The Persistent Visual Store as the Locus of Fixation Memory in Visual Search Tasks

David Kieras (kieras@umich.edu)
Electrical Engineering & Computer Science Department, University of Michigan
2260 Hayward Street, Ann Arbor, Michigan 48109 USA

Kieras, D. (2009). The persistent visual store as the locus of fixation memory in visual search tasks. In A. Howes, D. Peebles, R. Cooper (Eds.), *9th International Conference on Cognitive Modeling – ICCM2009*, Manchester, UK.

Abstract

Experiments on visual search have demonstrated the existence of a relatively large and reliable memory for which objects have been fixated; an indication of this memory is that revisits (fixations on previously fixated objects) typically comprise only about 5% of fixations. Any cognitive architecture that supports visual search must account for where such memory resides in the system and how it can be used to guide eye movements in visual search. This paper presents a simple solution for the EPIC architecture that is consistent with the overall requirements for modeling visually-intensive tasks and other visual memory phenomena.

Keywords: visual search; cognitive modeling; eye movements.

Introduction

Many everyday and work activities involve visual search, the process of visually scanning or inspecting the environment to locate an object of interest that will then be the target of further activity. Many human-computer interaction tasks require such visual search to be made in a visual environment that is must simpler than natural scenes. For example, a particular icon coded by color, shape, and other attributes must be located on a screen and then clicked on using a mouse. This domain combines relative simplicity of the visual characteristics of the searched-for objects with practical relevance: the task is a natural one in the sense that such activities are very common in current technology. Visual search is so heavily relied on in many computer-based systems that it probably is a major bottleneck in system performance. Thus understanding in detail how visual search works in such domains can lead to better system designs. In addition, if visual search can be understood in the context of a comprehensive computational cognitive architecture, then it will add to our knowledge of human perception, cognition, and action in the especially rigorous and coherent way characteristic of computational cognitive architectural modeling.

Visual Search and Active Vision

In a laboratory visual search task, a display of objects is presented, and the participant must locate a particular object specified by perceptual properties and make a response based on whether such an object is present or exactly which properties it has (e.g. the specific shape). In most experiments, the display is static and contains some number of objects, only one of which is the target that must be responded to; the others are distractors. The properties of the display or the displayed objects are manipulated, and reaction time (RT) and/or eye movements are measured. The empirical literature on this task was dominated for a long time by studies that measured only RT, and often for tachistoscopically presented displays that ruled out eye movements, but more recently the cost of eye movement data collection has decreased to the point that it has become much more common, and deservedly so. While any behavioral measurement only indirectly reflects the mental processes that produce it, RT is clearly much less diagnostic of what goes on during visual search than eye movements. Furthermore, tasks in which the eye is free to move about a static display in a naturalistic manner, typical of eye movement studies of visual search, will be more representative of the normal operation of the visual system and the role of attention in visual activity. This point was argued eloquently by Findlay and Gilchrist (2003) in presenting an *active vision* framework for understanding visual activity; it is markedly different from traditional approaches to visual attention which have ignored both the role of eye movements and extra-foveal information.

A key process in visual search is choosing the next object for inspection. A variety of studies (see Findlay & Gilchrist, 2003, for a review) have shown that this choice is not at all random; rather the color, shape, size, orientation, or other visual properties of objects influences which object is chosen for the next fixation; the phenomenon is called *visual guidance* or *eye 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.

Fixation Memory

An important fact about visual guidance in visual search tasks is that an object that was previously fixated will be only rarely selected for a new fixation. This is an old result in eye movement studies (e.g. Barbur, Forsyth, & Wooding, 1993), but it did not receive much attention until the controversial Horowitz and Wolfe (1998) claim that "Visual search has no memory." They compared search RTs of a static display with a changing display, in which the objects changed positions during search, and found no difference in RT. If the visual search mechanism remembered where it had already inspected, it should be disrupted if the objects changed location; the RT being unaffected argues that the search was not disrupted, which means in turn that there was no memory for the previous fixations. Peterson, Kramer, Ranxiao, Irwin, and McCarley (2001) countered with a study demonstrating that "Visual search has memory". They recorded eye movements during search of a static display,

and discovered, as earlier studies had noted, that revisits were rare, meaning that the previous fixations were remembered in some way.

