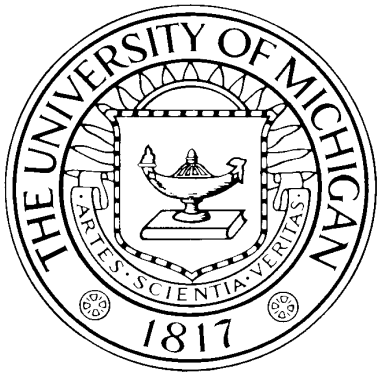
Visual Search Explained with a Computational Cognitive Architecture: Early Visual Processes, Eye Movements, and Task Strategies

David E. Kieras
University of Michigan



DTIC AD1114079

Report No. TR-20/ONR-EPIC-21

DOI: 10.13140/RG.2.2.23716.91520

August 20, 2020

This work was supported by the Office of Naval Research under grant number N00014-16-1-2560.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
20-08-2020	Technical Report	06/01/2016 - 12/31/2019

4. TITLE AND SUBTITLE
Visual Search Explained with a Computational Cognitive Architecture: Early Visual Processes, Eye Movements, and Task Strategies

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-16-1-2560

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
David E. Kieras

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Michigan
Division of Research Development and Administration
Ann Arbor, MI 48109

8. PERFORMING ORGANIZATION REPORT NUMBER
TR-20/ONR-EPIC-21

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
875 N. Randolph Street Suite 1425
Arlington VA 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Visual search is important in many practical situations such as user interfaces. The primary theory of visual search performance is based on covert visual attention, with no role for low-level or early visual factors, eye movements, or task strategies, although these have been empirically measured and can be easily characterized. This report presents models constructed using the EPIC computational cognitive architecture that show that these ignored factors actually suffice to account in quantitative detail for visual search performance, casting doubt on covert visual attention as an explanatory concept, and enabling the future development of more complete and accurate models of visual search.

15. SUBJECT TERMS
Cognitive Architecture, Human Performance Modeling, Visual Search, Attention

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UL	83	

Visual Search Explained with a Computational Cognitive Architecture: Early Visual Processes, Eye Movements, and Task Strategies

David E. Kieras

University of Michigan[1]

Abstract

Visual search is important in many practical situations such as human-computer interfaces. The current most popular theory of visual search performance is based on covert visual attention, with no role for low-level "early" visual processes, eye movements, and task strategies, although such explanatory entities have been empirically measured and can be easily characterized. This report presents detailed computational models of *simple* visual search tasks in which subjects respond whether a target is present or absent in stimulus displays of relatively small numbers of objects. These models were constructed using the EPIC computational cognitive architecture. The model simulation results show that despite having been ignored by covert attention theories, parsimonious combinations of early visual processes, eye movements, and task strategies actually suffice to account in quantitative detail for visual search performance. These findings thus cast serious doubt on covert visual attention as an explanatory concept in this domain, and are harbingers for the future development of more complete and accurate computational models of visual search.

1. Introduction

Visual search is a task that most people perform many times a day. Some object is sought, and the visual surrounding is examined until the object is found. Many real-world computer tasks, ranging from personal computers to complex military radar systems, require finding particular objects such as icons on the display. In a laboratory visual search experiment, the subject is presented with a display of objects and a *search task* which is a specification of the *target*, the sought-for object or objects. The other objects on the display are defined by the search task specification as *distractors*. The subject is asked to examine the display and locate the target (e.g. by clicking a mouse pointer on it), count the number of targets present, or simply to indicate whether or not it is present. In the review that follows, tasks that require locating the target or counting the number of targets are termed *complex* visual search, and those the require only a present-absent/response are termed *simple* visual search.

1.1. Complex Visual Search

Complex visual search. Visual search research inspired by real-world tasks has a long history in experimental psychology and human factors. For example, a National Research Council

[1] This work was supported by the Office of Naval Research, under N00014-16-1-2560. Thanks are due to David E. Meyer and Anthony Hornof for many helpful suggestions and comments.

symposium in 1959 (Morris & Horne, 1960) presented a substantial body of experimental and analytic research relevant to military visual search tasks. This work involved sophisticated treatment of the non-uniform resolving power of the eye, eye movement measurements, signal detection analysis, and analysis of strategies for how people should conduct visual search. Another body of research in human factors compared different coding schemes for effectiveness in visual search (see Sanders & McCormick, 1987 for a summary). For example, Smith & Thomas (1964) found that target color was a very efficient cue in tasks that involving counting the number of target objects in displays containing 20, 60, or 100 objects, while target shape was a much less effective cue, resulting in much longer response times and more errors. Many studies of visual search used eye movement tracking, even though it was extremely difficult until recently. For example, a classic film-based eye-movement study by Williams (1967) used displays of 100 unique numerically-labelled objects that differed in color, size, and shape. Subjects were required to find the object with a specified label given different cue combinations of the target's color, size, or shape. The results showed that some cues were more effective than others; in particular, color led to much faster searches than size or shape, and the eye movements showed that subjects fixated much more frequently on objects with matching colors than matching size or shape. In these classic, and more recent experiments, subjects are required to actually locate the target object, such as fixating it or pointing to it in some way. Such tasks are called *complex visual search* in this paper.

1.2. Simple Visual Search

Simple visual search. Concurrent with the classic and recent work on complex visual search, another approach to investigating visual search developed separately and had quite different characteristics. It was inspired by early influential work of Neisser (1963), who introduced visual search of alphabetic text as an information-processing topic. Unlike in the research surveyed above, Neisser and his followers (e.g. Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989) assumed that purely visual factors and eye movements are unimportant for understanding and explaining the phenomena of visual search. Thus, those factors have played little or no role in their theories of visual search.

As part of this different approach to investigating visual search, subsequent studies after Neisser (1963) began using variants of a highly simplified experimental paradigm, termed *simple visual search* in this paper. In experiments on simple visual search, the subject views a display of relatively few objects (or items) and simply responds with a keypress about whether a specified target is *present* or *absent*. At most a single target is present (constituting a *positive* trial) or not (constituting a *negative* trial). The number of objects in the display (*set size*) and the specification of the target vs. distractors (*search task*) are the main independent variables. The reaction time (RT) for correct responses as a function of set size and trial *polarity* (positive vs. negative) is the major dependent variable.

Excellent reviews of results from a wide variety of experiments on simple visual search appear in Findlay & Gilchrist (2003), Wolfe (2014), and Hulleman & Olivers (2017). Figure 1.1 shows some example displays and tasks from three simple visual search tasks[2] studied by Wolfe, Palmer, and Horowitz (2010), whose experimental data is modeled in this report, and will be described more below. For the left-most panel, the task is to respond with a key press whether a *2*

[2] The names for the search task conditions in this paper differ from those in Wolfe, Palmer, and Horowitz (2010).

shape (target) is present or absent in a field of *5* shapes (distractors), the *Shape* task. In the center panel, the task is to respond with a key press whether a red bar is present or absent in a field of green bars, the *Color* task. In the right panel, the task is to respond whether a red vertical bar is

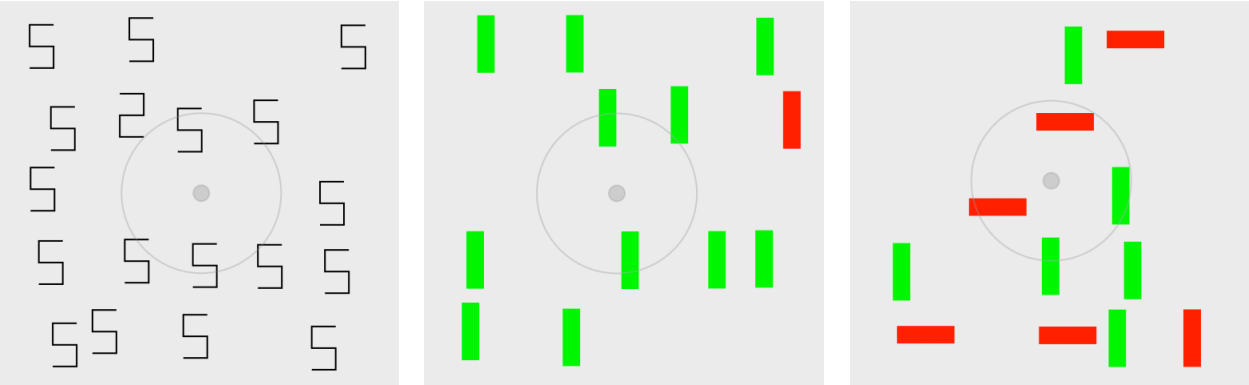


Figure 1.1. Example displays produced by the model for positive trials in each task condition used in the Wolfe, et al. experiment for set size of 18. From left to right, tasks are: Shape, Color, Conjunction. Gray circles indicate initial fixation point position. Outer circle: 10° diameter, Inner circle: 1° diameter.

present or absent in a field containing red horizontal bars and green vertical bars, the *Conjunction* task (the target is the conjunction of red and vertical). The number of objects (shapes or bars) in the display was varied from 3 to 18, and the target was present on half the trials, and absent on the rest.

The data typically resulting from such experiments show a roughly linear increase in RT with set size. Different visual properties and search tasks produce positive trial slopes ranging from essentially zero to roughly 50 ms/item or more for more apparently difficult search tasks. The negative trial slope is often roughly twice the positive trial slope, which suggests that a serial self-terminating search process is involved. Error rate (ER) usually increases with set size and apparent task difficulty, but is usually fairly low. Following the common practice in RT experiments, as long as the ER is "low enough", the RT for correct trials is believed to be uninfluenced by errors, and thus only the correct response RT is considered, and the ER can be ignored. This practice was criticized by Pachella (1974), but still continues.

Work by Treisman (1969; Treisman, Sykes, & Gelade, 1977) on the role of attention in visual search led to the seminal Triesman & Gelade (1980) paper, followed by Wolfe and his coworkers, starting with Wolfe, Cave, & Franzel (1989). The simplicity of the methodology and appeal of the attention-based theories offered by these authors led to a vast number of studies (see reviews in Wolfe, 2014, Hulleman & Olivers, 2017). Treisman and Gelade (1980) found that the slopes were zero or very small for tasks like the Color task, but larger and linear for tasks like Conjunction. However, Wolfe, Cave, and Franzel (1989) found that searches with apparently difficult discrimination requirements, like Shape, produced much larger slopes than Conjunction-like tasks, which could in fact be relatively fast. Wolfe (2014) thus discards the Triesman and Gelade (1980) dichotomy and placed search tasks on an efficiency scale based just on RT slope, with tasks like Color being *efficient* and those like Shape being *inefficient*, and tasks like Conjunction being intermediate.

1.3. Covert Attention Theory of Visual Search

Search seems to be too fast for eye movements. An obvious explanation for the linear RT slopes is that subjects move the eyes to each item sequentially to perform the search, which is a good approximation of the results in the complex visual search work summarized above. However, unless the objects are extremely difficult to discriminate, the typical RT slopes observed in simple visual search seem to be much faster than possible if eye movements would have to fixate each object separately (cf. Neisser, 1963; Triesman & Gelade, 1980; Wolfe, 2007, p. 106-107; Wolfe et al., 2010).

This discrepancy motivates the basic theoretical claim underlying most of the simple search theorizing, that the sequential search is done not by *overtly* moving the eyes, but instead by *covertly* moving selective attention from one internal object representation to another. The claim is that this covert shift of attention can happen much more quickly that overt shifts of the eye position. Judging from Triesman, Sykes, & Gelade (1977), this *covert selective attention* theory of visual search appears to have its roots in Neisser's (1967) assertion, based on extremely early computer vision speculation (Minsky, 1961), that "focal attention" is necessary to "bind" together primitive features into a visual object; this attention-based binding operation was advanced in the tremendously influential Triesman & Gelade (1980) Feature Integration Theory as an explanation for the different RT effects in Color and Conjunction-like tasks. In this theory, if no binding has to be done, as in the Color task, then a parallel process can simply detect the presence or absence of the target feature quickly and independently of set size, while the Conjunction task requires serially deploying attention to each object to detect an object that has both target features.

The dominance of the covert selective attention hypothesis has led to a remarkable dismissal of the role of visual factors like retinal nonhomogeneity and the relevance of eye movements (for additional discussion, see Findlay & Gilchrist, 2003; Rosenholtz, Huang, & Ehinger, 2012). This dismissal is all the more puzzling because Neisser (1967), Treisman & Gelade (1980), and Wolfe (2007, 2014) and the mainstream of visual search work descended from them, all refer to visual factors like retinal nonhomogeneity and eye movements. Yet these investigators did not consider them to be important compared to covert attention, even though many simple visual search experiments have demonstrated their relevance (e.g. Carrasco & Frieder, 1996; Wertheim, Hooge, Krikke, & Johnson, 2006; Zelinsky & Sheinberg, 1995, 1997).

1.4. Are Eye Movements Involved in Simple Visual Search?

A key and simple point is that if more than one visual object can be processed in a fixation, the argument that the RT slopes are too small to be consistent with eye movements simply disappears (cf. Findlay & Gilchrist, 2003; Hulleman & Olivers, 2017). In fact, a general result (e.g. Williams, 1967; Zelinsky & Sheinberg, 1995, Hulleman & Olivers, 2017) is that search time (RT) is strongly related to the number of fixations during the trial, which makes the general dismissal of visual factors and eye movements in the simple visual search literature hard to justify. This calls for a re-examination of the role of eye movements in simple search tasks, which reveals important methodological problems.

