, the search task is described. Second, a high fidelity model is built using the EPIC cognitive architecture [20] to explore perceptual parameter and other settings necessary for the task. Third, the key components of the model are extracted and carried over to a lower-fidelity modeling environment, GLEAN [21], which lends itself to use by practitioners. The paper provides a specific approach for predicting visual search performance in HCI tasks that could be integrated into easy-to-use predictive analysis tools, complements and advances other approaches for predicting visual performance in HCI visual search tasks, and demonstrates the value of building these models using a cognitive architecture that distinguishes the human invariants from the task-specific knowledge.

THE TASK

This modeling work applies contemporary modeling techniques to explain a data set from the literature, in this case a very early eye movement study of visual search published by Williams in 1966 [29, 30], who ventured into experimental territory commonly avoided even today. The task involved visual search of a very large number of objects varying in size, color, and shape, each with a unique two-digit label. The 100 objects consisted of all combinations of 4 sizes, 5 colors, and 5 shapes.

Figure 2 shows a re-creation of a small portion of a search field used in the experiment. The entire display is $39^\circ \times 39^\circ$ of visual angle. Search objects ranged from 0.8° to 2.8° in size and were randomly distributed across the field. Participants were precued with the numerical label of the

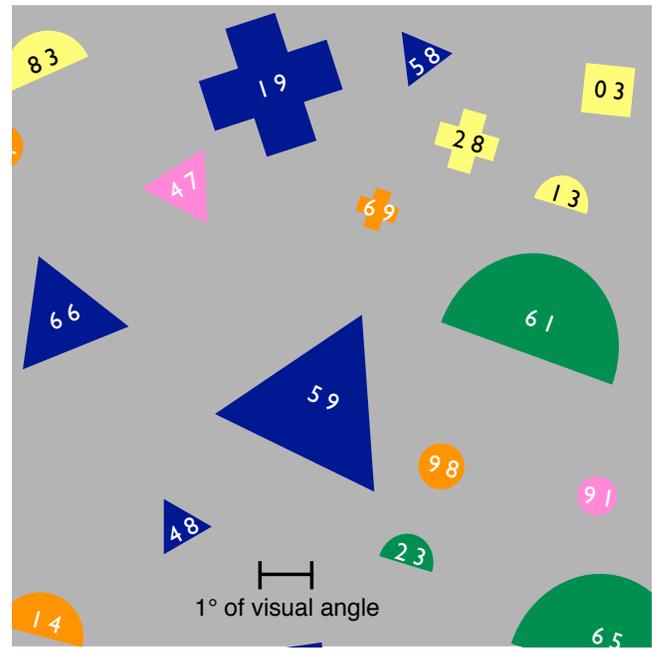


Figure 2. A reproduction of a portion of the display used in the Williams (1966) visual search experiment. The user's task was to search for an icon given a number and some combination of color, shape, and size. For example, for the trial shown here, the precue might be "small green 23".

target. The key to the experimental design is this: On most trials, participant were also precued with some combination of the color, size, and shape features of the target. For example, in Figure 2, if the target is the large triangle at the center, the precue could be just "59" (the number-only precue) or "blue 59" (the color-only condition) or "very large triangle 59" (the size and shape condition). All combinations of size, color, and shape precues were tested in addition to the number-only precue. The specifications appeared first in the center of the display; when a button was pressed, the search objects were added to the display. The participant pressed another button when he or she had located the specified object.

Eye movements were recorded with a corneal-reflection film camera system and scored by hand. The total number of fixations were counted, and classified by whether they fell on objects whose size, color, and shape matched the specifications. The hypothesis is that if a specification is an effective precue for visual guidance, more fixations should be on objects matching the precue than expected by chance. Williams approximated reaction time (RT) measures by counting the number of fixations and dividing by 3.25, the observed average number of fixations per second.

The Observed Data

Figure 3 shows the average number of fixations that participants needed to find the target for each of the eight precue conditions. The most pronounced effect is that the target can be found with the least number of fixations any time that color is specified in the precue. Specifying size helps less to reduce the number of fixations. Specifying shape helps only a little.

Figure 4 shows the proportion of fixations that landed on objects that had a feature specified in the precue, for each of the eight precue types. This shows the most interesting phenomenon captured by the Williams experiment, that during visual search participants tended to look at objects with whatever features were included in the precue. They especially did this for color. When no features were specified in the number-only condition, there was a chance likelihood that fixations landed on the correct color, size, or shape (which was respectively 20%, 25%, and 20%). When color was specified, roughly 60% of the fixations were on objects with the target color. Specifying the size also helped participants to increase their fixations on objects of the target size. Specifying the shape did not help much. Fewer fixations were needed because the fixations that were made were more likely to land on the target.

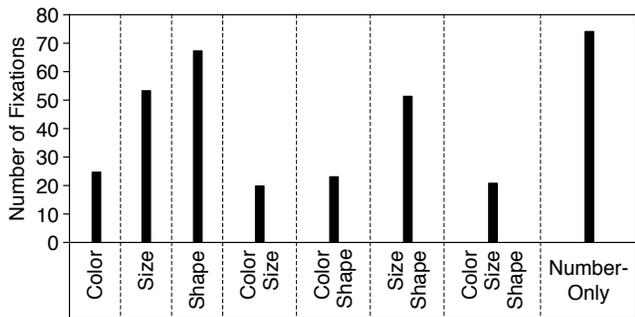


Figure 3. The average number of fixations observed for each precue type from the Williams experiment [29, 30].