Encoding failures trigger revisits. Peterson et al. went further with a detailed analysis showing that most revisits were made immediately after only one intervening fixation, which rules out memory failure as the cause of a revisit. Rather, Peterson et al. proposed that revisits were due to *encoding failures*: soon after fixating an object and moving on to the next, the person would realize that the previous object had not been fully encoded, and so would revisit it. Using a Monte-Carlo model, they demonstrated that this explanation accounted for the statistical structure of the revisits considerably better than either a no-memory or memory-failure model.

Search strategies dominate. Several subsequent studies (e.g. von Mühlenen, Müller & Müller, 2003; Geyer, von Mühlenen, & Müller, 2006) using RT, eye tracking, and changing displays make it clear that the Horowitz and Wolfe results were an artifact of how the changing displays would force a change in task strategy. If the display is changing, the only way to perform the task successfully is use a strategy that compensates, such as to "wait and see" whether the target appears in a subset of the display. In other words, the changing display paradigm forces a strategy that produces a no-memory effect. Regardless of the methodological issues and the merits of the results, an important implication is that making use of memory for previous fixations is not "hard-wired" in the visual system, e.g. at the oculomotor level, but rather is an optional feature of a visual search task strategy.

Objects, not locations. Additional studies (e.g. Beck, Peterson, & Vomela, 2006) have attempted to determine whether what is remembered on each fixation is the location, the identity, or the properties of the objects. However, it should be clear that in a changing-display paradigm, if objects are identified in terms of their properties (e.g. shape), then they are "teleporting" from one location to the next, which is not a natural input to the visual system. Hulleman (2009) performed the most elegant and naturalistic test of whether fixation location was remembered simply by having the objects move around on the display during search similar to the Pylyshyn & Storm (1988) multiple object tracking paradigm. He observed almost no difference in search rates compared to a static display. This strongly suggests that fixation locations themselves were not remembered, since the objects were continuously changing location. The conclusion would seem to be that previously fixated *objects* are being remembered, where object identity persists over changes in location. In a naturally static display, such as the Peterson et al. (2001) paradigm, the issue does not arise: objects retain their location and properties.

Large capacity. The consensus of the empirical literature at this point is that memory for previous fixations exists. Moreover, it has a fairly large effective capacity. The Peterson et al. study involved twelve objects, half of which would have to be visited on the average. Results described in Kieras and Marshall (2006) involved 48 objects for two targets, with low revisit rates. Takeda (2004) estimated the capacity as high as 20 objects. This effective capacity is much more than the typical estimates for working memory, and so-called visual working memory in particular (e.g. Luck & Vogel, 1997) which has been estimated as holding only about four objects in a change-detection paradigm.

The locus puzzle. From the point of view of cognitive architecture, this result presents a serious quandary. Where is this capacious and reliable memory situated, and how does it work? Is it a special-purpose memory, or is it simply a by-product of some other memory function? These questions were addressed as part of program of detailed quantitative modeling of visual search tasks using the EPIC architecture, which was developed to represent perceptual-motor constraints on performance as fully as cognitive constraints, and so is well-suited to the goal. This work with EPIC visual search models focussed on representing how multiple stimulus attributes could guide visual search through conjunctive feature guidance, and how to represent their differential availability at the retinal level. These models were successful at accounting for detailed results in very simple tasks such as Findlay's (1997) first-saccade conjunctive search, searching very large displays of 100 multiattribute objects as in Williams (1967), and searching dense displays of 48 complex objects (Kieras & Marshall, 2006). However, in these models, the memory for fixations was represented in an unsatisfactory *ad hoc* manner. This paper presents a detailed model for the Peterson results to show how the fixation memory could be a side function of a memory system that is already present.