There is a small literature that directly compares simple search tasks with and without eye movements. Isolated papers from this literature have sometimes been cited (e.g. by Wolfe, 2007 p. 106-107, Wolfe et al., 2010) as supporting the generalization that the same RT effects are

obtained regardless of whether eye movements are made. However, the following brief survey shows that this generalization appears unjustified and highly questionable.

There are five experimental procedures that have been used in simple visual search tasks. In all of them, the subject is asked to fixate a central point before the display appears.

1. The *short display time procedure* seeks to eliminate the effects of eye movements by making the search display exposure time so short (e.g. 180 ms) that the eyes cannot be moved from the central fixation point before the display disappears. This means that many factors can be studied without the complications of eye movements or the difficulty and expense of eye tracking. However, clearly subjects respond on the basis of some kind of evanescent stored representation of the display rather than the actual display.

2. The *eyes-free procedure* allows the subject to make eye movements as they choose, while the display remains visible until the subject responds. This is obviously the most natural procedure as well as the simplest to implement.

3. The *eyes free with eye movement tracking* is an especially useful procedure because it provides direct information on the role of eye movements. However, eye tracking methodology has been more difficult and expensive to implement than simple RT experiments, and so it is unusual in the simple visual search literature, although there is a long-established literature on eye movements in complex search tasks.

4. The *instructed eyes-fixed procedure* asks subjects to maintain fixation on the central fixation point while performing the visual search task. The display remains visible until the response. It is often assumed that subjects adhere to these instructions and no check is made on whether subjects actually comply. This procedure was used in Wolfe, et al. (2010, p. 1306).

5. The *instructed eyes-fixed with monitoring procedure* uses eye tracking equipment to monitor whether the subject succeeds in maintaining fixation. However, because of the accuracy limitations of eye tracking, a threshold such as 1 degree is set for the maximum eye movement that is allowed. If the threshold is exceeded, the trial is either replaced or simply dropped from the analysis. This implies that a better characterization of this procedure is that rather than eliminating eye movements completely, it eliminates fixations that are on or close to the search objects.

There is a consensus that the short display time procedure produces results that while well-behaved, differ substantially from those of the eyes-free or instructed eyes-fixed procedures. For example, Klein & Farrell (1989, Exp. 1) and McElree & Carrasco (1999), found short display time produced RT and ER patterns that were both very different than the typical eyes-free or -fixed results.

There are a few direct comparisons of instructed eyes-fixed and eyes-free procedures, e.g., Klein & Farrell (1989), Motter & Simoni (2008), Scialfa & Joffe (1998), Wertheim, Hooge, Krikke, & Johnson (2006), and Zelinsky & Sheinberg (1997). The results are inconsistent and problematic. In some studies the effect of procedure on RT differed depending on set size and whether the search task was efficient or inefficient. Across these studies, instructed eyes-fixed produced both faster and slower RTs than eyes-free. Some studies do not report ER at all; however, in most cases, the ER tends to be higher in eyes-fixed than in eyes-free. Clearly, faster

RT and higher ER in eyes-fixed could just be a speed-accuracy tradeoff — guessing a response with eyes fixed would simply take less time than moving the eyes to determine a more accurate response. However, because the ERs differed between procedures, the presence (or absence) of similar RT effects is at best ambiguous (Pachella, 1974).

What accounts for the inconsistencies between these experiments? There are purely methodological factors, all contributing to the likelihood of Type I and Type II errors, leading to inconsistent results. Most studies used stimuli that were either highly discriminable or at small eccentricities, which means that extra-foveal vision might often be adequate to discriminate them with few or no eye movements, substantially weakening the procedure manipulation. Most studies had fairly unpracticed subjects, and relatively few trials/condition. Instructions and incentives for stable speed vs. accuracy performance were not used. Most studies did not control crowding effects due to display density, which together with small sample sizes could cause high variability in RT for larger set sizes. The number of subjects was often very small (as few as two) compared to the normal practice for RT experiments. The procedure factor was usually between-subjects or even between-experiment, decreasing the statistical power. In most cases, it is difficult to assess the statistical issues because of unclear method descriptions and incomplete or inadequate statistical analyses and reporting. Finally, the claim of "no difference" in RTs between eyes-fixed and eyes-free procedures requires accepting the null hypothesis; since the experiments were methodologically and statistically weak, such a claim has no force, especially since the results are inconsistent.

A few of the studies used the eyes-fixed with monitoring procedure. As noted above, small eye movements were allowed, and apparently made. For example, in Klein & Farrell (1989), the eccentricity of the display objects from the fixation point was constant at 2.4 degrees, and eye movements of up to 0.5 degrees either horizontal or vertically were allowed, giving up to 0.71 degrees diagonally, corresponding to 30% of the stimulus eccentricity. In other experiments the allowed eye movements could be up to 1.27 degrees of eye movement from the fixation point. Wertheim, et al. (2006, Exp. 2) actually report the number of these small movements and provide some sample eye movement traces.

Small eye movements in the vicinity of the fixation point could be functional. Recent work on *fixational eye movements* (e.g. microsaccades; see review by Martinez-Conde, Macknik, & Hubel, 2004) suggests that these small eye movements could cause a stimulus object to change receptive or integrative fields in extra-foveal vision enough to make it discriminable. Thus, even if subjects succeed in following the eyes-fixed instructions, there is still the possibility that small fixational eye movements, either voluntary or involuntary, would allow nearby or highly discriminable extra-foveal objects to be discriminated. Thus as noted above, at best the eyes-fixed procedure eliminates full fixations on the objects, rather than eliminating *any* role of eye movements. Under these circumstances, it is hard to justify attempting to impose an artificial eyes-fixed constraint on subjects rather than let them use their eyes as they normally would.

The conclusions from this brief survey are that (1) If a manipulated factor causes ERs to differ, but not RTs, the claim that the factor has no effect is unjustified (cf. Pachella, 1974). (2) Claiming that some type of eye movements is not involved in simple visual search is not justified empirically. (3) The visual discriminability of the search objects, and the characteristics and need for eye movements, are closely related; one can not be studied without reference to the other. (4)

Especially in the instructed but unmonitored eyes-fixed procedure, it should be assumed that subjects could still be making eye movements.

Thus despite the historical, but unjustifiable, claim to the contrary, assuming that eye movements can be involved is a better first step in constructing sound explanations of simple visual search effects.

1.5. Explaining Visual Search

This paper will apply three key explanatory concepts to visual search that differ from the covert attention account. The first two were described in the excellent review by Findlay and Gilchrist (2003) under the heading of *Active Vision*; Newell (1973) stated the third very emphatically:

- ***"Early" visual factors***. These are sensory and perceptual aspects of the visual stimulus such as object eccentricity, size, spacing, color, orientation, and shape. These are termed *early* in this paper because the major processing of these factors appears at the level of the retina and the first levels of visual perception. An important example is that due to *retinal non-homogeneity*, resolving power is highest in the fovea, but extra-foveal or peripheral vision nonetheless provides useful information. Moreover, different visual features, termed *properties*[3] in this paper, differ in how well they can be detected in central versus peripheral vision. For example, color can be detected quite well in peripheral vision, but detailed shape, as in normal-sized letters, often requires foveal vision for recognition.
- ***Eye movements and visual guidance.*** The human moves the eyes around the visual scene as needed to bring visual objects into the high-resolution portion of the retina. However, the information from extra-foveal or peripheral vision is not ignored; rather it provides guidance about where the eyes should be moved next, and might actually suffice to complete the task.
- ***Task strategy.*** Subjects will acquire and apply task-specific strategies for moving the eyes as needed to complete the task and making the appropriate response. The task strategy coordinates the perceptual and motor systems in a visual search task.

The concept of task strategy needs some elaboration at this point. The presence of a task strategy of some sort has been implicit in almost all theoretical work, as implied by the obvious ability of humans to respond to different instructions about what constitutes the task, such as the type of the target, and do so promptly in an experimental setting. But long ago convincing arguments were advanced, but largely ignored, that strategies devised and executed by subjects need to be *explicitly represented* in a model. Reitman (1970, p. 501) pointed out that strategies and the underlying memory and processing mechanisms were quite distinct and needed separate representation in a model. Later, Newell (1973) made a similar point, and in a more pithy way, using *method* to refer to the subject's task strategy:

---

[3] The conventional term *feature* is ambiguous between a stimulus dimension (e.g. *color*), and a value on that dimension (e.g. *red*); the models in the paper use a convention inherited from LISP of *properties* and *values*; a *property* and *value* pair names a dimension and a value on that dimension.

> The most fundamental fact about behavior is that it is programmable. That is to say, behavior is under the control of the subject to shape in the service of his own ends. ….
>
> *First Injunction of Psychological Experimentation:* Know the method your subject is using to perform the experimental task.
>
> In short, we are totally engaged, in psychological experimentation, in the discovery and verification of the specific methods used by the subject in doing the experimental tasks. (p. 293 - 294)

Similarly, Meyer & Kieras (1997a, 1997b, Kieras & Meyer, 2000; Schumacher, Seymour, Glass, Fencsik, Lauber, Kieras, & Meyer, 2001) showed by detailed modeling that human performance effects attributed to structural constraints such as a "response selection bottleneck" were more correctly, and fruitfully, accounted for in terms of flexible strategies operating with constrained perceptual and motor mechanisms.

Computational models based on these three explanatory concepts account well for data from eye-tracking experiments on complex visual search (Kieras & Marshall, 2006; Kieras, 2010, 2016; Kieras & Hornof, 2014, 2017; Kieras, Hornof, & Zhang, 2015; Hornof & Halvorsen, 2003; Halvorsen & Hornof, 2011; Hornof & Kieras, 1997, 1999). However, the simple visual search results present a challenge for EPIC: Not only is the task strategy very different, but critically, t*he data are claimed to support theories in which the mechanisms that EPIC does not include (such as covert attention) are central, while mechanisms that EPIC does include (early visual factors and eye movements) are deemed irrelevant*. Thus if an EPIC model can account for the results in quantitative detail, it follows that hypothetical cognitive mechanisms such as 'covert attention shifting', perceptual object-feature 'binding', and so forth, which have pervaded the theoretical literature on simple visual search, are not needed to explain simple visual search performance.

1.6. Overview of the Paper

In the next main section, Section 2, this paper provides a reanalysis of a very high-quality dataset, made available by Wolfe, Palmer, & Horowitz (2010), concerning performance in three classic simple visual search tasks, which are illustrated in Figure 1.1. Following this presentation is a brief critique of the key theoretical and methodological assumptions in the current work on simple visual search.

Section 3 describes the EPIC computational cognitive architecture and then subsequent sections present an active vision model constructed with EPIC to account for the Wolfe, et al. (2010) results, both at the level of the aggregated data (Section 4), and at the level of clusters of individual subjects who performed similarly (Section 5). The basic results are that the EPIC models account extremely well for the aggregate data, both qualitatively and quantitatively, using a combination of simple perceptual and motor mechanisms and simple rational strategies that are well-adapted to the perceptual differences between the tasks. The individual subject cluster data is similarly well accounted for; in most cases subjects followed the same strategy that fit the aggregated data, but had different perceptual parameters (e.g. acuity); a few subjects followed a different strategy in the Conjunction task, the most complex. Section 6 states conclusions and implications for future work.

2. The Visual Search Experiment

This section is a reanalysis of a very high-quality dataset, made available by Wolfe, Palmer, & Horowitz (2010), concerning performance in three classic simple visual search tasks, which are illustrated in Figure 1.1, and used for the modeling work in this paper. The data were collected by Wolfe, et al. (2010), and made available for download at http://search.bwh.harvard.edu/new/data_set_files.html. This dataset is exceptional because of the relatively well-specified stimuli and very large number of trials from very well-practiced subjects. For completeness, the experimental method is re-stated here in the context of how the experiment was simulated in the model.

2.1. Method

Tasks. There were three different present/absent search tasks; Figure 1.1 shows a sample target-present display produced by the model for each task condition.

The three tasks are as follows; the labels differ from Wolfe, et al., but are used here for greater clarity.

- **Shape task**: The objects are "digital 5" and "digital 2" shapes made up of vertical and horizontal line segments similar to these digits on a traditional seven-segment display. The distractors are always *5*s; if present, the single target is always *2*.
- **Color task**: The display contains vertical green or red bars. The distractors are always green. If present, the single target is always red.
- **Conjunction task**: The display contains bars that are vertical or horizontal, and red or green. Half of the distractors are green verticals and half are red horizontals. If present, the single target is always a red vertical bar.

Stimuli. Wolfe, et al. provide a good level of detail about the stimulus properties, but do not describe how the individual displays were created, and the download data set does not contain information about the actual display used in each trial, so for purposes of modeling, the display had to be generated for each simulated trial using the process described in what follows. The example displays in Figure 1.1 above were generated by the model.