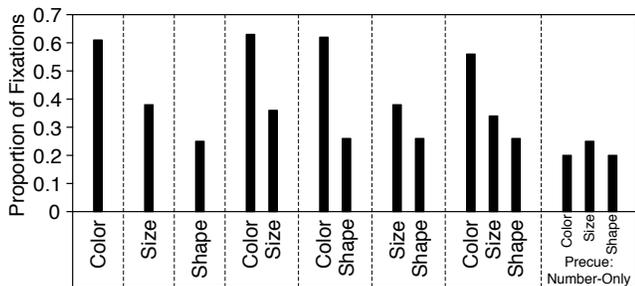


Figure 4. The proportion of fixations landing on objects with the features specified in the precue, for the eight precue types.

It is clear from the results that color is the strongest cue for visual guidance, resulting in the highest proportion of fixations on matching objects (about 0.6) and the fewest fixations (about 20). Size comes next, and shape is a distant third. There is a tendency for each precue to have a smaller effect if a stronger precue is also present. If only the label is provided (the number-only precue), the fixations on objects that match the target properties is at chance level, and the number of fixations is large (about 72).

The data reveal a high number of repeat fixations. If every object that was examined received exactly one fixation, on average half of all candidate objects would be examined. For example, in the number-only condition, on average fifty objects would be fixated. This implies some form of inefficiency or unreliability in the search process, wherein objects are looked at repeatedly and at a very high rate. For example, the 72 fixations in the Number-only condition

implies a repeat rate of about 30%. Williams reports a small number (3%) of immediate repeat fixations, but does not report repeat fixations over longer time periods.

In contrast, recent observation (such as Peterson et al. [24]) and modeling [15] of repeat fixations suggests that repeat fixations are relatively rare, around 5%, implying a good memory for previous fixations, and almost all are performed immediately, being due to recognition (encoding) failures rather than failures of the memory for previous fixations. However, these results were obtained in search tasks involving many fewer objects and that took much less time than Williams' task. Perhaps the much higher repeat rate in these results is due to time decay of the fixation memory. In fact, in Peterson's task, repeat fixations at long lags become more frequent if the trial has gone on for an usually long time [Peterson, personal communication]. This issue will be important in modeling the Williams data.

Williams did not report confidence intervals nor apply conventional statistical analyses for what is basically a descriptive study. However, the reported details imply that there are many thousands of fixations underlying the reported proportions, so their binomial confidence intervals would be almost invisibly small on these graphs. Thus we can take the data as being very reliable statistically.

THE MODELING APPROACH

Active Vision

The modeling approach used here closely follows the notion of *active vision* which is argued eloquently by Findlay and Gilchrist [6] and which characterizes visual behavior in terms of the eye movements that are made to reorient the high-resolution vision at the center of the gaze to different items of interest. Active vision is markedly different from traditional approaches to visual attention which have ignored both the role of eye movements and extra-foveal information (information within roughly 2° to 10° of the point of gaze) and which have relied primarily on RT rather than eye movements as the primary dependent measure.

A key process in any visual search task is choosing the next object for inspection. A variety of studies (see [6] for a review) have shown that this choice is strongly influenced by the color, shape, size, orientation, and other visual properties of nearby objects. This phenomenon is called *visual guidance*. In the active vision framework, these properties are available in extra-foveal or peripheral vision to some extent, meaning that visual attention, which in the context of normal visual activity is almost synonymous with where the eye is fixated, is a process of selecting for detailed examination one of a large number of objects currently perceived to be in the visual scene, and doing this selection on the basis of the visual properties available in extra-foveal vision.

Though the importance of color in visual search is well-documented [26], in the active vision framework, color is not specially privileged in some way, but rather, various

direct measurements simply show that the color of an object is visible over a wide range of eccentricity and object sizes (e.g. [8]), and so can often serve as an effective cue about where to look next. The relative ineffectiveness of shape is likewise not due to a fundamental problem with shape, but rather that in many cases recognizing the shape requires resolving detailed features that can only be seen close to the fovea. As an extreme of shape recognition, the text label involves very fine features requiring foveation unless the text is very large [1].

The EPIC Cognitive Architecture

The Williams results can be modeled using the EPIC (Executive Process-Interactive Control) architecture for human cognition and performance [19, 20]. EPIC provides a general framework for simulating a human interacting with an environment to accomplish a task, and directly supports an active vision approach to visual search as demonstrated in [9].

The EPIC architecture consists of software modules for the simulated task environment or device that interacts with a simulated human, which consists of perceptual and motor processor peripherals surrounding a cognitive processor. The device and all of the processors run in parallel with each other. To model human performance in a task, the cognitive processor is programmed with production rules that implement a strategy for performing the task. When the simulation is run, the architecture generates the specific sequence of perceptual, cognitive, and motor events required to perform the task, within the constraints determined by the architecture and the task environment.

Figure 5 shows the visual system of EPIC. The *eye processor* explicitly represents differential retinal availability in terms of *acuity functions* that specify whether each visual property (feature) of each object is currently

visible as a function of the size of the object and its eccentricity. The currently available visual properties for each object are represented in the *sensory store*; the *perceptual processor* then encodes the properties of each object, possibly in relation to other objects, and passes the encoded representation on to the *perceptual store* where they are available to the cognitive processor to match the conditions of production rules. The perceptual store thus contains the current representation of the visual world that cognition can make decisions about, including decisions about where to move the eyes next by commanding the *voluntary ocular motor processor*. The perceptual store retains the representations for all objects currently visible, with more information and detail available for objects that have been fixated.

There are two senses of *size* in this model: the physical size of the object in degrees of visual angle and its encoded size which corresponds to the precue specification from *small* to *very large*. The acuity functions use the physical size of the object to determine whether visual features (including the encoded size) is perceptually available given the object's eccentricity, and the strategy uses the encoded size to focus the search when that feature is available.