The EPIC Cognitive Architecture

The EPIC architecture for human cognition and performance provides a general framework for simulating a human interacting with an environment to accomplish a task. Due to lack of space, the reader is referred to Kieras & Meyer (1997), Meyer & Kieras (1997), or Kieras (2004) for a more complete description of EPIC. Figure 1 provides an overview of the architecture, showing perceptual and motor processor peripherals surrounding a cognitive processor; all of the processors run in parallel with each other. To model human performance of 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 2 expands the visual processor shown in Figure 1. The *eye processor* explicitly represents differential retinal availability in terms of acuity functions that specify which visual properties of objects are currently visible as a function of the current position of the eye and the size of the object. 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

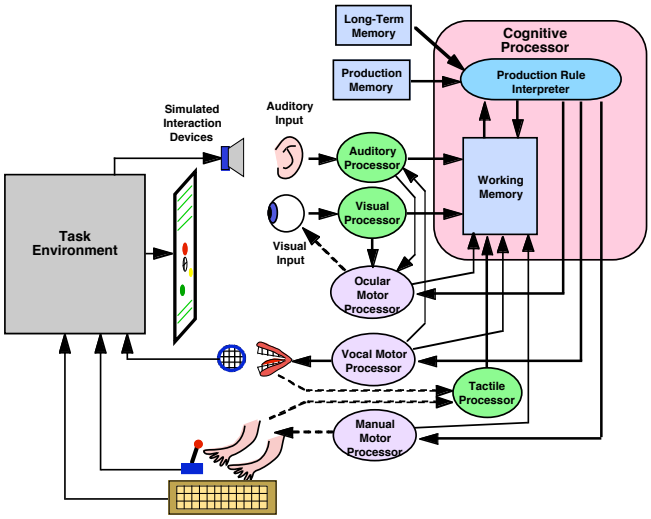

Figure 1. The overall structure of the EPIC architecture.

reason and make decisions about, especially decisions about where to move the eyes next by commanding the *ocular motor processor*. The perceptual store retains the representations for *all objects currently visible*, with more information and detail about those that have been fixated.

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

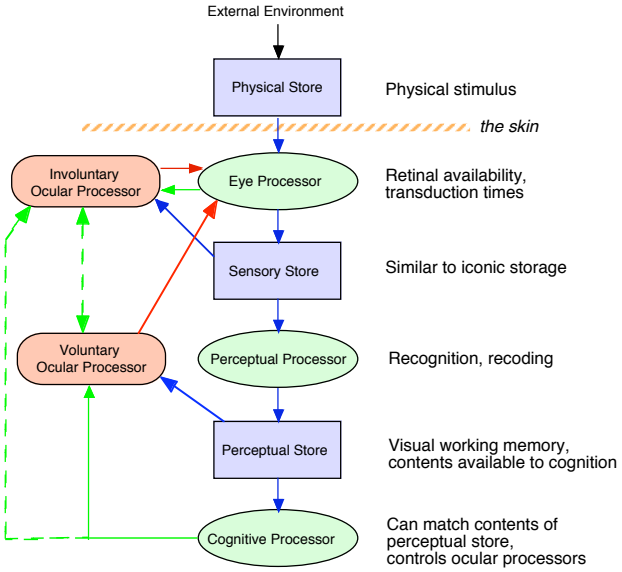

Figure 2. EPIC's visual system.

The concept is that as the eyes move around the visual scene, a complete and continuous representation of the objects currently present in the visual situation will be built up 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 in supported by studies summarized by Henderson & Castelhano (2005) in which a visual scene is continuously present, but using a gaze-contingent forced-choice paradigm, subjects are tested for their memory of a previously fixated object in a naturalistic scene; retention times at least several seconds long were observed.

Modeling Fixation Memory

The earliest attempts to fit models with the EPIC architecture for visual search in several tasks determined that some kind of fixation memory is required in order to account simultaneously for basic measures such as the number of fixations, search time, and distribution of fixations on objects with different properties (e.g. Kieras & Marshall, 2006). In order to include fixation memory, these earliest models simply "tagged" each object in memory to designate that it had already been fixated and then made an occasional random fixation to produce a revisit. This is an unsatisfactory ad-hoc solution.