The search display was an area 22.5° × 22.5°, containing 25 invisible cells of 5° × 5°; Wolfe, et al. state that each object appeared in a random location within one of the cells, but did not state whether or how touching or overlapping objects were prevented. Assuming that such displays were not allowed, the random location within a cell was constrained to keep the horizontal or vertical edge of an object at least 0.25° away from the cell boundary, ensuring a minimum separation of 0.5° between edges of adjacent objects. Throughout the modeling, the location of an object was defined to be the center point of the bar or digit bounding box. Set sizes were 3, 6, 12, and 18. In the model, a display was generated for each trial as follows: the set-size number of distractors were first placed in randomly chosen display cells. With probability of 0.5, the trial polarity was then determined; if the trial was positive (target present), a randomly chosen distractor was replaced with a target object.

In the Shape task, the objects were 1.5° × 2.7° character-like shapes. In the model, the object size is defined in the models as average of the horizontal and vertical bounding box dimension.

giving 2.1° for the Shape object size. The target was a *2* and the distractors were *5*s. In the Color task, the objects were 1° × 3.5° vertical bars (size defined as 2.25°); the target bar was red, distractor bars were green. In the Conjunction task, the objects were also 1° × 3.5° bars, red or green, oriented either horizontally or vertically. The target was a red vertical bar, and distractors were red horizontal and green vertical bars. Half of the distractors were chosen to be of each type, with set size 3 special-cased so that at least one distractor of each type was present. Since a positive trial display was produced by replacing a random distractor with a target, over trials, each type of distractor would appear equally often.

Design. There were 10 subjects in the Conjunction task condition and 9 in the other two. One subject was in both Conjunction and Shape, but the data set does not identify this subject, so the task condition was treated as a purely between-subject manipulation.

Procedure. Each trial began with a centered fixation cross. Subjects were instructed to "keep their eyes focussed on this cross" but because eye movements were not monitored, subjects could have moved their eyes, and based on the studies reviewed above, it is likely that they did so, at least in some conditions. The search display was presented and remained visible until the subject pressed a key for target present or target absent. Subjects were instructed to respond "as quickly and accurately as possible," and correct/incorrect feedback was presented for 500 ms after each trial. But no explicit incentive such as a payoff function (Edwards, 1961; Sternberg, 2016 Appendix B) was provided to make the task instructions more specific or implementable by the subjects. Unlike many visual search experiments, the subjects were very well practiced, with about 500 trials per subject for each combination of set size and positive/negative trial polarity, a total of about 4000 trials. This suggests that generally, subjects maybe have been biased towards being as fast as they could with whatever they considered acceptable accuracy, but otherwise, with no incentive to make this tradeoff in any more specific way.

2.2. Results

The downloaded data consisted of the RT and correct/incorrect status for each trial for each subject at each set size and trial polarity. RT outliers were removed from the data following the description in Wolfe et al. (2010). Following common practice in RT experiments, the data were reduced as follows: For each task condition, for each subject, the mean RT for correct trials and the proportion of errors (error rate, ER) for that subject was calculated for positive and negative trials at each set size, giving a total of 8 data points for RT and 8 data points for ER for each subject. These subject means were then averaged to produce the data points plotted in Figure 2.1. Throughout this report, positive (target present) trials are shown as red points and lines, negative (target absent) trials in black; the Shape task is plotted with squares, Conjunction with triangles, and Color with circles. The 95% confidence intervals around each data point in Figure 2.1 were calculated by determining the standard error of that mean using the 9 or 10 individual subject means contributing to that point, thus reflecting between-subject variability, but not within-subject variability.

Wolfe, et al. did not report any overall statistical tests of these results. Therefore, unequal-*n* ANOVAs were performed on the reduced data using the **R ez** package (R Core Team, 2017; Lawrence, 2016). For RT, the main effects of task condition, trial polarity, set size, and all two- and three-way interactions were significant ($p < .05$). For ER, whose overall average was 2.4%,

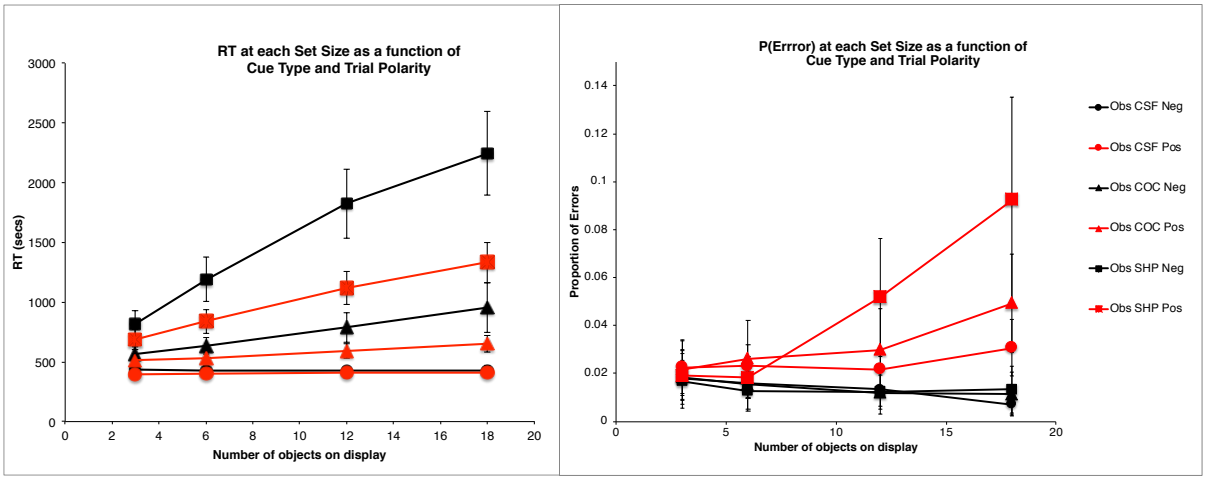


Figure 2.1. Observed RT (left panel, scale 0-3000 ms) and ER (right panel, scale 0-0.15) in each task condition. Shape: squares, Conjunction: triangles, Color: circles. negative trials are plotted in black, positive trials in red. The error bars are 95% confidence intervals based on the standard error of the individual subject means underlying each plotted mean and thus reflect between-subject variability.

the task condition main effect was not significant ($p > .1$) but the trial polarity and set size main effects, and all two- and three-way interactions were significant ($p < .05$).

Examination of specific within-subject effects was done with Fisher Least Significant Difference (FLSD) values, which to avoid clutter are not shown on the graphs. For within-subject (within-condition) comparisons of the 24 mean values plotted in the graphs, the FLSD values are 68 for RT and 0.011 for ER. This shows that for the Color task, of course the RTs are not different for either trial polarity or set size. For Conjunction and Shape both the FLSD value and the between-subject confidence intervals indicate that the increasing trends for the RTs, and the tendency for negative RT to be greater than positive are reliable effects. For ER, these values indicate that the differences between the negative trial ERs at set size 3 and 18 are significant. For positive trial ER, for the Color task, the set size 18 point is not quite reliably different from the smaller set size points; for Conjunction, the set size 18 ER is higher than all smaller set sizes ERs, but the ER for set size 12 is not reliably higher than for the smaller set sizes. Finally, for Shape positive trials, set size 18 ER is higher than all smaller set sizes, and ER for set size 12 is higher than the smaller set sizes ER. Overall, these comparisons can be summarized as: most of the apparent within-subject/within-condition effects in the graphs are reliably different even if the between-subject confidence intervals overlap.

Table 3.1 presents some summary statistics; given the great importance in the literature attached to the linearity and slope of the RT functions, this table provides the intercept, slope,

Table 3.1 Observed Aggregated Data Statistics

	negative				positive					
Task Condition	Intercept	Slope	r^2	ER	Intercept	Slope	r^2	ER	ER Max	Slope ratio
Color Task	436	-1	0.68	0.014	395	1	0.90	0.025	0.031	-0.69
Conjunction	480	26	1.00	0.014	483	9	1.00	0.032	0.049	2.84
Shape	589	95	0.99	0.014	574	43	0.99	0.045	0.093	2.21

and r^2 of a linear fit to the mean RT data in each condition, along with the ratio of the negative trial RT slope to the positive trial RT slope. Since the Color task slopes are essentially zero, the slope ratio is meaningless in this condition. Note how the Shape slope ratio is roughly 2, the classic value for a self-terminating serial search, but the slope ratio is definitely larger than 2 in the Conjunction condition. Also shown is the mean ER and the maximum ER in each condition.

2.3. Discussion

Reaction Times. The RT results follow the classic pattern obtained in most simple visual search experiments. The RT functions are essentially flat in the Color task (positive trial slope is about 1 ms/item). In Conjunction and Shape, positive and negative trial RTs have a substantial slope, with the negative trial slope very roughly twice that of the positive trials, the classic indicator of a serial self-terminating search. Treisman & Gelade's (1980) Feature Integration Theory focussed on explaining why conjunctive searches (as in this Conjunction task) have these sloped linear RTs compared to single-feature searches (like the Color task) with their flat RTs. But Wolfe, Cave, and Franzel (1989) showed that this central claim did not stand up to further empirical work: conjunctive searches can be relatively fast (as they are here, at 9 and 26 ms/item), and single-feature searches can be either very fast or very slow (as in Shape), depending on the specific visual properties involved. A fine point to note is that there is a hint of negative acceleration in the steepest RT function; much stronger curvilinear effects are commonly observed (see Wolfe et al., 1989) but have usually been ignored (but see Buetti et al., 2016).

Error Rates. Some intriguing patterns appear in the ERs, but for the most part, Wolfe et al. (2010) chose neither to analyze nor to theoretically interpret the observed ER pattern. Instead, their remarks regarding ERs were limited to mentioning just two facts: (1) the overall mean ER across search tasks, display sizes, trial polarity, and subjects was only 2.4% ; (2) a small but reliable downward trend in the frequency of errors on negative trials as a function of display size.

We show subsequently that this lack of greater attention to the observed ERs was unfortunate. Nor was it at all unusual. Indeed, error-rate (ER) effects typically have been ignored in the literature on simple visual search when overall ERs were as low as the 2.4% found by Wolfe et al. (2010), and an overall speed-accuracy tradeoff was not present, as shown by overall ERs being positively correlated with overall RTs. Thus, under these conditions, both Wolfe et al. and other past investigators typically proceed as if ignoring the errors and focussing only on the RTs from correct trials was acceptable (cf. Pachella, 1974).

Such an approach to data analysis and interpretation embodies a long-standing practice in the study of RT effects as articulated by Sternberg (1969). Namely, RTs may be assumed to manifest internal stages of processing, so we can construct models and theories about cognitive processing based on patterns in RT effects. However, unlike for correct responses, there are many possible ways subjects could make an error. Consequently, the theoretical interpretation of error RTs is ambiguous, and possible interpretations have little theoretical interest. The standard practice thus has been to consider RTs from correct trials only, and as long as the ER is "low enough", it has been assumed that erroneous behavior does not seriously contaminate the correct trial RTs.

However, despite the conventional justification for ignoring the error rates, it is clear that the ER effects in the Wolfe et al. data are highly systematic. In particular, there are very few errors on negative trials (False Alarms), and their rate does not appear to depend at all on the task

condition and very little on set size (apparently declining slightly with set size). The average across all task conditions and set sizes is only 1.4%. In sharp contrast, the errors on positive trials (Misses) are more frequent than the False Alarms, and definitely depend on the task, being lowest in the Color task and highest in Shape. The Miss rate in the Conjunction and Shape tasks strongly increases with set size. As noted above, the 2- and 3-way interactions are strongly significant in spite of large individual differences. Because the ER effects are pronounced and statistically reliable, they deserve to be explained along with the RT effects rather than ignored. Furthermore, rather than postulate that ER depends on "task difficulty" and represent it by error rate parameters that increase with "task difficulty," it would be better to explain this effect in terms of the visual and strategy mechanisms at work in the different tasks. Therefore, in what follows, ER is considered as a first-class data source alongside RT for developing and evaluating models of visual search.

3. The EPIC Cognitive Architecture

A cognitive architecture is basically a modern version of the computer metaphor for human information processing. Computers have a fixed *hardware* structure, called the *computer architecture*, with various specific interconnected processing mechanisms. They can be programmed with *software* that is stored and executed by the hardware, in order to perform various different tasks through using the hardware mechanisms appropriately. The hardware architecture is *fixed* in that it doesn't change its structure or characteristics depending on the task to be performed; rather the software has to deal with the specifics of the task. Without a software program, the hardware does nothing; but a program has to be written in terms of what the hardware makes possible. How well the system actually performs the task depends both on the characteristics of the hardware architecture and how the programming makes use of the architecture.

3.1. The Concept of a Cognitive Architecture

These concepts can be mapped to human information-processing systems. By analogy, humans have "hardware" for perception, cognition, and motor activities, with mechanisms whose characteristics are task-independent and thus fixed. The “software" consists of task-specific strategies involving procedural knowledge that is acquired and executed by components of the cognitive architecture. The Atkinson & Shiffrin (1968) model of memory, in which human memory systems were viewed as configurable to meet the requirements of a particular task, was an early precursor of contemporary cognitive architectures. Another precursor was the Model Human Processor presented by Card, Moran, & Newell (1983), which represented a human as a set of perceptual and motor processors and memory stores with fixed characteristics, and a cognitive processor that functioned in a task-specific fashion, thereby resulting in models that could produce practically useful predictions of task execution times.