Persistence of the visual perceptual store

When the eyes move away from an object, the properties of the object persist in the sensory store for a short time (e.g. 200 ms) after which the perceptual processor notes that the property in the perceptual store no longer has sensory support. After a relatively long time, the property is removed from perceptual store. However, if the the object disappears completely, the object and its properties will be removed from the perceptual store fairly quickly.

The concept is that as the eyes move around the visual scene, a complete and continuous representation of the objects in the scene is built and maintained in the perceptual store, allowing the cognitive processor to make decisions based on far more than the properties of the currently fixated object. The notion that this information persists for a considerable time as long as the scene is present is supported by studies summarized by Henderson & Castelhamo [10] in which subjects are tested for their memory of a previously fixated object and retention times of at least several seconds long are observed.

Figure 6 illustrates how visual information flows through EPIC's physical, sensory, and perceptual stores by showing the contents of these memories at a single point in time. The left panel shows the physical store—an encoding of all objects and features in the world. The gray concentric circles at the bottom left show the current location of EPIC's eyes, currently on the yellow cross 88. The small gray circle has a 1° radius corresponding to the fovea; the larger gray circle is a calibration ring of 5° radius. This view of the physical store permits a person running a model to monitor its progress; high fidelity with the original visual display is not needed. The middle panel shows the sensory store—the objects and features that are currently visible to the eyes. Feature availability is determined by the acuity

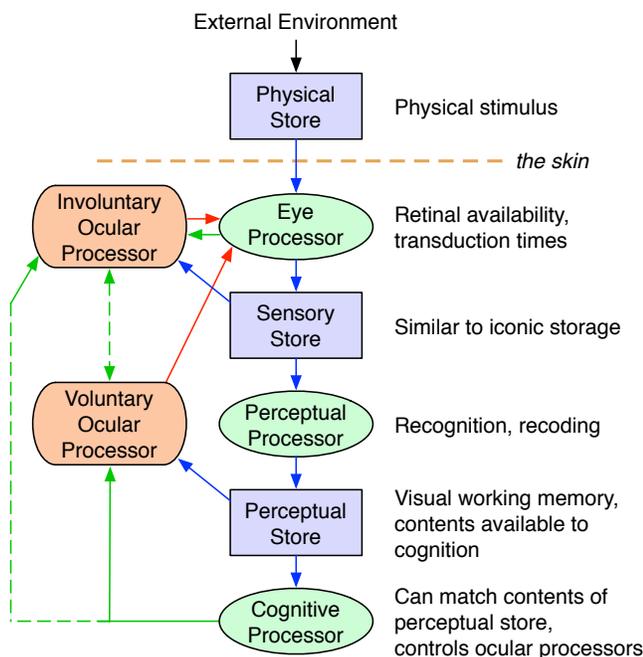


Figure 5. EPIC's visual system.

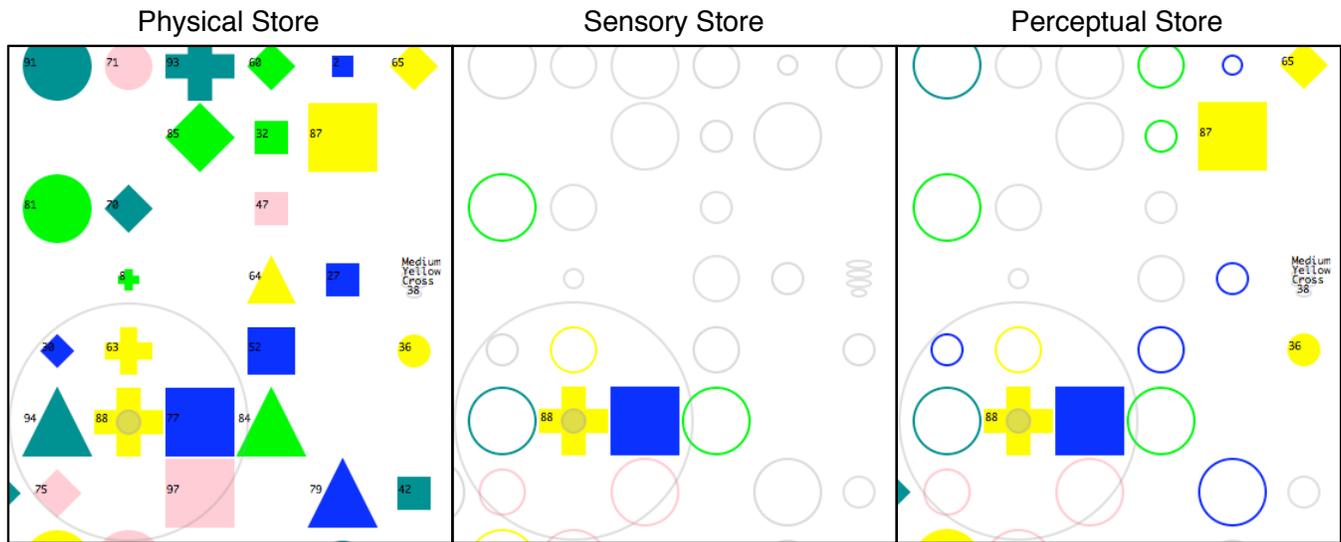


Figure 6. A snapshot of EPIC's physical, sensory, and perceptual stores. The eyes are currently fixating the yellow cross 88.

function. The right panel shows the perceptual store—the visual information that has persisted from the last few gaze fixations on the other objects whose numbers are visible.