The model presented here examines a more interesting possibility, namely that the perceptual store, which represents the current visual scene, could serve as a memory for fixations. That is, if the object has been fixated, then its representation would include the relevant property of the object; if the object was the target, the search would stop as soon as this was determined. But if it was not, then the next object to be examined can be chosen from the set of objects currently lacking information about the property in question. Thus by choosing objects whose properties are unknown, previously fixated objects will not be revisited.

However, since the encoding of the fixated objects is not perfectly reliable, there will be occasions when a previously fixated object will be lacking the target property, and so will get visited again. This concept is the basis for the simple statistical model presented by Peterson et al. (2001); the explicit cognitive architectural model presented here provides a generalization to other visual search tasks, and in addition, clarifies some aspects of their results.

Model for the Peterson Task

Figure 3 shows the EPIC model display of the physical visual situation consisting of the stimuli for a single trial in the Peterson task after several fixations. The stimuli on each trial were twelve objects presented in random locations on a static display; eleven were distractors, consisting of rotated L-shapes, and one was the target, a T-shape rotated either to the left or to the right. The participant's task was to locate the T shape and press a key depending on whether it was the left- or right-rotated shape. Figure 3 shows how the objects were quite small, being 0.19° in visual angle size, and were widely spaced, a minimum of 4.9° apart. Participants with

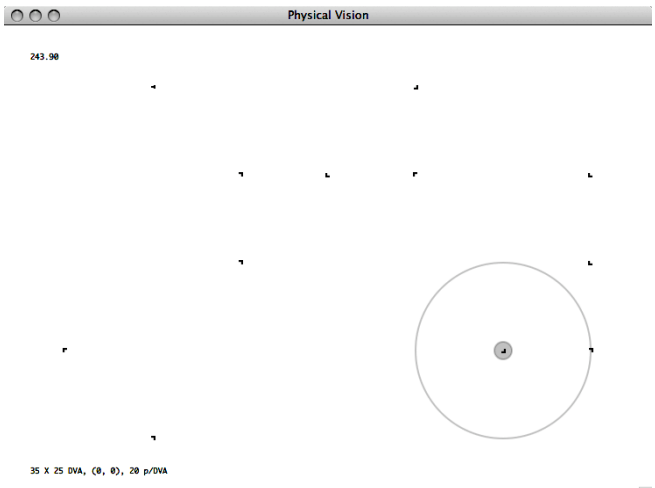

Figure 3. An example of the physical situation in a Peterson et al. (2001) task trial after several fixations as depicted in EPIC's display. The concentric circles show the current location of the eyes; the small inner circle has a 1° radius corresponding to the conventional fovea size; the outer circle is a calibration ring with 5° radius. The sizes of the overall display and the search objects are shown to scale, so the objects are indeed very small.

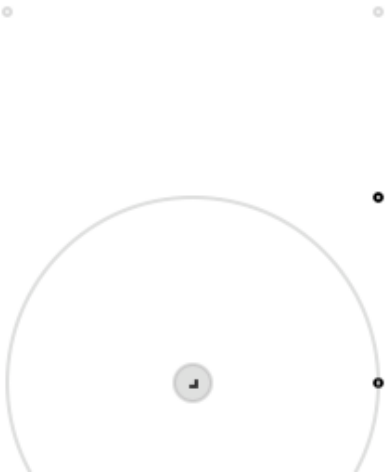

Figure 4. An example of the contents of the sensory store corresponding to the lower right corner of Figure 3. Objects whose location, but no other properties, are known are represented as light gray open circles (top two). Objects which are close enough to the current fixation point to have their black color available, but not their shape, are represented as black open circles (right hand two). Both the shape and the color are available for the currently fixated object.

normal vision would thus have to fixate each object individually to recognize it. Because of space limitations, the very small shapes are obscured in the figure.

The EPIC model to fit the data comprised a choice of (1) visual acuity parameters, (2) an encoding process in the visual perceptual processor, (3) a parameter for the encoding failure rate, (4) a parameter for the *decay time* of visual properties in the perceptual store that are no longer sensorily supported, and (5) a set of production rules that implemented the visual search strategy. Each of these model inputs will be described briefly.