As pointed out by Newell (1973), a special property of human behavior is that it is *programmable*, which means that in addition to the fixed components of the architecture, some kind of flexible *control process* is require to represent and apply a *task strategy*, the procedural knowledge for performing the task, giving a *complete model* of ‘end-to-end’ performance from perception to action. Newell further suggested that *production systems* were a good candidate for

representing the control process. This concept gave rise to the specific architectures such as ACT (Anderson, 1983), Soar (Laird, Newell, & Rosenbaum, 1987), and EPIC (Kieras & Meyer, 1997; Meyer & Kieras, 1997a, b; Kieras 2016) which represent the control process as *production rules*, which are if-then condition-action rules which are acquired as procedural knowledge and executed by the central cognitive system.

Such cognitive architectures are more comprehensive and explicit than previous attempts to model human cognition and performance. In particular they are computational, which means they can be implemented as explicitly defined computer simulation systems, and so can produce rigorous and quantitative predictions of empirical results.

Constructing a model in such an architecture is a much more disciplined process than prior modes of theorizing. Especially important is that the architecture provides constraints — behavior has to be produced using a specified set of mechanisms and components, especially for the control process. This is a striking contrast to earlier computational models, which often contained arbitrary code and components (e.g. many of the models in Norman, 1970).

Another area of greater discipline is in specifying the control process. In traditional information-processing models of human performance, the model consisted of "boxes" such as memory stores, with some mathematically-defined characteristics, but the movement of information from one store to another, or from perception, or to motor systems, happened in an unspecified manner — apparently the boxes were magically "wired up" with patch-cords to perform the task by some unspecified outside agent, an "executive process" (a homunculus?). This unspecified control process is still characteristic of many cognitive models; examples for simple visual search are the Feature Integration Theory (Triesman & Gelade, 1980), and Guided Search (Wolfe, Cave, & Franzel, 1989; Wolfe, 2007). These theories propose that perceptual input activates feature representations that are examined and combined by an implicit process that depends on the search task and which may or may not require a time that depends on the number of objects on the display. Such models are essentially pre-cognitive-architecture "box models" (see Kieras, 1980, 2007) because even if some of the components are described mathematically, their connections are described verbally. Thus, these models do not have a component that represents a task strategy in any explicit, and therefore, "programmable," fashion. So explaining how the internal representations are "wired" to responses differently in the Shape, Color, and Conjunction tasks is completely implicit.

3.2. General Structure of EPIC

3.2.1. Overview

The current major production-rule architectures differ in terms of emphasis: As originally proposed, ACT and Soar were mainly concerned with learning or problem-solving, while EPIC, inspired by the Model Human Processor, focussed on human performance, where learning is less important, but a more complete representation of perceptual-motor mechanisms is required compared to other cognitive architectures. This emphasis emerged in the original development of EPIC with the goal of modeling multitask performance, mental workload, and human-computer interaction. It became clear early in this work that the perceptual-motor constraints and requirements are fundamental constraints on task performance that absolutely had to be included to produce comprehensive and accurate models of human performance in important tasks.

Thus models built in EPIC are *complete models,* in Newell's (1973) terminology, that are "end to end" or "embodied" in the sense that they start with simulated sensory input and produce simulated physical movements; while these are abstracted and simplified, they constrain the model to match the requirements of the complete human task, rather than focus only on the internal purely cognitive processes, as had been the case with almost all previous computational cognitive modeling work (c.f. Kieras & Meyer, 1997).

Extensive presentations of EPIC are available elsewhere (Meyer & Kieras, 1997a,b; Kieras & Meyer, 1997; Kieras, 2007; Kieras, 2016), and has been applied already to a large variety of tasks including verbal working memory (Kieras, Meyer, Mueller, & Seymour, 1999), multiple-task performance (Meyer & Kieras, 1999), executive control (Kieras, Meyer, Ballas, & Lauber, 2000) and multiple-channel speech processing (Kieras, Wakefield, Thompson, Iyer, & Simpson, 2016), and as mentioned above, complex visual search. In what follows, the EPIC architecture will be described only in enough detail to support the rest of the paper, but with some detail on the architectural components and mechanisms to support visual search in general, namely the active vision concepts of visual factors, eye movements, and task strategy. The example displays and tasks in Figure 1.1 will be used as examples in this description.

Figure 3.1 shows the overall structure of the EPIC architecture. In overview, EPIC provides a general framework for simulating a human interacting with an environment to accomplish a task. The environment is represented as device that provides simulated sensory input to the simulated human and responds to simulated motor actions from the simulated human. The simulated human consists of memory systems, perceptual processors, and motor processors, surrounding a cognitive processor. The device and all of these processors run in parallel with each other. The EPIC simulation software consists of modules for the device and each of human processors and

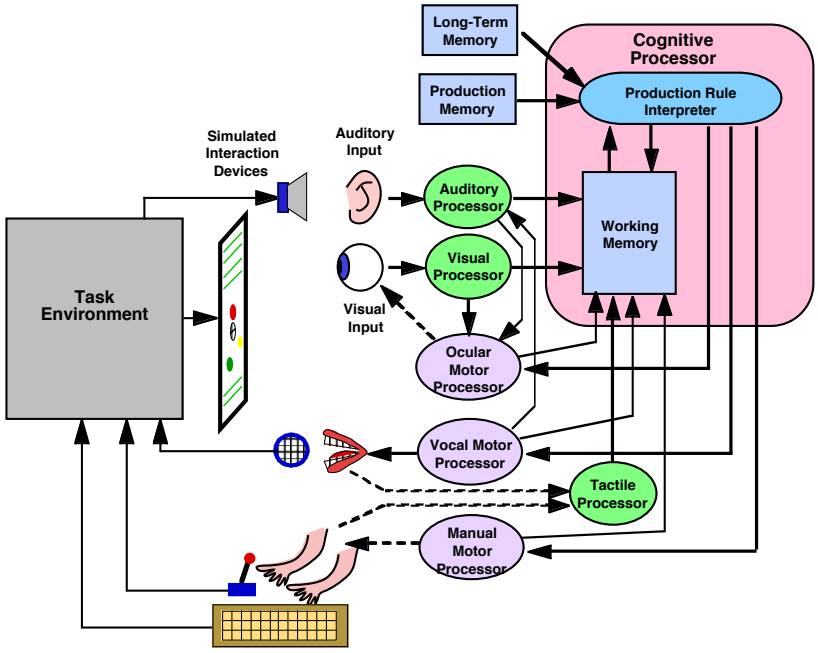


Figure 3.1. The EPIC architecture in simplified form. The simulated environment, or device, is on the left; the simulated human on the right.

memory systems connected in a fixed configuration. This defines the "hardware" of the architecture. The "software" takes the form of production rules, condition-action if-then rules that represent the procedural knowledge of how to perform a task, that are executed by the cognitive processor. To simulate human performance in a task, the modeler specifies production rules that implement a strategy for performing the task, and also may set parameters for the perceptual and motor processors. When the simulation is run, the architecture generates the specific sequence of perceptual, cognitive, and motor activities required to perform the task, within the constraints determined by the architecture and the task environment. Monte-Carlo runs of the simulation produce precise quantitative predictions of human performance as actual behavior sequences, which similarly to human data, are aggregated statistically to produce predicted mean RT and ER in each experimental condition.

The fixed sensory-perceptual and motor mechanisms are represented with mathematical models in the perceptual and motor processors; the perceptual mechanisms deliver symbolic perceptual descriptions to the cognitive processor, which uses a production-rule representation to perform symbolic operations such as making inferences, selecting relevant information, and following task procedures. The timing and accuracy of motor movements is similarly represented by mathematical models. This hybrid combination of mathematical models for low-level phenomena and symbolic models for high-level cognition works especially well in modeling how perceptual and motor mechanisms operate in the context of whole tasks. As such, EPIC has served as an efficient and elegant means to summarize a large body of experimental results in the perceptual, cognitive and motor domains as they relate to human performance in complex tasks. An important application area is providing a simulation testbed for optimizing human-computer interfaces.

More specifically, the cognitive processor consists of a production rule interpreter that uses the contents of production memory, long-term memory, and the current contents of the production-system working memory (PSWM) to choose production rules to fire. Auditory, visual, and tactile processors deposit information about the current perceptual situation into working memory; the motor processors also deposit information about their current states into working memory. The cognitive processor runs on a cyclic basis with a 50 ms cycle duration. At the beginning of each cycle, the conditions of all of the rules are tested in parallel against the contents of PSWM, and those whose conditions match are fired and their actions executed at the end of the cycle. The actions can modify the contents of working memory, which may change which rules will match on the next cycle, or the actions can instruct motor processors to carry out movements. The total number of production-rule cycles required to produce a response contributes to the time to complete the task. The motor processors control the hands, speech mechanisms, and eye movements. In addition to the parallelism across perception, cognition, and action, there is also parallelism within the cognitive processor because multiple production rules can "fire" simultaneously, giving cognitive processes that can be "multithreaded" analogous to modern computer systems. The ability of the rules to apply in parallel is an important feature; it simplifies the EPIC models for high-performance tasks such as multitasking, in which an executive process can be implemented in production rules that can control multiple task processes. This is reflected in the acronym: **E**xecutive **P**rocesses **I**nteract with and **C**ontrol the rest of the system by monitoring their states and activity (Kieras, 2007, 2016, Kieras & Meyer, 1997, Kieras, Meyer, Ballas, & Lauber, 2000; Meyer & Kieras, 1997a,b, 1999).

A fundamental assumption of EPIC and similar production-rule cognitive architectures is that subjects in experiments create a set of production rules when first instructed in the task, and then refine those rules as they gain experience in the task; clearly, feedback or incentives will play a role in how the strategy is refined. EPIC does not attempt to represent the underlying learning mechanisms, but can represent a strategy assumed to be in effect after some practice, and can characterize transfer of training effects in terms of overlap in the production rule sets, and thus can account for differences in procedure learning time (e.g. Kieras & Bovair, 1986). In the models described in this paper, it is assumed that the amount of practice is extensive enough that subjects have developed a stable strategy that can be chosen so as to fit the data within the constraints of the architectural mechanisms that EPIC provides.

3.2.1. The "EPIC Philosophy"

Notice what is *not* in this architecture, especially not in the cognitive processor: Consistent with the focus on human performance, there is no learning mechanism in the usual sense. But relevant to human performance, there is no concept of limited processing or memory capacity, no processing "bottleneck," no attentional selection mechanism, no reliance on an "activation" concept. These may well exist, but EPIC follows a minimalist approach: the need for such concepts must be demonstrated by modeling work rather than being assumed *ab initio*. Thus EPIC's assumptions about cognition are as parsimonious as possible; in fact, the *only* thing the current EPIC cognitive processor does is execute production-rule task strategies in a pervasively parallel way. Thus in contrast to the mainstream of cognitive theory, the emphasis is rely as fully as possible on the known characteristics of perception and motor mechanisms, with cognition providing only a minimal hypothetical mechanism for explicitly representing and executing a task-specific strategy that connects perception to action.

To clarify the sense of "attention" here: The concept of attention is clearly associated with overt behaviors such as eye movements, but the concept of covert attention is generally associated with some kind of top-down direct internal involvement of cognition in perception such as selecting, influencing, enhancing, or constructing the perceptual information to be supplied to cognition. That is, covert attention is some variant of *early selection* (e.g. the original Broadbent (1958) "filter" model, and later elaborations by Triesman (1969)). EPIC has no such mechanism. Rather, in a task like visual search, *all* of the available perceptual information is delivered to cognition, where the strategy decides when a response can be made, or what object needs to be fixated to collect more information. Note that this does not imply an unlimited capacity system because the amount of perceptual information is "automatically" limited by the visual system itself, such as the limits on peripheral vision or visual memory, which is why eye movements may be required to complete the task. Thus if one insists on using the language of attention, then EPIC has a *very late selection* concept of attention, and a more sophisticated process in which additional actions may be taken before the actual response is made.

This approach is the "EPIC Philosophy" — a unifying principle of EPIC modeling — a research strategy that leverages architectural commitments in EPIC to arrive at parsimonious, well-defined, executable, and accurate computational models of human performance. The key is to set priorities on which architectural mechanisms will be included in the model in order to take maximum advantage of what is well-established empirically and best-understood theoretically, such as perceptual and motor mechanisms, before including mechanisms whose status is more hypothetical and less observable, such as cognitive mechanisms beyond strategy execution. This

research strategy will be described more completely Section 4 below, in the context of *explanatory sequences*.