THE EPIC MODEL

Constructing an EPIC model for the Williams task required a choice of (1) visual acuity parameters, (2) a parameter for the decay time of visual properties in the perceptual store that are no longer sensorily supported, and (3) a set of production rules that implement the visual search strategy. The strategy and the values for the parameters were varied in a manually conducted iterative search to maximize the goodness of fit to the data. Each of these model inputs will be described briefly.

Acuity functions

Visual acuity functions were specified for the size, color, shape, and text properties to model how visual properties can be recognized in peripheral vision depending on the eccentricity, size, and property involved. The functions are derived from the literature on visual perception (such as [1, 8]) which show how, in the periphery, the features of larger objects are more recognizable than the features of smaller objects. An object's color is more visible in the periphery than the object's shape, which in turn would be more visible than the object's text label whose small size would require a fixation to be recognized. Text acuity is thus specified as being available within 1° of the current eye position, corresponding to the conventional definition of foveal vision. For other features, a quadratic psychophysical acuity function determines the availability of object properties based on the eccentricity and the size of the object, with some random noise added. Thus, the size, color, and shape functions are quadratic threshold functions for object size *s* that depend on eccentricity *e* and a random noise component *X*, specified as follows:

$$\begin{aligned}
 & \text{threshold} = ae^2 + be + c \\
 & P(\text{available}) = P(s + X > \text{threshold}) \\
 & X \sim N(s, vs)
 \end{aligned}$$

Since comprehensive parametric data on extra-foveal acuity for different visual properties is lacking in the empirical literature, the parameters for these functions were determined with an iterative search in which the acuity function parameters and retention time were adjusted to maximize the goodness of fit with the observed data. Because the differences in availability are strongest at larger eccentricities and because these differences are primarily determined by the quadratic coefficient of the acuity function, the other parameters of the acuity functions were kept constant to simplify the parameter search.

Figure 7 shows the availability functions for color, size and shape that were used in the model. The functions were set with parameter values of *v*=0.7, *b*=0.1, *c*=0.1 for all three features, and the parameter of *a* was set to 0.035 for color, 0.2 for size, and 0.3 for shape. The functions converge to 0.9 at 0° to provide for an optional 10% encoding error, analogous to the encoding errors used in previous models [9, 15, 18], though this was not used in this model. Instead, an additional step function was imposed at 1° such that, like

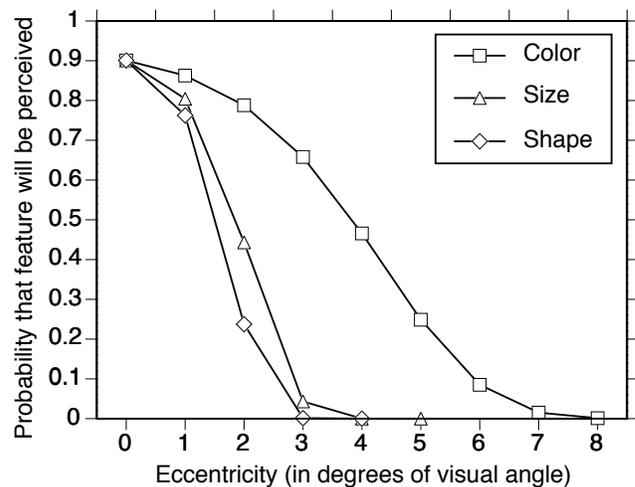


Figure 7. The acuity functions used in the EPIC model.

the text property, color, size, and shape were also fully available in the fovea.

The availability for every property of every object is recomputed whenever the eye is moved. The sensory store in Figure 6 shows what is currently available around the fixation point after several fixations. Objects whose location, but no other properties, are known are represented as light gray open circles. Objects which are close enough to the current fixation point to have their color available, but not their shape, are represented as colored open circles.

Perceptual store persistence time

Once a property of an object is visible, that property is attached to the object representation in the visual perceptual store where it can serve to match conditions of production rules. The visual perceptual store is persistent in that as long as an object is within the visual field, its properties, once acquired, will persist for up to several seconds even after the eyes move [10], and can thus serve as a memory for previous fixations [15]. The perceptual store in Figure 6 shows EPIC's perceptual store several seconds into the visual search. The duration parameter was estimated iteratively, starting with the 4 s lower bound determined in [15]. The reported results here used a value of 9 s.

Task strategy

The visual search strategy in the model is an application of a basic strategy, shown in Figure 8, that has been used in several EPIC visual search models (such as [11]). There are two threads of execution: (1) Nomination rules in the first thread propose objects to fixate based on available visual properties, and also nominate a random choice. Choice rules then pick a single candidate from the nominated objects according to a priority scheme, and launch an eye movement to the chosen candidate. The nomination thread then either starts over, or terminates if the other thread had found the target. (2) Comparison rules in the second thread wait for all relevant properties of the fixated candidate to be fully visible and either respond if it is a target, or discard the candidate if not. Given the typical 100 ms transduction and encoding times for visual properties and the 50 ms production rule cycle time, the overlapped processing provided by the two threads enables the time between successive eye movement initiations to be short, about 250 to 300 ms, which is commonly observed in high-speed visual search tasks.