(1). The visual acuity parameters for this situation are very simple, specifying that the shape of an object was available only in the fovea, while the location of an object is available throughout the visual field, meaning that any object can be selected as a fixation target. The object color plays no role in the task, but its availability was left at the default value. Figure 4 shows the effects of the acuity functions for the same display as in Figure 3.

(2, 3). The perceptual processor encodes the objects in terms of the recognized shapes for distractors and targets, which are then stored in the visual perceptual store where they become available for production rules to match on. The Peterson et al. encoding failure concept is represented as follows: with some constant probability, the encoding could fail and result in a partial encoding that retains some information about whether a distractor or target was present, but not enough to identify the actual shape. For example, a partial encoding for a distractor could be that two line segments were joined at the ends, while a partial encoding for a target could be that one line segment joined another in the middle. For purposes of display in the model, these partial encodings are represented by partially rotated L and T shapes. The probability of partial encoding of targets and distractors is assumed to be the same.

(4). After encoding, if the eye is then moved to a different object, the actual shape quickly becomes unavailable, and the encoded shape is marked as no longer having sensory support. The encoded property then disappears from the perceptual store after the time specified by the decay time. In accordance with the Henderson and Castelhano (2005)

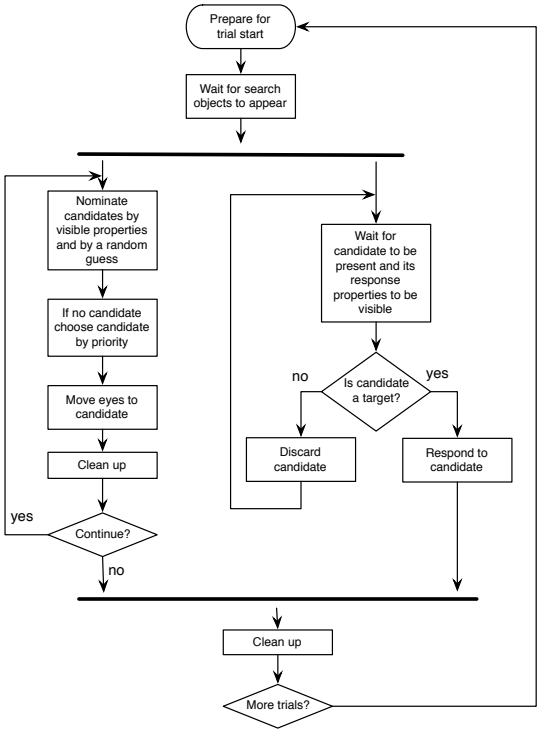

Figure 5. Flowchart for the search task strategy. Two threads overlap the process of choosing the next candidate and checking whether the current candidate is a target.

results, this parameter is assumed to be a few seconds in magnitude, though for purposes of model fitting for this data, it was made as small as possible.

(5). The visual search strategy in the model is an application of a basic strategy, shown in Figure 5, that has been used in several EPIC visual search models. There are two threads of execution. 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 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. The overlapped processing provided by the two threads enables the time between successive eye movement initiations to be short, about 250 ms, which is commonly observed in high-speed visual search tasks.

For the Peterson model, the strategy chooses objects for the next fixation according to the following simple scheme: Only objects not being currently inspected are considered. If an object is partially encoded as a target, it is given first priority for the next fixation, followed by an object not encoded as a distractor (either no encoding at all or partially encoded as a distractor), followed by an object chosen at random. Thus the strategy favors possible targets, then unvisited or partially encoded objects, and avoids objects fully known to be distractors. Figure 6 summarizes the model by showing the contents of the perceptual visual store corresponding to Figure 3, right before a target revisit.