3.3. Specific EPIC Mechanisms Relevant to Simple Visual Search

3.3.1. Visual Mechanisms

Visual architecture. Figure 3.1 shows the overall structure of the EPIC architecture, but in simplified form; each of the processors in that diagram is fairly complex. Figure 3.2 shows the detailed architecture of the visual system. The *physical store* represents the current visual environment, e.g. what is on the screen of a display. Changes in the state of the physical visual environment are sent to the *eye processor*, which represents the retinal system and how the visual properties of objects in the physical store are differentially *available* depending on their physical properties, such as color and shape, and their *eccentricity* —the distance in degrees of visual angle from the center of the fovea (see review in Findlay & Gilchrist, 2003). The resulting "filtered" information is sent to the *sensory store*, where it persists for a fairly short time and comprises the input to the *perceptual processor*, which performs the processes of recognition and encoding. The output of the perceptual processor is stored in the *perceptual store*, which is the essentially the visual working memory and whose contents make up the visual modality-specific partition of the production system working memory. Thus, production rule conditions can test visual information by matching the current contents of the perceptual store. The visual stores are slaved to the visual environment as filtered by the current eye position. If the eye moves or the physical objects appear, disappear, change location, color, or size, the visual perceptual store will be updated to reflect the current visual scene. Because production rules can test the perceptual store contents, they can respond to this constantly updated representation of the current visual environment.

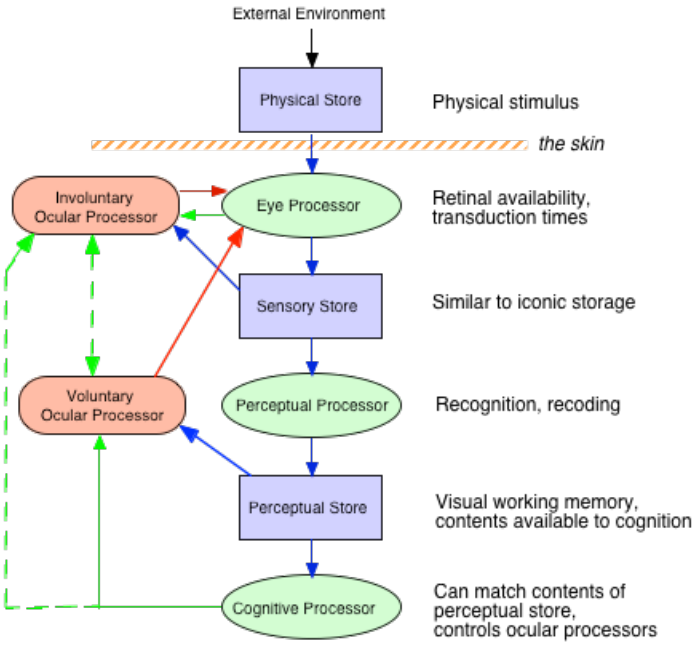


Figure 3.2. The detailed structure of the visual architecture in EPIC.

The appearance or disappearance of an object, or changes to its properties (such as its color or location), will be quickly updated in the perceptual store, but if the information is no longer supported by visual input due to an eye movement away from the object, the information persists for some time, on the order of seconds (see Henderson & Castelhano, 2005). In this way the perceptual store integrates over eye movements and maintains a cohesive representation of the current visual situation—corresponding to our subjective experience of a continuously present and integral visual surround.

A fundamental property of EPIC's visual system is that there is no built-in limit to the number of objects or their properties that can be held in the perceptual store. However, the position of the eye and the properties of the early-vision system determine which objects and what properties of them are actually present in the visual perceptual store, and thus the total information present is limited.

This flow of information through visual architecture (Figure 3.2) is the central concept in the EPIC models for visual search. Figure 3.3 illustrates this process with some sample displays from an actual EPIC model.

Visual Availability

Retinal availability functions. A common hidden assumption about vision in cognitive psychology seems to be that only foveal vision needs to be considered (cf. Findlay & Gilchrist,

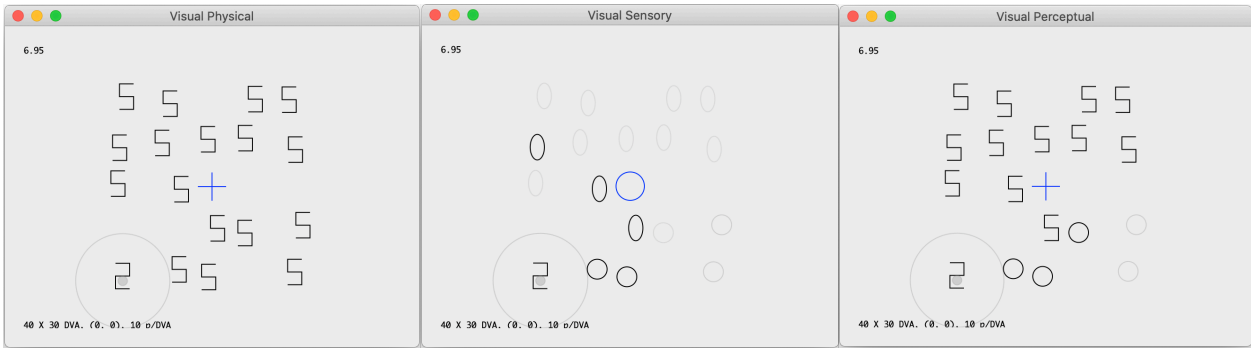


Figure 3.3. Sample model displays in a simulated positive trial for the Shape task with set size of 18, after several fixations using the Basic Search strategy (Section 3.3.3 below). These automatically-generated displays show the flow of information through the visual architecture (Figure 3.2) and are intended to support debugging and monitoring the model; they use a coding scheme that does not necessarily correspond to a person's perceptual experience.
Left panel: *the visual physical store contains the actual visual input which is provided to the eye processor. The concentric circles in the lower left show the current eye location on the target '2' shape.*
Center panel: *the visual sensory stores briefly the current object properties provided by the eye processor using the availability functions at the current eye location. Black or blue circles or ellipses are objects whose color is currently available, but not their shape. Light gray circles are objects whose properties are unavailable but location is known. Note that shape is available only for the fixated object, and color is available only for objects less than several degrees away.*
Right panel: *the visual perceptual store receives the available properties from the sensory store and retains them for several seconds, thus accumulating object properties over fixations. This contents of the perceptual store are provided to the cognitive processor production rules, so the strategy rules can test available properties of all of the objects. The strategy initially fixated several objects in the upper half of the display, whose shape and color are still retained, and finally fixated the target, whose shape then triggered the target-present response rule. The strategy will move the eyes only to an object whose shape is not present in the visual perceptual store. Note that several objects have not yet been viewed closely enough for their shape or color to be known, but the strategy halts immediately on finding the target.*

2003). However, classic psychophysical measurements make it clear that considerable information is available outside the fovea, and even into peripheral vision. More specifically, what can be perceived of an object depends not only on the eccentricity of the object but also on its size. Anstis (1974) provides some example measurements and comparisons that show that a single letter can be identified in the periphery if it is large enough. For example, the recognition threshold size of a single letter is about 0.2° at eccentricity of 5°, and about 1.3° at about 30° eccentricity. Moreover, different visual properties are differentially available in peripheral vision; for example, color is very available (e.g. Gordon & Abramov, 1977). A long-proposed neural mechanism for this relationship between eccentricity and size is *cortical magnification*: a constant amount of visual cortex (presumably supporting a certain number of receptive fields) is required for performing discrimination at a certain level, and since anatomically, the density of cortical representation declines with distance from the fovea, the size of the stimulus must increase with eccentricity to involve the same amount of cortex. Such cortical magnification functions have been measured in psychophysical experiments; a typical result (e.g. Virsu & Rovamo, 1979) is that to maintain discriminability, the required size increases linearly up to a moderate eccentricity (e.g. 30° in Anstis, 1974) and then quite sharply in the further periphery, with a cubic function providing a good fit.Visual search studies such as Carrasco & Frieder (1996) using short exposure duration show that if object size is constant, then targets at greater eccentricity are found more slowly and less accurately, but if peripheral objects are magnified in size according to the empirical magnification functions, search time becomes flat with eccentricity.

Unfortunately, the available psychophysical literature does not use a standard set of stimulus properties. For example, orientation is a common property used in both vision and visual search experiments, but the range of possible orientations is very large, and the stimuli for such experiments range from very short line segments slightly tilted to left or right, to very large horizontal and vertical bars, and even circular Gabor patches. It is rare to see experiments from different laboratories using even similar, much less the same, stimuli so it is impossible to combine empirical results into a set of parametric functions that describe the detectability of object properties as a function of size and eccentricity. Thus EPIC's retina processor uses availability functions of a certain form based on the psychophysical literature, and the parameters of the functions have to be estimated to fit the data being modeled. However, the psychophysical results do set some constraints — for example, the color of an object can be expected to be more available than its orientation, and much more available than its shape.

In the models reported here, the maximum eccentricities are less than 30°, which the results in Antis (1974) suggest that a simple linear relationship between detection threshold for object size and eccentricity can used. The availability function is thus a Gaussian detection function that gives the probability that a specific property will be available (or detected) for an object with size s at eccentricity e:

$$P(detection) = P(s > N(\mu, \sigma)),\ \mu = \theta e,\ \sigma = 0.5$$

The value μ, the mean of the Gaussian function, can be interpreted as the 50% threshold for object size. It depends linearly on the eccentricity proportional to θ which is the *availability threshold coefficient*. Thus small values of θ correspond to the property being more available (lower threshold, more eccentricity possible for a given size), and large values of θ mean that the object is less available (higher threshold, must be closer for a given size). The standard deviation

σ determines the "steepness" of the detection function, and was held constant at 0.5. Thus only the θ parameter was adjusted to fit the observed data. For simplicity, the specific values for each property are assumed to have the same availability; e.g. a Color[4] property value of Red is assumed to have the same θ as a Green value. Finally, the size of the objects used in these displays is defined as the average of the vertical height and horizontal width (bounding box) in degrees.

Figure 3.4 shows the availability functions for some representative values for θ used in models for the Wolfe et al. (2010) experiment presented below; namely $\theta_C = 0.1$ for Color, $\theta_O = 0.2$ for Orientation, and $\theta_S = 0.4$ for Shape. Figure 3.4 also shows a couple of useful eccentricity metrics relevant to these models. The *average initial eccentricity* corresponds to the eyes being on the initial fixation point when the display appears. The *average pairwise eccentricity* is the average distance, for all possible fixated objects, between the fixated object and all of the other objects. Color is very available; its detection probability is high throughout the eccentricity range. Orientation is significantly less available; at the average pairwise eccentricity, the probability of detection is only about 0.4. Finally, Shape is not very available at all; even at the average initial eccentricity, the probability of detection is almost zero. Thus fixations within a few degrees will be required to reliably detect the Shape. It will be argued below that the Shape property, with its apparent complex substructure of line segment features, can be treated in the models as if it were a *unitary* simple property like Color or Orientation; the fact that it is more difficult to discriminate is reflected in its larger availability threshold coefficient.

No bottleneck on availability for multiple objects. In this model of visual availability, the properties of more than one object can be available in the perceptual system for a given eye location. No "attentional limit" is assumed to operate at this level of the visual system. If the availability is high enough (a low θ threshold), then the properties of multiple objects, even at large eccentricities, will be available. For example, Hornof & Halverson (2003) recorded eye movements in a menu-search task, and determined that 2-3 of the textual menu items were examined in each fixation and were able to account well for the eye movement and search time

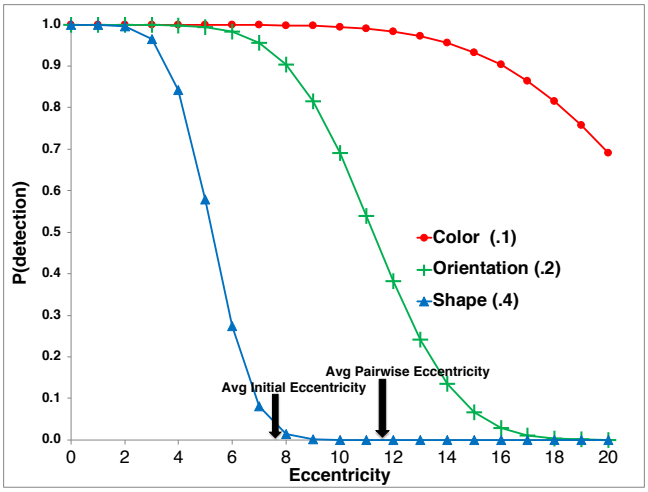


Figure 3.4. Example availability functions for the Color, Orientation, and Shape properties used in the simple visual search task.

[4] Names of perceptual properties and their values, such as Color, Red, Orientation, Horizontal, or Shape, that are explicitly represented in the architecture and models are capitalized.

data with an EPIC model whose visual availability parameter was set to cover this number of textual objects in a fixation.

This feature of EPIC's visual system is similar to the long-standing concept of the *functional visual* (or *viewing*) *field* (FVF, reviewed in Hulleman & Olivers, 2017), in that both concepts allow more than one object to be processed in a single fixation. However, the FVF concept assumes that there might be other factors that govern the number of objects covered by a fixation, including possible attentional limitations, or purely visual factors, such as *crowding effects*, to be discussed next. In contrast, the EPIC availability functions determine only whether object properties are available for further processing given their size and eccentricity, independent of any other visual factors such as crowding effects.