For the Williams model, the strategy nominates candidate objects that have the precued properties, such as the precued color or shape. The fixation memory effect is implemented by only nominating objects whose text property is unknown—not currently present in the visual perceptual store—either because the object was never fixated or because it was fixated a long time ago and the representation has decayed. The priority scheme for choosing a fixation target was originally implemented as picking the object with the most matching properties, which required a large number of rules for each combination. However, the acuity functions dictate that if color is available, shape and size are not very relevant. So a more

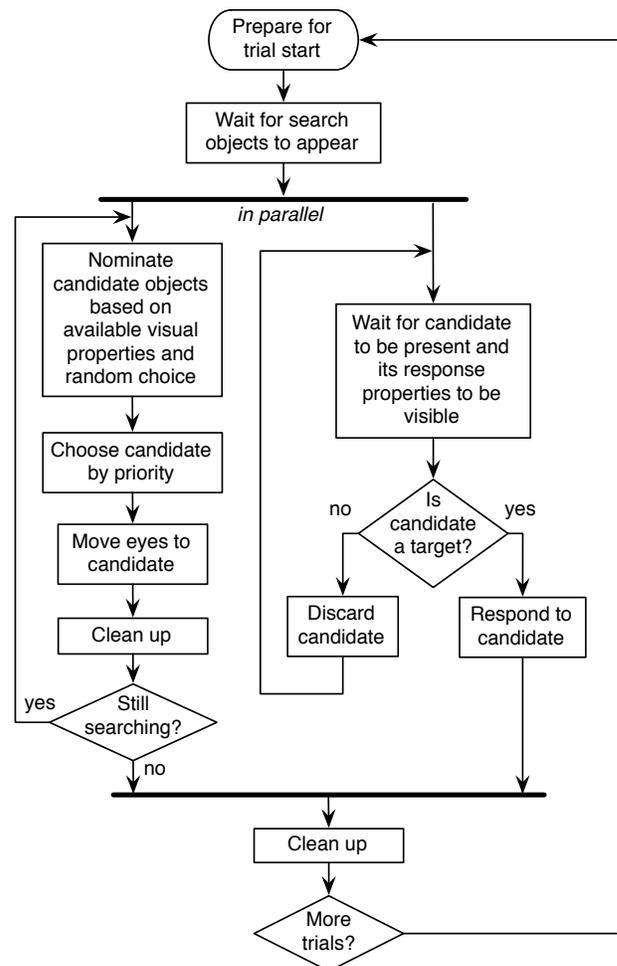


Figure 8. Flowchart for the search task strategy.

parsimonious single-property priority scheme that mirrors the relative acuities is used in this version of the model: An object with a matching color is chosen over one with a matching size over one with a matching shape. This scheme is an optimization to favor the most-available information over the-less available information.

EPIC MODEL RESULTS

The model was run for 500 trials in each experimental condition, which was determined to produce stable predicted values, and the predicted eye movement and response time data were compared to the observed data.

Figure 9 shows the predicted and observed proportion of fixations that landed on objects that had a feature specified in the precue, corresponding to Figure 4. Clearly the fit is very good using the acuity function and perceptual store persistence parameters listed above; $R^2 = .99$; average absolute error (AAE) = 3%.

Figure 10 shows the predicted and observed number of fixations. Again there is a very good fit ($R^2 = 0.98$, AAE = 12%). The observed and predicted RTs in Figure 11 fit well ($R^2 = 0.98$ and AAE = 9%), although there is a general tendency for the model RTs to run slightly longer than Williams' results. Given the unusual methodology used to

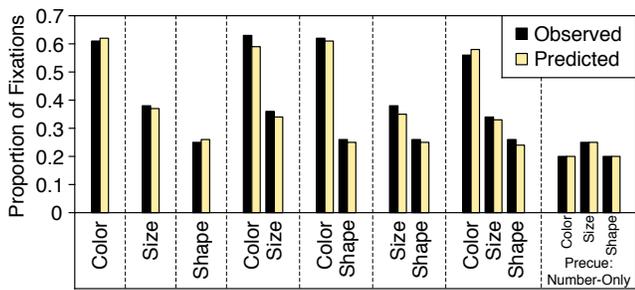


Figure 9. Observed and predicted proportion of fixations landing on objects with the features specified in the precue, for the eight precue types.

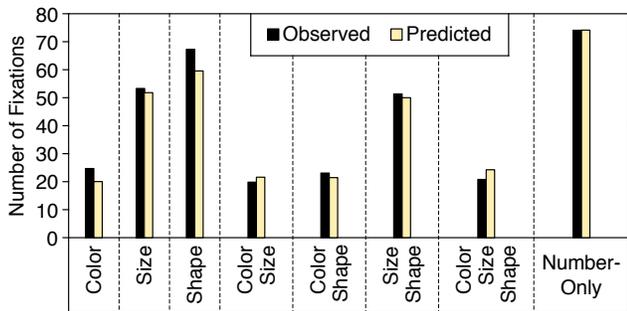


Figure 10. Observed and predicted number of fixations for each precue type.

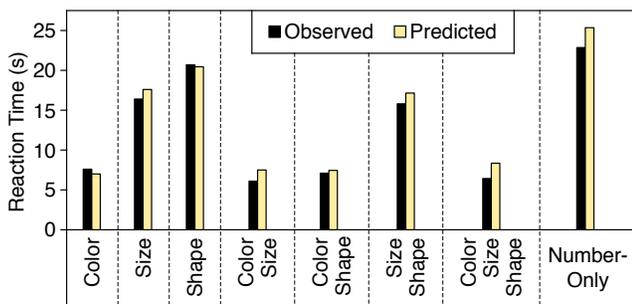


Figure 11. Observed and predicted RTs for each precue type.

measure the RTs, it is not clear that attempting to improve the fit would be worthwhile.

In an analysis of the model output, the proportion of repeat fixations was found to increase substantially as the perceptual store duration was decreased, and the number of fixations increased. The persistence parameter was adjusted to produce the overall good fit on the number of fixations shown in Figure 10, and the proportion of repeat fixations on search objects was then determined with the final parameter value. The range was 11% repeats in the best condition to 33% in the Number-only condition. This proportion was highly linear with the predicted number of fixations, with an intercept = $-.03$, slope = 0.01 , and $R^2 = 0.95$. Thus the loss of fixation memory over time accounts for the excess number of fixations in the data.