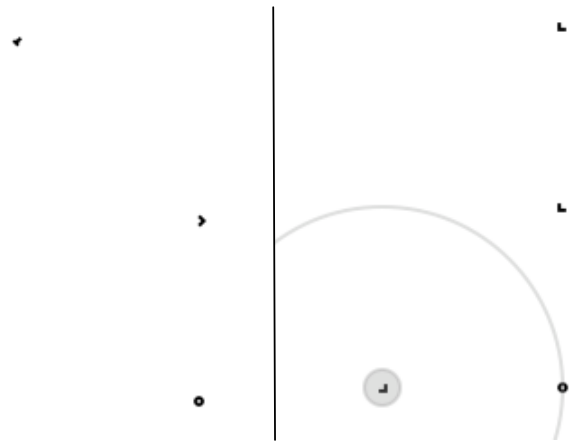

Figure 6. An example of the contents of the perceptual store after several fixations corresponding to the upper left corner (left panel) and lower right corner (right panel) of Figure 3. Two objects whose color is known to be black, but whose shape is unknown are represented as black open circles. Previously fixated objects have encoded shapes available. In the right panel, three distractors have been fixated, including the current one. In the left panel, there is a partially encoded target at the top left, and partially encoded distractor in the center right, represented as partially rotated shapes. The strategy is about to move the eyes back to the previously visited target.

Results

Figure 7 shows the observed and predicted results for this model, with the observed data from Peterson et al. (2001) shown as solid points and lines with 95% confidence intervals. The graph shows the proportion of fixations that are revisits as a function of *lag*, the number of fixations between the original and the revisit. Thus most of the revisits occur after fixating one intervening object. The total number of revisits is shown in the upper curve, and the number of revisits on targets in the lower curve.

The predicted values from the model are shown as open points and dotted lines. The model parameter values were chosen by iteration to produce a good fit with 10,000 simulation trials per run. The fit of the model predictions is very good; almost all of the predicted values are within the confidence intervals; the R^2 and standard error of prediction is 0.986 and 0.001 for Revisits, and 0.999 and 0.000 for Target Revisits. The parameter values producing this fit are 0.14 for the probability of encoding failure, and 4000 ms for the decay time of properties in the perceptual store. Any shorter decay time produces an increase in the number of predicted revisits at very long lags.

A comparison to the Peterson et al. 2001 model is useful. Although they reported the number of target revisits, they modeled only the total number of revisits, and so did not attempt to account for the fact that most of the immediate revisits are due to revisits to the target. Exploration with a variety of strategies and parameter values makes it clear that to fit both curves, the model must make the distinction between partially encoded targets and partially encoded distractors. Partially encoded targets must be favored for revisits, and partially encoded distractors treated similarly to unvisited objects — otherwise, there is no way to fit both curves simultaneously. That is, if possible targets are not favored for a revisit, then parameters that fit the overall rate of revisits far underpredict the proportion of target revisits.

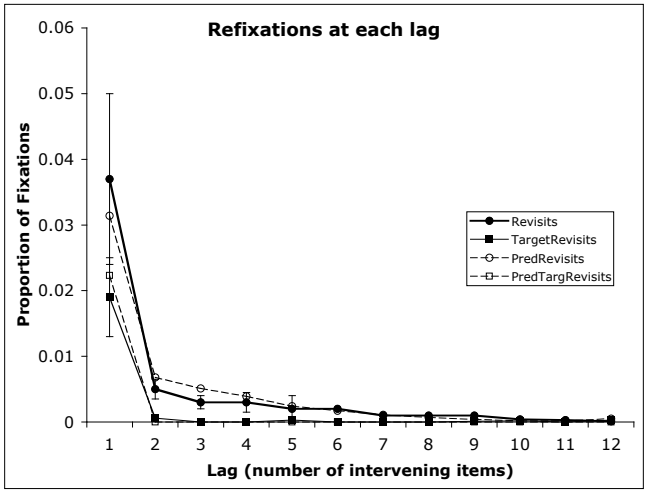

Figure 7. Solid points and lines and confidence intervals are the observed proportion of fixations at each lag for total Revisits and Target Revisits. Open points and dotted lines are model predictions.

In fact, according to the model, the revisit data for all objects are the sum of two underlying functions: Partially encoded distractors are revisited only because they are treated like unvisited distractors, yielding a shallow descent in revisits as a function of lag (imagine the total revisits curve for lags 2 to 12 extrapolated back to lag 1). But partially encoded targets are revisited immediately, producing the sharp descent from lag 1 to lag 2. The sum of these two trends produces the sharp-then-shallow curve for total revisits which was modeled by Peterson et al. The current EPIC model always revisits partially encoded targets immediately, and never favors partially encoded distractors over unvisited distractors. It might be possible to improve the fit slightly by using different encoding failure parameters for targets and distractors, and a more subtle choice strategy, but the current model fits the data acceptably well with few free parameters and a simple strategy.