Availability and eye movements. The availability for each property is independently resampled for all objects when the display first appears and whenever the eye is moved. As the eye moves around, the available properties of a particular object can fluctuate, and will not be reliably available from one fixation to the next. However, the properties, once acquired, will remain for some time in the visual perceptual store, where production rules can match objects with the combinations of properties appropriate to the task. Figure 3.3 above provides an overview of this process.

The availability and eye movement mechanisms are thus relevant to modeling the Wolfe et al. (2010) data. That the Shape, Orientation, and Color properties have different availabilities is a good candidate to explain the great difference in search rates between the task conditions, because less availability means that more eye movements would be required to make the properties available. As pointed out above, because the eye movements were unmonitored, there is no reason to assume that subjects followed the instructions to remain fixated on the supplied fixation point, or that they could have performed the task with any accuracy if they did.

Visual Crowding

Crowding effects. *Crowding*, also known in older literature as the *flanker effect* or *lateral interference*, refers to the phenomenon in which the perception of an object is impaired if it is surrounded by closely-spaced objects, but the same object is perceived accurately if the spacing is larger or the eccentricity is smaller (for reviews, see e.g. Levi, 2008; Pelli & Tillman, 2008a, b; Rosenholtz, 2016). Crowding was first described for the recognition of characters in reading by Bouma (1970), and Anstis (1974) provides examples for character recognition. Pelli & Tillman (2008a,b) provide many demonstrations of the effect. Crowding effects appear if the center-to-center spacing is less than a *critical spacing*, which for a wide variety of visual properties turns out to be approximately half the eccentricity of the object in question, a relationship first reported by Bouma (1970). Thus if a set of closely spaced objects is viewed at a large eccentricity, crowding might impair perception of them, but by moving the point of fixation closer, the critical spacing becomes smaller, and the crowded objects could be perceived correctly. This simple characterization of crowding will suffice for the work in this paper, but Strasburger (2020) describes important complicating issues and clarifications for crowding and peripheral vision.

Crowding in simple visual search. Levi (2008), Pelli (2008), and Rosenholtz (2016) argue that taking crowding into account is essential to understanding extra-foveal vision because it appears to be more responsible for the limitations on peripheral vision than simple loss of

resolution. The reason why crowding might be relevant to simple visual search tasks is that these experiments *almost always confound the number of objects on the display with object spacing*. The typical experiment, e.g. Treisman & Gelade (1980), Wolfe, Cave, Franzel (1989), Wolfe et al. (2010), varies the set size while keeping the display area constant, placing the objects at random within the display area, and thereby producing higher object density at higher set sizes. The few studies attempting to separate crowding and set size effects suggest that much of the reported set size effects in visual search could in fact be due to crowding rather than simply the number of objects (e.g. Motter & Simoni, 2008; Wertheim, Hooge, Krikke, & Johnson, 2006).

In contrast, Wolfe et al. (2010) simply asserted that the stimuli "can be easily identified outside of the fovea" in these tasks (p. 1305) but report no measurements about whether this is true in their higher-density displays; this seems unlikely at least in the Shape condition, and even the Color task condition produces more Misses than False Alarms, which implies that there is some source of systematic difficulty in what would otherwise seem to be a trivial task.

To assess whether the Wolfe, et al. displays confound crowding with set size, two measures of crowding were computed for a large number of displays randomly generated by the model. The first measure assumes that the eyes were at the initial fixation point, and the eccentricity from this point was calculated for every object on the display. The second is a pairwise measure: assume that the eyes were fixated on one of the objects, and then for every other object, the number of surrounding objects within the critical spacing (defined as half of the eccentricity) was counted. This number was computed for each object being the assumed point of fixation, and the results averaged. Figure 3.5 shows the resulting average number of crowding objects for each measure. When the eye was assumed to be at the initial fixation point, the mean initial eccentricity from the central initial fixation point was 8.8° and the number of crowding objects increased from an average of 0.1 at set size 3 up to 1.1 for set size 18. When the eyes were assumed to be on each object, the average pairwise eccentricity was 12.3°, and the number of crowding objects increased from 0.15 at set size 3 to 2.5 at set size 18.

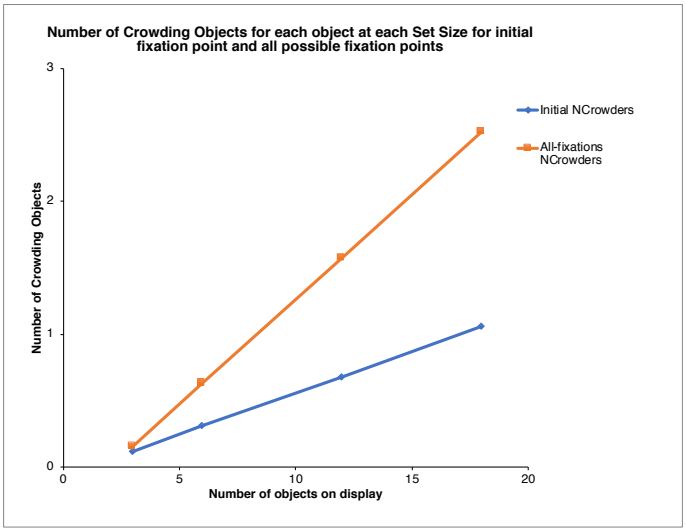


Figure 3.5. Crowding in the Wolfe, et al. (2010) displays. The upper curve shows the average number of objects that crowd another object assuming a fixation on each possible object. The bottom curve shows the average number of objects that crowd another object assuming the initial fixation point.

Thus crowding effects are likely to present in this simple visual search task — at small set sizes, almost no crowding is present, and at larger set sizes, some crowding is present at the initial fixation point, and substantial crowding appears at once the eyes are moved away from the initial fixation point to other objects. Simple visual search experiments have indeed confounded their basic manipulation with an unacknowledged, but powerful, factor in visual perception, especially if the eyes are moved from the initial fixation point.

Accordingly, models of simple visual search should incorporate crowding effects. Note that experiments on complex visual search tend to use a large and constant number of objects, so that crowding effects would tend to be uniform and probably indistinguishable from availability effects.

How does crowding work? The mechanism of crowding effects remain unclear despite the extensive literature, but there seems to be a consensus on a basic idea (e.g., Pelli, et al. 2004; Pelli & Tillman, 2008a; Levi, 2008; Rosenholtz, 2016; Rosenholtz & Keshvari, 2019) that the visual system attempts to form visual objects by integrating information over integration fields, areas whose size increases with eccentricity (c.f. the cortical magnification concept). If more than one stimulus object occupies a single such integration field, the integration process will be disrupted in some way. But if the point of fixation is closer, the smaller size of the integration fields will allow the same visual objects to be correctly formed. The problem is that the empirical work has not clarified the basic rules for the integration process, the actual results of crowding disruption, and how it relates to the visual availability of the relevant properties.

More specifically, a basic psychophysical measurement of crowding presents a target object with at least one crowding object on each side, and measures how well subjects can identify the target object properties as a function of eccentricity, object spacing, and the visual properties of the target and crowing objects. These experiments generally ask subjects to report on the target object, which is specified by its central location in the display — so the measurement depends on the whether a property is perceived at that location.

Many experiments measure the *amplitude* of the crowding effects in terms of an elevation of detection or discrimination threshold of the relevant property of a crowded target object as a function of crowder spacing. Different properties have differing threshold elevations; the available literature suggests that color has a low crowding amplitude and is thus relatively resistant to crowding effects, while the detailed shape of characters has a high crowding amplitude. In addition, the greater the similarity of flankers and targets, the larger the crowding amplitude. Other experiments simply ask the subject to identify the property (e.g. color) of a central target object surrounded by flankers, and vary the eccentricity or spacing, and report differences in accuracy.

What is unclear is what these experiments are actually measuring; if performance decreases in the presence of crowding, is it due a simple loss of availability of the target object property (i.e. an increased θ in the presence of crowding), or is it that a flanker property tends to be confused with the target property? This second possibility is a popular hypothesis in the literature: The existence of the crowded object is still detected, and its basic perceptual features also are detected, but the disrupted integration process associates those features with the wrong object, such as a flanking object, and vice-versa — the features are essentially *scrambled* between the objects that crowd each other. Such scrambling seems apparent in demonstration

examples of crowding (e.g. Pelli & Tillman, 2008a, b) , and would account for the detection or discrimination tasks results. Better evidence is found in studies that examine the nature of the errors when reporting the target property; misreporting a flanker feature as a target feature accounts for most of the errors, although there are complications (e.g., Keshvari & Rosenholtz, 2016; Põder & Wagemans, 2007; Yashar, Xiuyun, Jiageng, & Carrasco, 2019).

In terms of a specific computational mechanism for crowding effects, Rosenholtz and coworkers (Balas, Nakano, & Rosenholtz, 2009; Rosenholtz, Huang, & Ehinger, 2012; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012; Rosenholtz, Yu, & Keshvari, 2019) propose that the "blurring" of peripheral vision can be described as a transformation that preserves the overall *texture*, or statistical structure of the image, but can result in images in which object features have been lost, recombined, fragmented, or displaced in location, similar in some ways to the scrambling concept. They report that the discriminability of such transformed images predicts the search time effects in simple visual search tasks. Also, subtle changes to the stimuli that decrease the disruption produced by the transformation results in faster search times (Chang & Rosenholtz, 2016). But a texture-based computational cognitive model that predicts RT (and ER) directly from the visual input still remains to be developed. One obstacle is that the texture-preserving transformation is very computationally intensive.

In summary, the actual mechanism of crowding appears to be very complex and subtle. However, any computational cognitive modeling necessarily requires simplifications in return for the insights that can result. Accordingly, the modeling in this report with the EPIC architecture took advantage of how the Wolfe et al. (2010) displays consist of simple discrete objects with simple discrete properties, and used simple strategies for search; all of which enable the implementation and testing of a very simple form of the feature scrambling concept.

A simple crowding mechanism. In this simplified approach, the properties of the visual objects are determined in two steps that are applied whenever the visual situation changes, specifically when the display appears or the eyes are moved. First, the visual architecture applies the availability functions to determine which properties are currently available for each object from the current eye position. Second, a crowding mechanism in the perceptual processor randomly scrambles the values of each property between objects that are within the critical spacing of each other. To parameterize the magnitude of the crowding effect, the scrambling is performed with a certain *crowding probability* ϕ for a property of each object. Note that if $\phi = 0$, then no crowding scrambling for that property will occur. The value of ϕ can differ depending on the property involved, but is the same for all values of that property. As noted above for availability, and will be argued below, the models can treat Shape as a unitary property like Color and Orientation, and thus the same scrambling mechanism for crowding applies to Shape as well as Color and Orientation. Representative estimated values for the models in this report are $\phi_C = 0.025$ for Color and also $\phi_O = 0.025$ for Orientation, both of which have two very distinct values, and $\phi_S = 0.1$ for Shape.

More specifically, in the actual visual system it is reasonable to suppose the scrambling will happen simultaneously in some way for all of the objects, but the simple scrambling mechanism works sequentially as follows: (1) Using the availability functions, the available perceptual property values for each object are determined for its current eccentricity. If a property is unavailable for an object, the property is assigned a *blank* value. (2) Using the current eye position and object locations, the objects that are within the critical spacing of each object are

collected into a *crowding group* for that object. The crowding relationship can be asymmetric: Suppose object A has a smaller eccentricity than object B; Object A might not have B in its crowding group, but object B might have A in its crowding group. (3) For each property (e.g. Color), and each object in order of increasing eccentricity, a "coin flip" is performed for that object, and with that probability ϕ the property values of all of the objects in that crowding group are collected (including blank property values), randomly shuffled, and then assigned back to the objects in the crowding group. Thus the actual numbers of available and unavailable property values are not changed; rather those values are scrambled for the objects in the crowding group. (4) If an object has no crowders (i.e. it is closely fixated or relatively far from other objects), and its properties are available, these properties then become "sticky" in the visual perceptual store, and will not be lost or replaced by a blank property, but could still be scrambled in the future with available properties of crowding objects.

Note that the crowding probability ϕ applies at the level of each object. For example, a crowding group of four objects will have the crowding probability "coin flip" and scrambling applied a total of four times, once for each of the objects. Thus the probability of at least one scrambling operation being applied increases with the number of objects in the crowding group.