Generality of the strategy and obtained parameter values

The three inputs to the model were the acuity functions, the perceptual store persistence time, and the task strategy. Each of these contribute to the fit of the model: The task

strategy makes use of the most visible properties of the objects to guide the search efficiently and uses the persistent information in the visual store to avoid repeat visits to objects, but the effectiveness of this depends on how long the information persists. This task strategy should generalize to visual search tasks in which the visual field lacks an overall structure to guide the search, and the target has to be located based on its visual features. The strategy, acuity function parameters, and persistence time parameter should apply to similar tasks such as with radar displays using military-standard icons (such as in [23]).

But other displays would likely require different acuity parameters and a variation on the task strategy. For example, some of the complex icons used in desktop and mobile devices, such as the older icons in Figure 1, lack a dominant color. Many icons also lack a distinguishing size or shape. If no distinguishing features are visible in the periphery, then a strategy that uses such features would be less useful, and performance might be more like the number-only condition in this study.

High-Fidelity Model Success

The success of this EPIC model to account for the Williams data means that a very demanding search task can be described with an architecture and model based on active vision principles—different visual properties are differentially effective in visual search primarily because they have different visual acuity characteristics. In addition, the capacious but limited memory for previous fixations will fail more often if the search task goes on for a long time, resulting in inefficient repeat fixations on objects. All these factors can be explained with the EPIC model, using architectural features, parameter values, and strategies similar to those used to model other search tasks.

THE CHALLENGE FOR PRACTICAL MODELING

Applying models of human performance in the evaluation of interface designs has been an important goal for HCI theory since Card, Moran, and Newell's seminal work [4, 17]. However, there is a difficult tradeoff between the high-fidelity models like EPIC that are central to the research effort of modeling human performance, and practical low-fidelity easy-to-use models necessary for HCI application. GOMS modeling is the most common example of such a practical approach, and has a track record of model successes that are both scientifically valuable and practically useful [13, 14]. The key important characteristic of these practical model techniques is *ease of model construction* which requires first, a simple modeling language (as in GOMS) that is far easier than production rules, and second, that the details of psychological processes must be thoroughly encapsulated in the modeling system [17]. So can visual search times be predicted to a useful degree with a very simple model?

Guidelines for Low-Fidelity Prediction of Visual Search

The empirical literature on visual search, together with the modeling results of this task and others in the EPIC framework, identifies several characteristics of visual search that can serve as guidelines for a simple and

practical method for predicting visual search. These guidelines are as follows:

1. *In general, sample objects without replacement so that they are seldom revisited.* The empirical results, explained by the model mechanisms of encoding failures and memory for previously fixated objects, is that in small search tasks completed in a few seconds, the revisit rate is about 5%. The model results for this task, which could take a much longer time, produced a higher revisit rate, about 11-33%, mostly due to memory failures.
2. *Treat the color feature as widely available if the objects are reasonably large and the color is prominent,* such as for objects 1° or larger with a single color.
3. *Treat other object properties such as encoded size and shape as narrowly available except for very large objects.* Generally, the object must be fixated for these properties to be visible.
4. *In general, text must be fixated to be recognized.*
5. *Fixate adjacent items only if their available visible properties have not ruled them out as candidates.* While a common observation is that the closest relevant object tends to be fixated next, this could simply be due to the shape of acuity functions resulting in closer objects being more likely to have their visual guidance properties available.
5. *Set each fixation time to a constant in the range of 250 to 300 ms.*

An Approximate Algorithm for Predicting Visual Search

These guidelines for predicting visual search can be combined into a simple algorithmic model of visual search such as the following:

1. Start with a list of all the to-be-searched objects in the visual field each marked as unknown color and unvisited.
2. Set n , the number of fixations, to zero.
3. Repeat the following:
 - If a color is specified:
 - For each object, use the availability function to determine whether its color is available. If it is, add the color to that object in the list.
 - If the specified color is available for one or more unvisited objects in the list, choose one of those objects at random.
 - If the specified color is not available for an unvisited object, choose an object with unknown color at random.
 - If a color is not specified, choose an unvisited object at random.
 - Make the chosen object the current object, make its location the current eye location, mark it as visited, and increment n . If it is not the target object, repeat with the new eye location. If it is, the search is done.
4. Estimate the proportion of repeat fixations with the formula $(n * .01 - .03)$, floored at 0. Add this proportion to n , and multiply by 300 ms/fixation to yield the approximate search time.

The GLEAN Model

A GOMS model of the Williams task was implemented using the GLEAN cognitive modeling system [21]. GLEAN is a simulation environment similar to EPIC, but with a much simpler cognitive architecture directly inspired by the Card, Moran, and Newell [4] Model Human Processor, and whose cognitive processor is directly programmed in terms of procedural GOMS models using GOMSL, a formalized version of the earlier NGOMSL notation. GOMSL includes a visual search operator `Look_for` which takes a list of property value pairs (e.g. `Color is Red`) and examines the content of the visual processor's store of visual objects, locates the first object that has the specified property-value pairs, and stores the name of the object in working memory under a supplied tag. This tag can then be used to identify the object for later operators, such as pointing a mouse to the object. In the GOMS tradition, the time required for the `Look_for` operator is a single *Mental* operator with a time constant of 1.2 or 1.35 s [14].