Conclusion

The Peterson et al. (2001) experiment is fundamental in that it well isolates a set of basic processes underlying visual search that a successful cognitive architecture must be able to explain naturally. The present EPIC model demonstrates a how memory for fixations can emerge from the operation of a strategy for choosing the next object based on a persistent visual store of information about previously fixated objects. In this task, the only relevant properties of the objects is their location, whose wide availability makes it possible to choose an previously unvisited object for fixation, and the shape, visually available for only the one object foveated at a time. This model works by relying on the persistence of the perceptual encoding in the visual store and a simple strategy that maximizes task performance by making the most efficient use of partial encoding results.

The persistent visual store needs to be present in the architecture to allow cognition to reason about the entire visual situation. Its persistence is required for this architectural function, and is consistent with other empirical results such as those surveyed by Henderson and Castelhano (2005).

Thus the architectural puzzle posed by the existence of fixation memory can be solved by relying on this otherwise-required store; no special architectural mechanism is need to account for fixation memory. Models currently being refined for other visual search tasks (such as that described in Kieras & Marshall, 2006) show that this concept of fixation memory scales to more complex displays, objects, and search tasks.

Acknowledgment

This work was supported by the Office of Naval Research, under Grant No. N00014-06-1-0034.

References

Barbur, J.L., Forsyth, P.M., & Wooding. D.S. (1993). Eye movements and search performance. In D. Brogan, A. Gale, & K. Carr (Eds.) *Visual Search 2*. London: Taylor & Francis. 253-264.

Beck, M.R., Peterson, M.S., Vomela, M. (2006). Memory for where, but not what, is used during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 3**2**, 235-250.

Findlay, J. (1997). Saccade target selection during visual search. *Vision Research*, **37**, 617–631.

Findlay, J.M., & Gilchrist, I.D. (2003). *Active Vision.* Oxford: Oxford University Press.

Geyer, T.,von Mühlenen, A, & Müller, H.J. (2006). What do eye movements reveal about the role of memory in visual search? *Quarterly Journal of Experimental Psychology*, 60, 924-935

Henderson, J.M. & Castelhano, M.S. (2005). Eye movements and visual memory for scenes. In G. Underwood (Ed.), *Cognitive processes in eye guidance*. New York: Oxford University Press. 213-235.

Horowitz, T.S. and Wolfe, J.M. (1998). Visual search has no memory. *Nature*, 394, 575–577.

Hulleman, J. (2009). No need for inhibitory tagging of locations in visual search. Psychonomic Bulletin & Review, 2009, 16 (1), 116-120.

Kieras, D.E. (2004). EPIC Architecture Principles of Operation. Web publication available at ftp://www.eecs.umich.edu/people/kieras/EPIC/EPICPrinOp.pdf

Kieras, D.E. (2007). The control of cognition. In W. Gray (Ed.), *Integrated models of cognitive systems*. (pp. 327 - 355). Oxford University Press.

Kieras, D.E, & Marshall, S.P. (2006). Visual Availability and Fixation Memory in Modeling Visual Search using the EPIC Architecture. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 423-428.

Kieras, D. & Meyer, D.E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12, 391-438.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, **104**, 3-65.

von Mühlenen, A. Müller, H.J., & Müller, D. (2003). Sit-and-wait strategies in dynamic visual search. *Psychological Science*, **14**, 309-314.

Peterson, M.S., Kramer, A.F., Ranxiao, F.W., Irwin, D.E., & McCarley, J.S. (2001). Visual search has memory. *Psychological Science*, **12**, 287-292).

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, **3**, 179-197.

Takeda, Y. (2004). Search for multiple targets: Evidence for memory-based control of attention. *Psychonomic Bulletin & Review*, **11**, 71–76.

Williams, L.G. (1967). The effects of target specification on objects fixated during visual search. In A.F. Sanders (Ed.) *Attention and Performance*, North-Holland. 355-360.