Illusory targets, distractors, and blanks in simple visual search

As stated above, the availability mechanism and then the crowding mechanism is applied when objects first appear, and then again after each eye movement. During the course of a simple visual search trial, as the eyes are moved around, the sticky perceptual store properties accumulate, and more objects acquire properties, either from becoming available due to nearby fixations, or from scrambling from nearby objects. During this process, the property value for an object might get replaced by some other object's property value. One case is that a distractor object might get a target property and thus become an *illusory target*. Another case is that the target object might get a non-target value, becoming an *illusory distractor*. A final, more subtle case is that a target object might get a blank property even though its actual property value would otherwise be available, and thus becoming an *illusory blank*. Illusory distractors are more likely than illusory blanks, because if the point of fixation is close enough for the target property to be available, there is a good chance that a nearby distractor object also has its property available for swapping.[5]

3.3.2. Motor Mechanisms

Oculomotor accuracy and movement time

Production rules can send commands to the *involuntary* and *voluntary ocular processors* shown in Figure 3.2 that control the position of the eyes. The involuntary ocular processor generates "automatic" eye movements, such as reflex saccades to a new or moving object, or

---

[5] Despite the similar terminology, the illusory objects produced by this model of crowding are different from the *illusory conjunctions* concept presented by Triesman (Triesman & Schmidt, 2007; see also Wolfe, 2014). In the present model, features of objects are assigned early in visual integration and can be scrambled due to purely visual factors that disrupt the integration, all independently of attention. The illusory conjunction concept assumes that attention is required to perform this integration, and in its absence, free-floating features might be erroneously attributed to objects. Such illusory conjunctions appear to be more common if the objects are closely spaced, which suggests the concept is simply an attempt to explain crowding effects as a function of attention rather than visual processing (cf. Pelli, Palomares, & Majaj, 2004).

smooth movements to keep the eye foveated on a slowly moving object. The involuntary processor is not involved in the models presented here and so will not be further described. Critical to the models in this paper, the voluntary ocular processor is directly controlled by the cognitive processor. A production rule action can command an eye movement to the location of a designated object represented in the perceptual store.

The voluntary oculomotor processor includes motor "noise" that affects saccade accuracy. A variety of studies have shown that saccades tend to fall short of the actual fixation target, and the standard deviation of the saccade length tends to be proportional to the length (see Abrams, Meyer, & Kornblum, 1989, and the review in Harris, 1995). Thus, the oculomotor processor samples the length for a saccade to an object at eccentricity *e* from a Gaussian distribution:

$$saccade\ length = N(\mu, \sigma),\ \mu = g \cdot e,\ \sigma = s \cdot \mu$$

Typical empirical values for *g* (*gain*) range from 0.85 - 0.95, and *s* (*spread*) is typically around 10%. Harris (1995) did some modeling work that showed that given the variability in saccade length, and the resulting need to make multiple saccades to ensure fixation on an object, optimum total eye movement times to a target were obtained with $g = 0.95$, $s = 10\%$, which are consistent with observed values of these parameters. These values are used in EPIC as the default values for oculomotor noise along the line of flight of the saccade. Angular error might also be present; a saccade might not only fall short, but it might also miss to one side. Unfortunately, there are very few studies on angular error; a simplified model inspired by van Opstal and Gisbergen (1989) samples the saccade polar angle from a Gaussian distribution whose mean is the actual angle and whose standard deviation is a constant, currently defaulting to simply 1°.

The time duration of a saccade was determined using the classic linear function described by Carpenter (1988):

$$saccade\ duration(ms) = 21 + 2.2 \cdot saccade\ length(degrees)$$

The fixed intercept in this function was taken to represent the time to initiate the saccade; therefore the movement initiation time parameter in the original EPIC oculomotor processor was set to zero.

Manual motor time and accuracy

The time to make a keypress response is represented by an architectural mechanism first proposed in EPIC for use in models of the psychological refractory period task (Meyer & Kieras, 1997a). In these high-speed tasks, the subject has a finger poised over each response key so only a rapid finger flexion is required. The manual motor processor uses a motor programming concept that producing a movement of this type requires selecting motor features that specify the hand and finger, then initiating the actual movement, which takes a certain amount of time to close the key switch (see Kieras, 2009, for additional discussion of this concept). Each feature was estimated as requiring 50 ms to select; the initiation time was also estimated at 50 ms, and the movement time at 25 ms. The motor features can be reused if identical, making the next movement faster.

Wolfe et al. (2010) do not provide any details about the response keys actually used, making it necessary to assume the parameters. Since present and absent responses were approximately equally probable, it is reasonable to suppose that the motor time on the average for present and

absent responses would be the same in terms of the average number of feature reused. Thus the time contributed by the manual motor processor was set at 125 ms for both present and absent responses in all task conditions and set sizes.

The models presented here account for error rate (ER) as well as RT. As discussed more below, one component of ER is a low and constant rate of *action slips*, or "oops" errors, a basic error mechanism that is often postulated in human performance research. Namely, a person has the correct intention for a response, but at random, an incorrect motor action will be triggered. This most basic of error sources can be represented very simply for the manual present/absent responses; it is as if one of the motor features were "flipped" to a different and frequent value. Thus, when the strategy calls for a present or absent response, the *opposite* response is made with a slip error probability *SlipER*. This will be included in the discussion below on how the models make errors in the simple visual search task.

3.3.3. Cognitive Task Strategy Mechanism

As discussed above, in the EPIC architecture, the only thing the cognitive processor does is execute a task strategy in order to support task-specific programmable behavior. The strategy is represented in terms of production rules. In the visual search task, these rules will examine the contents of the visual perceptual store and make decisions about what actions to take, such as noting which objects are possible targets, moving the eyes to an object of interest, or making a keypress response. This component of the architecture is thus a "mechanism" like those already presented, but with the understanding that rather than setting numerical parameters to modify the activity as in the case of the perceptual mechanisms, specific strategies are loaded into the cognitive processor to govern how the task is performed. Thus describing how the strategy mechanism can be applied to visual search tasks requires presenting the possible relevant strategies for doing simple visual search search.

The strategies for simple visual search presented here are based on previous EPIC models for visual search (Kieras, 2016; Kieras & Marshall, 2006; Kieras & Hornoff, 2014). These strategies can be easily summarized in pseudo-code, making it unnecessary to provide the technical details of the production rule syntax and the specific production rules used in the models (see Kieras, 2016 for detail and some examples). The following sections presents the different strategies used in the simple visual search models to be described below; this is to enable a compact presentation of how the models were fit to the data.

The Basic Search strategy

The simple visual search task can be performed with the *Basic Search* strategy shown as pseudocode in Box 3.1. Once the display objects appear on the screen, after a delay time *VDelay* held constant at 100 ms in all of the models, the available properties of the objects are present in the visual perceptual store, and the strategy production rules then alternate between two phases: In the Step 3 *nomination phase*, *nomination rules* fire to nominate objects (including those in peripheral vision) that are either the *actual target*, or *possible targets* because at least one relevant target property is blank (unknown). In the Step 4 *choice phase*, specific rules fire depending on the nominations to take one of three actions: (a) If an actual target object has been nominated, a target-present response is immediately made via a command to the manual motor

processor. (b) If there are no nominations, meaning that all objects appear (even in peripheral vision) to be distractors, then a target-absent response is immediately made. (c) Otherwise, there are only possible-target nominations, so an oculomotor processor eye movement command is issued to move the eyes to the *closest* nominated object. Once the eye movement is complete, the nomination phase starts again at Step 3. The implementation is such that each numbered step in Box 3.1 corresponds to a single production rule cycle; thus the strategy is implemented in the minimum number of production rule cycles possible in the architecture and the production rule syntax.

1. Start the trial with the eyes on the fixation point and wait for object display to appear.
2. Delay for some time.
3. Nominate target or possible target objects using their available properties.
4. Choose an action:
a. If a nominee matches the target, respond present, trial is done.
b. If there are no nominees, respond absent, trial is done.
c. Otherwise, choose the closest candidate object from the nominees, and move the eyes to it, and go to step 3.

Box 3.1. Pseudocode for the Basic Search strategy.

Basic Search as an optimal strategy. The Basic Search strategy is essentially the "fastest reasonable" way to perform the task. It is "fastest" because it takes advantage pf extra-foveal vision: it may not be necessary to fixate each object to know whether it is the target or a distractor. Thus a target-present response is produced as soon as a target is detected, even if it is not fixated, and a target-absent response is produced as soon as all objects appear to be distractors, regardless of whether they have all been fixated. This strategy is "reasonable" because the response will be accurate, limited only by the accuracy of the perceptual information and slips in the motor action.

Thus, as fixations are made, information about the objects accumulates in the perceptual store (as illustrated in Figure. 3.3 above) until either the target object is detected, or the available properties of all objects show that none of them could be the target. The RT depends on how many eye movements are made in this process. Unlike many models which do some form of time-out for making target-absent responses (cf. review in Hulleman & Olivers, 2017), the Basic Search strategy states simply that if there are no possible-target nominations (i.e., *everything looks like a distractor*), then a target-absent response should be made immediately — no time-out is required.

Strategy variations

In general, the choice of strategy has a large effect on whether a model can fit the data, and a satisfactory fit can only be obtained by a good choice of both parameter values and strategy. As argued above, the Basic Search strategy is a "fastest reasonable" strategy that produces fast responses which will be accurate to the extent that the perceptual and motor systems are accurate. However, some variations on this strategy will be needed for good fits to the data; these are presented here to make the later model-fitting descriptions more compact.

Fixed-Eye Strategy. The extremely simple strategy shown in Box 3.2, is a stripped-down version of the Basic Search strategy. The eyes are left on the central fixation point and the target-

1. Start the trial with the eyes on the fixation point and wait for object display to appear.
2. Delay for some time.
3. Nominate target and possible target objects using their available properties.
4. Respond present if target nominated, respond absent if not, trial is done.

Box 3.2. Pseudocode for the Fixed-Eye visual search strategy.

present or target-absent response is chosen after a single nomination phase. This strategy corresponds to the common instructions in visual search experiments, namely to keep the eyes on the fixation point. Whether subjects can or do follow this instruction and apply the Fixed-Eye strategy is another question.

Basic Search with Confirm-positive strategy. This is a more elaborate version of the Basic Search strategy that provides protection against erroneous target-present responses when crowding has produced an illusory target. It is a simple and well-specified form of "double-checking" that is sometimes proposed in visual search models (cf. Hulleman & Olivers, 2017). As shown in Box 3.3 in Step 4a, if the apparent target object is already fixated (eccentricity <= 1°), a present response is made; if not, the strategy confirms that an apparent target is the actual target by moving the eyes to it, which would mitigate any crowding, and responds present if the apparent target appears to be an actual target, or continues the search if not. This confirmation takes extra time, both for the eye movement and additional production rule cycles.

1. Start the trial and wait for object display to appear.
2. Delay for some time.
3. Nominate target and possible target objects using their available properties.
4. Choose an action:
a. If a nominee matches the target, and it is already fixated, respond present, trial is done. If not already fixated, confirm by moving the eyes to the nominee: If the object is the target, respond present, trial is done; otherwise, go to step 3.
b. If there are no nominees, respond absent, trial is done.
c. Otherwise, choose a candidate object from the nominees, and move the eyes to it, and go to step 3.

Box 3.3. Pseudocode for the Basic Search with Confirm-positive strategy.

Basic Search with Confirm-both strategy. This strategy, shown in Box 3.4, elaborates a bit more on "double-checking" to handle the possibility that all objects erroneously appear to be distractors because crowding has made the actual target into an illusory distractor. But unlike the Confirm-positive case, it is undefined which object is possibly illusory. A simple approach is that if no fixations have yet been made, the strategy moves the eyes to the most eccentric object, which is mostly likely to be affected by crowding, and confirms that it is a distractor. The Confirm-both variation is only useful in a context where Confirm-positive is relevant.

Limited-fixations strategy option. This variation, shown in Box 3.5, provides a way to speed up the Basic Search strategy or one of the Confirmation variations. Instead of moving the eyes an unlimited number of times until a response is made, the strategy "times out" and responds absent if some time has elapsed without finding the target. This approach is often proposed as an error mechanism in other models (cf. Hulleman & Olivers, 2017). Since Basic Search moves the eyes at a roughly constant pace, the "time out" was implemented more simply as a limit, *NFix*, on the

1. Start the trial and wait for object display to appear.
2. Delay for some time.
3. Nominate target or possible target objects using their available properties.
4. Choose an action:
a. If a nominee matches the target, and it is already fixated, respond present, trial is done. If not already fixated, confirm by moving the eyes to the nominee: If the object is the target, respond present, trial is done; otherwise, go to step 3.
c. If there are no nominees, and more than one fixation has already been made, respond absent, trial is done. If no fixations yet, confirm by moving the eyes to the most eccentric object: If the object is a distractor, respond absent, trial is done; otherwise, go to step 3.
d. Otherwise, choose a candidate object from the nominees, and move the eyes to it, and go to step 3.

Box 3.4. Pseudocode for the Basic Search with Confirm-both strategy.

number of eye movements. This option can be applied to any variant of Basic Search strategies at each step prior to initiating an eye movement.[6] Note that the initial fixation on the fixation point at the beginning of the trial is not counted.

Before making an eye movement, check:
a. If the number of fixations made thus far is greater than *NFix*, respond absent, trial is done.
b. Otherwise, proceed.

Box 3.5. Pseudocode for the Limited-fixations strategy option.

Visual properties and the Basic Search strategy

Nomination rules. The above presentation of availability functions and the crowding algorithm implied a choice of the relevant visual properties for the Wolfe et al. (2010) tasks. It seems obvious that for the Color and Conjunction tasks, the relevant properties are the traditional Color and Orientation features, that if available, have one of two values. The nomination and choice rules are thus very simple for the Color task because only a single object property is involved: An object is nominated as the target if it has a Red Color, or as a possible target to be fixated if it has an unknown Color. If all objects have a Green Color, then no objects can be a target, so there are no nominations, leading to an immediate absent response.