When expressed in GOMSL, the complete model for the Williams task is very simple, as shown in Figure 12. This version uses all four properties to identify the target object; if fewer properties are specified, the `Look_for` operator would use only the specified properties. For brevity, the `Acquire Probe` method is not shown; it simply examines each field of the probe and stores the size, color, shape, and label under the corresponding tags. The brevity and simplicity of this model compared to the intricacy of the EPIC or other full-fledged cognitive architecture models is striking. This simplicity is the principal argument for trying to develop computational architectures such as GLEAN that are based on GOMS [17].

Two GLEAN models were run. The two models were identical to each other from the perspective of the analyst (the person) using GLEAN—both used the exact same GOMSL code shown in Figure 12. The first model used the current GLEAN system in which every `Look_for` operator is assigned a constant time of 1.2 s. The second model used

```
Define_model: "Williams 67 task"
Starting_goal is Perform Search_task.

Method_for_goal: Perform Search_task
Step 1. Accomplish_goal: Perform Trial.
Step 2. Goto 1.

Method_for_goal: Perform Trial
Step 1. Accomplish_goal: Acquire Probe.
Step 2. Keystroke X.
Step 3. Wait_for_visual_object_whose
Event_type is New
and_store_under <new_object>.
Step 4. Look_for_object_whose
Color is <color>, Encoded_size is <size>,
Shape is <shape>, Text is <label>
and_store_under <found_object>.
Step 5. Keystroke X.
Step 6. Return_with_goal_accomplished.
```

Figure 12. The GOMSL model for the Williams task when all precue properties are specified.

an augmented version of the GLEAN system in which the time required for a `Look_for` operator was determined using an implementation of the “approximate algorithm” described above, and in which the same acuity function described for the EPIC model was used to determine the availability of each object's color feature.

GLEAN Model Results

The first model, using the current GLEAN system with a constant `Look_for` operator time, was run for a single iteration. (All subsequent runs would have been identical.) The model produced a short constant RT of 1.92 s regardless of the precue specifications. This RT was primarily determined by the Mental operator time regardless of the cue specifications, the number of objects, or their visual characteristics. The model is seriously inaccurate, with an R^2 of 0.0 and an overall AAE = 80%.

The second model, using the GLEAN system augmented with the enhanced `Look_for` operator and the color-only acuity function, was run for 500 trials on each condition. Figure 13 shows the model's RT predictions. The predictions are much less accurate than the EPIC model and consist of only two values—a short one for color-guided search and a long one for a random search. Because the size and shape acuity functions were not used, the model is not accurate in these conditions. Despite this oversimplification, the R^2 over this data is a respectable 0.91; the AAE for the color-guided and Number-only search is reasonably good at 11%, and overall is 26%. Despite the serious over-prediction for the non-color precues, the predictions are much more accurate than when predicting search time using the Mental operator.

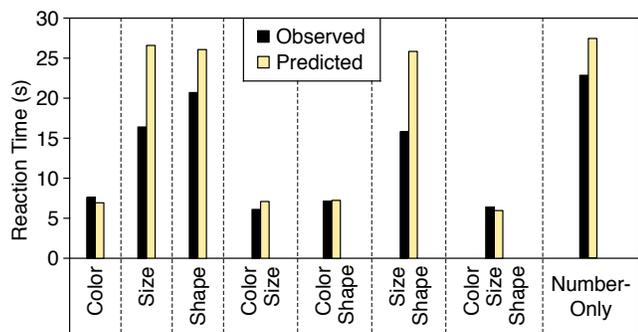


Figure 13. Observed and predicted RTs for the augmented GLEAN GOMS model.

Low-Fidelity Model Success

Even in this preliminary highly simplified form, the augmented version of the `Look_for` operator would produce much more accurate results than the original GOMS Mental operator version. Further accuracy could be obtained by working with additional data sets, and by including the object size and shape in the algorithm using representative acuity functions. The best approach to further this line of development of practical models for visual search would be to develop additional EPIC models for visual search tasks of realistic stimuli, harmonize them with the model for the Williams task, and update the algorithm

and its parameters accordingly. To avoid misleading appearances of accuracy, the GLEAN implementation would need to gracefully and visibly default to the standard Mental operator predictions when outside the bounds supported by the EPIC models.

CONCLUSION

This work presents extensions to existing cognitive modeling frameworks that permit models to more accurately characterize the processes of “active vision” in HCI tasks such as by incorporating into the modeling frameworks visual acuity functions which account for the gradual decrease in feature availability for objects as they appear at greater eccentricities from the point of gaze. These extensions will make it possible for contemporary modeling approaches to predict the benefits of simplified icons, such as in Figure 1, that can be distinguished based on peripherally-visible features rather than just fine details. These extensions can (and we believe should) be incorporated into both high fidelity modeling frameworks such as ACT-R [7] and easy-to-use modeling tools such as CogTool [27] and Distract-R [25].

This work shows that practical modeling of visually intensive tasks can be made more accurate without requiring additional effort from the analyst. It also demonstrates the benefits of decoupling the human invariants such as visual acuity functions from the encoding of the procedural skill needed to do a task—the augmented GLEAN provided a substantially better fit using the exact same encoding of procedural knowledge.

There is great value in the cognitive-architecture approach of modeling the human, task, and device each as separate entities. This will contribute to modeling frameworks that can predict performance across a wide range of existing and not-yet-imagined HCI tasks. This may be difficult to accomplish with models that characterize overall HCI performance simply as combinations of smaller-scale empirical relationships as in [5], which does not provide as clear a decomposition into the components responsible for the task performance. For example, the mediocre fit to some of the conditions shown in Figure 13 is easy to explain—the architecture simply does not make use of the possible extra-foveal availability of size or shape cues. Adding these would be straightforward and would improve the fit.

Cognitive modeling is well within reach of most interface designers as demonstrated by both unpublished and published [2, 3, 12, 13] successes and as supported by efforts to make such models not only easy-to-generate [14, 21, 25] but also increasingly flexible and veridical, such as with the work presented here which takes us further along the road towards accurate and practical predictive models of active-vision-based visual search.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research under grants N00014-10-1-0152 and N00014-13-1-0358 to the first author.

REFERENCES

1. Anstis, S. M. (1974). A chart demonstrating variations in acuity with retinal position. *Vision Research*, 14(7), 589-592.
2. Bellamy, R., John, B., & Kogan, S. (2011). Deploying CogTool: Integrating quantitative usability assessment into real-world software development. In *Proceedings of the 2011 International Conference on Software Engineering*, 691-700.
3. Callander, M., & Zorman, L. (2007). Usability on patrol. In *CHI '07 Extended Abstracts*, 1709-1714.
4. Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
5. Cockburn, A., & Gutwin, C. (2009). A predictive model of human performance with scrolling and hierarchical lists. *Human-Computer Interaction*, 24(3), 273-314.
6. Findlay, J. M., & Gilchrist, I. D. (2003). *Active Vision: The Psychology of Looking and Seeing*. Oxford University Press.
7. Fleetwood, M. D., & Byrne, M. D. (2006). Modeling the visual search of displays: A revised ACT-R/PM model of icon search based on eye tracking data. *Human-Computer Interaction*, 21(2), 153-197.
8. Gordon, J., & Abramov, I. (1977). Color vision in the peripheral retina. II. Hue and saturation. *J. Opt. Soc. Am.*, 67(2), 202-207.
9. Halverson, T., & Hornof, A. J. (2011). A computational model of "active vision" for visual search in human-computer interaction. *Human-Computer Interaction*, 26(4), 285-314.
10. Henderson, J. M., & Castelano, M. S. (2005). Eye movements and visual memory for scenes. In G. Underwood (Ed.), *Cognitive Processes in Eye Guidance*. (pp. 213-35). Oxford University Press.
11. Hornof, A. J., & Kieras, D. E. (1997). Cognitive modeling reveals menu search is both random and systematic. In *Proceedings of CHI '97*, 107-114.
12. John, B., Starr, S., & Utesch, B. (2012). Experiences with collaborative, distributed predictive human performance modeling. In *CHI '12 Ext. Abs.*, 437-452.
13. John, B. E., & Kieras, D. E. (1996). Using GOMS for user interface design and evaluation: Which technique? *ACM Transactions on Computer-Human Interaction*, 3(4), 287-319.
14. John, B. E., Prevas, K., Salvucci, D. D., & Koedinger, K. (2004). Predictive human performance modeling made easy. In *Proceedings of CHI 2004*, 455-462.
15. Kieras, D. (2009). The persistent visual store as the locus of fixation memory in visual search tasks. In R. Cooper & D. Peebles (Eds.), *Proceedings of the Ninth International Conference on Cognitive Modeling*, UK.
16. Kieras, D. (2010). Modeling visual search of displays of many objects: The role of differential acuity and fixation memory. In D. D. Salvucci & G. Gunzelmann (Eds.) *Proceedings of the 10th International Conference on Cognitive Modeling*, August 6-8, 2010, Philadelphia.
17. Kieras, D. E. (2007). Model-based evaluation. In J. A. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook* (2nd ed.). (pp. 1191-208). Mahwah, NJ: Lawrence Erlbaum Associates.
18. Kieras, D. E., & Marshall, S. P. (2006). Visual availability and fixation memory in modeling visual search using the EPIC architecture. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 423-428.
19. Kieras, D. E. (2004). *The EPIC Architecture: Principles of Operation*. Retrieved from <http://www.eecs.umich.edu/~kieras/docs/EPIC/EPICPrinOp.pdf>
20. Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12(4), 391-438.
21. Kieras, D. E., Wood, S. D., Abotel, K., & Hornof, A. (1995). GLEAN: A computer-based tool for rapid GOMS model usability evaluation of user interface designs. In *Proceedings of UIST '95*, ACM, 91-100.
22. Maeda, J. (2013, June 12). The future of design is more than making Apple iOS flat. *Wired Magazine*. Retrieved from <http://www.wired.com/opinion/2013/06/the-future-of-design-is-more-than-making-apple-ios-flat/>
23. Osga, G., Van Orden K., Campbell, N., Kellmeyer, D., and Lulue D. (2002). *Design and Evaluation of Warfighter Task Support Methods in a Multi-Modal Watchstation*. Space & Naval Warfare Center, San Diego, Technical Report 1874.
24. 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.
25. Salvucci, D. D., Zuber, M., Beregovaia, E., & Markley, D. (2005). Distract-R: Rapid prototyping and evaluation of in-vehicle interfaces. In *Proc. of CHI '05*, 581-589.
26. Sanders, M. S., & McCormick, E. J. (1987). *Human Factors in Engineering and Design* (6th ed.). New York, New York: McGraw-Hill.
27. Teo, John, & Blackmon. (2012). CogTool-Explorer: A model of goal-directed user exploration that considers information layout. In *Proc. of CHI '12*, 2479-2488.
28. Vera, A. H., John, B. E., Remington, R., Matessa, M., & Freed, M. A. (2005). Automating human-performance modeling at the millisecond level. *Human-Computer Interaction*, 20, 225-265.
29. Williams, L. G. (1966). *A Study Of Visual Search Using Eye Movement Recordings*. Technical Report, Honeywell Inc., St. Paul Minnesota, Feb. 28, 1966. NTIS AD0629624.
30. Williams, L. G. (1967). The effects of target specification on objects fixated during visual search. *Acta Psychologica*, 27, 355-360.