In contrast, for Conjunction there are four possible nominations: a target nomination for a Red Vertical bar, and three cases for nominations for possible-target objects: Red Color & blank (unknown) Orientation, blank Color & Vertical Orientation, and blank Color & blank Orientation. If a target has been nominated, a present response is made; if one or more possible-target nominations are present, the strategy should choose one to fixate in the descending priority order as just listed. Finally if there are no nominations because all objects appear to be either Horizontal Red or Vertical Green, an immediate absent response is made.

Shape can be treated as a unitary property. The description of the availability functions in Section 3.3.1 above imply that Shape is treated as a unitary property just like Color and Orientation, only less available, and the presentation of the crowding mechanism states that it

[6] This option can be "turned off" for the Basic Search strategy variants simply by setting NFix to an extremely large value (e.g. 99) that is never exceeded in the model runs. While the Fixed-Eye strategy corresponds to NFix = 0, it is simpler to implement it without reference to the number of fixations.

applies identically for Shape as it does for Color and Orientation, only with a higher probability. This seems counterintuitive, given that the shapes used have a detailed substructure of line segments. However, when the Basic Search strategy is in effect, Shape functions as if it were a single and unitary property.

This can be demonstrated as follows: Start with the assumption that an object's Shape property consists of subfeatures. For example, a common idea is that characters are made up of line segment sub-features, each of which must be detected and then subject to crowding scrambling. So perhaps the partially available Shape for the "digital 2" and "digital 5" should consist of a subset of the seven possible line segments where each, if available, is either horizontal or vertical and in a particular location within the object, producing many possible *partial* shapes. Alternatively, perhaps the subfeatures are a leftward- or rightward-facing "digital c" each with a top or bottom location. Availability and crowding scrambling would produce different partial shapes such as a *c,* reversed *c*, a *3*, or *E*. Unfortunately the vision and crowding literature does not provide much guidance on the relevant properties of characters — defining the hypothetical features of visual objects has always been fairly speculative (c.f. Wolfe, 2014); any assumptions along these lines will be arbitrary and lead to complexity in the model.

However, the Basic Search strategy justifies a great simplification: Since these partial and jumbled shapes match neither a target nor a distractor, the strategy arguably should treat them as *possible targets* to be fixated to determine what they actually are, just as if they had a blank shape. This means that the Shape property can functionally be treated as a unitary property such that each perceived object has a Shape property with a value that is either *2*, *5,* or *blank*. The Shape task can thus be modeled as a single-feature task with nomination and choice rules for the Basic Search strategy that are as simple as those for the Color task: If a *2* is visible, a target is nominated and a target-present response is made; if a blank is visible, it is a possible-target nomination and a possible candidate for fixation. If all objects appear to have a *5* shape, then no nominations are made, and the response is target-absent. The availability of Shape can thus be represented with a detection function whose threshold θ_S is higher than that putatively involved with detecting the hypothetical individual subfeatures. Crowding will scramble these unitary Shape property values according to the same algorithm as for Color and Orientation.

How the visual search models make errors

As pointed out in the discussion of the Wolfe et al. results, there are clearly systematic effects in the ER data as well as the RT data that should be accounted for. A model that attempts to account for both RT and ER is not common in human performance research. Since this effort attempts to use both RT and ER as equal-status indicators of visual search processes, it is important to be clear on how the models make errors. As mentioned above, the Basic Search strategy will not itself make an error, so errors can happen only from the following three sources in the models to be presented:

Action slips. The presented strategies will not "deliberately" attempt to respond present on a negative trial, so False Alarm errors must be due to some factor different from the visual and strategy mechanisms already described which corresponds to the low and very stable rate of False Alarm errors across set size and search task. These errors can be attributed to the *action slips* or "oops" errors described in Section 3.3.2. In these models, if the strategy calls for a present or absent response, the *opposite* response is made with probability *SlipER*. This will

produce both False Alarms and Misses, but with a constant probability across search tasks, trial polarity, and set size. Note that slip errors do not affect the correct trial RT distributions since they are independent of the actual correctness of the response.

Premature termination of search. Misses could also be produced by one of the Basic Search strategy variations that limits the number of fixations and terminates the search with an absent response after some number of eye movements. Thus a Miss error results if the strategy terminates a positive trial with an absent response before the target has become visible. This is more likely if there are more objects on the display, and fewer visible properties. This means that Miss ER would increase with set size and apparent task difficulty. The Fixed-Eye strategy can be considered an extreme version of such a strategy in that zero eye movements are allowed before responding.

Illusory distractors from crowding. An important source of Misses is when the strategy rule that detects the absence of nominations fires when the target is in fact present on the display. This would happen if all of the objects appear to be distractors. This will be exactly the situation if crowding scrambling turned the target into an illusory distractor and at the same time, all of the other objects appear to be distractors. Thus Misses would increase with set size due to more crowding, and possibly with less available properties.

Thus the basic logic of the strategies together with some combination of action slips, limited fixations, and crowding effects could account for the basic features of the ER results, namely the stability of False Alarm rates, the increase of Misses with set size, and the increase of Misses with apparent task difficulty.

4. Models for the Aggregate Data

All of the theoretical concepts have been presented, so the remaining sections of this paper present how the models built using the cognitive architecture can account for the data. However, some important methodology issues must be discussed before commencing with the modeling results.

4.1. Model Development Methodology

Three methodological topics are presented in this section. The first concerns the conceptual approach to developing models that fit the data following the EPIC Philosophy. The second and third concern technical details on deriving and evaluating predictions from the models

4.1.1. Fitting Models using Explanatory Sequences

It is common in modeling work to simply show the final good-fitting model with the predicted and observed data, the corresponding parameter values, and the goodness-of-fit metrics, and declare victory. However, modeling work is very much more informative if the good fit can be shown systematically to be due to a particular combination of the architectural mechanisms in the model, the parameter values, and the strategy.

To demonstrate the basis of a good fit requires presenting models that *fail* to fit the data because they don't have the necessary mechanisms, parameter values, or strategy[7]. These make it clear that the final model fits the data not because it has so many "degrees of freedom", but rather because it has the minimum number of mechanisms and fewest adjustable parameters necessary. This can be shown with an *explanatory sequence* of models that shows the effects of adding mechanisms, parameter adjustments, and strategy variations, *one at a time*, to demonstrate how each contributes to an increasingly better fit. While the presentation of a model with an explanatory sequence is lengthy, the sequence makes it clear that the resulting model is neither arbitrary nor haphazard, but is really the best explanation of the data in terms of the architectural assumptions.

In cognitive architecture models, we start with a set of "hardware" mechanisms and their parameters and use them in conjunction with the "software" of task strategies. In EPIC, the "hardware" mechanisms are the given "black-boxed" and parameterized sensory-perceptual and motor systems, and the well-defined and limited production-system engine of the cognitive processor, while the "software" is the cognitive task strategy represented as the specific production rules applied by the cognitive processor to choose what motor actions to perform based on the perceptual data in order to meet the task requirements. In this approach, the "hardware" is assumed to be fixed, but the parameters have to be estimated, and the strategy, although it is constrained by the task requirements, is free to vary due to both procedural details of the experiment and the preferences of the subjects. So an explanatory sequence of EPIC models will show the effects of changing both the parameters that govern the given architecture components, and the strategy used in the task. Parameter estimates and strategy choice can interact; it is often very informative to hold one constant while varying the other. But a key idea from the EPIC Philosophy is to give first priority to including the known effects of the fixed perceptual and motor mechanisms with their estimated parameters, followed by varying the strategy as needed to match the data.

Constructing an Explanatory Sequence

In more detail, an explanatory sequence is constructed in steps as follows:

Setup: Create an initial model for performance in the task. This model contains the minimum and simplest architectural mechanisms necessary for the task. These include a strategy executed by the cognitive processor that uses specified outputs of the perceptual processors to decide how to command the motor processors in order to perform the task. The perceptual-motor processor parameters can be initially estimated from the literature or previous models of similar tasks, and the strategy can be chosen on the basis of performing the task to meet the requirements nominally or optimally.

Refinement: Iterate over choice of parameter values, strategies, and additional mechanisms. It is unlikely that the initial model performance matches the observed data. In order to fit the data, make systematic and step-by-step modifications to the model in the following *explanatory priority order*:

1. First priority: Include known perceptual and motor mechanisms and modify their parameter values over plausible ranges. That is, as much as possible, attribute

[7] Thanks are due to Susan Chipman for pointing this out: "Show me a model that *doesn't* fit the data."

characteristics of human performance to perceptual and motor abilities and limitations, especially if they have been studied in tasks independent of the one being modeled. This takes advantage of the relatively large corpus of empirical results and well-developed theory on perception and motor movements.

2. Second priority: Try alternative strategies to fully account for the observed effects, adjusting perceptual-motor parameter values if necessary. In general, the effects of different strategies on task performance can be profound and so a good choice is required of both parameters and strategy.
3. Third priority: Try changes or additions to the perceptual or motor mechanisms that are well supported by empirical data independent of the task being modeled. These may entail changes to both the strategy and the parameter values.
4. Fourth priority: The last choice would be making changes or additions to cognitive processor mechanisms beyond simple strategy execution. This is the last choice because such changes are the most hypothetical and most difficult to support by independent empirical data. For example, adding a selective attention mechanism to EPIC's cognitive processor would be justified only if none of the higher priority modifications could account for the data.

The results of applying this approach step-by-step is an *explanatory sequence* of models that increasingly match the data, and provide an explanation of the effects that are present. In terms of philosophy of science, this approach is a way of performing the logical process of *abduction*, which in this sense is *inference to the best explanation* (Douven, 2017), where the *best* explanation in this context must be in terms of the architecture mechanisms, and should rely as much as possible on empirically well-supported mechanisms, and as little as possible on more hypothetical mechanisms

In summary, to explain a specific observed empirical effect with EPIC, first priority is given to independently known properties of human perceptual and motor systems, and second priority is given to examining whether a different task strategy can account for the effect. Then if necessary, the adequacy or accuracy of the perceptual or motor mechanisms can be reassessed in the light of the empirical effects and justified changes made. Only as a last resort will the cognitive processor mechanism be extended beyond its basic strategy-execution role. A capsule summary of this approach is that *perceptual-motor mechanisms come first, task strategies come second, and additional cognitive mechanisms as little as possible*.

Two early applications of this approach was the original EPIC modeling of the psychological refractory period, a fundamental dual-task paradigm (Meyer & Kieras 1997a, b), and EPIC modeling of computer menu selection (Hornof & Kieras, 1997). Examples of architectural modifications that follow these principles are the implementation of covert speech used in verbal working memory (Kieras, Meyer, Mueller, & Seymour, 1999), the efficiency of visually-aimed eye or hand movements (Kieras, 2009; cf. Wright, Marino, Chubb, & Mann, 2019), and visual crowding effects in the early stages of this work on modeling simple visual search. Thus, EPIC models constructed on these principles do a good job of accounting for phenomena that formerly were explained by hypothetical cognitive mechanisms such as a response-selection bottleneck, short-term memory capacity limitations, or attentional capacity limitations.

Applying the approach to simple visual search

In what follows, the initial model for a task requires a strategy that connects some relevant perceptual output to some appropriate motor commands that performs the task in a plausible way. In simple visual search, this means responding present or absent to a presented display with speed and accuracy that is at least roughly in the vicinity of the data to be modeled. The explanatory sequences below each present such an initial model, using availability, eye movements, manual movements, and the Basic Search strategy. Then the sequences first adjust parameters, and then add additional mechanisms or adjust parameters chosen from action slips, crowding, and strategy variations, until a satisfactory fit is obtained. In Section 3 above we outlined the available components for use in a visual search model. Table 4.1.1 summarizes these mechanisms and the parameters whose values were adjusted during the model fitting. These will be used in the explanatory sequences to follow. All other parameters are fixed at their "standard" literature-based or assumed values, as described in Section 3.

If a model based on these mechanisms suffices to account for the effects, there is no need to propose additional architectural components, such as cognitive mechanisms with relatively ill-defined properties such as covert attention. In fact, it is the claim of this work that despite its traditional popularity, the cognitive mechanism of covert attention is not required to explain these visual search effects once known visual mechanisms, eye movements, motor errors, and the task strategy are represented in the model.

Throughout all of the model fits presented, preference was given to minimizing the number of different strategies invoked to fit the data, and trying to hold as many parameters at the same and simple values as possible. Since accounting for ER was a novel undertaking in visual search

Table 4.1.1

	Architecture Mechanism	**Adjusted parameters**
Visual Perception	Availability	θ_S, θ_C, θ_O
	Crowding	ϕ_S, ϕ_C, ϕ_O
	Perceptual store	
Motor Movement	Saccade time, accuracy	
	Manual time, accuracy	*SlipER*
Cognitive Task Strategy	Fixed-Eye	
	Basic Search	
	Unlimited-fixations	
	Limited-fixations	*Nfix*
	Confirm-positive	
	Confirm-both	

models, in the fits to be reported, some preference was given to a close fit for ER at the expense of the RT fit.

4.1.2. Generating model predictions

Implementation Details

The scrambling model of crowding requires that each visual object be processed in the context of the other objects on the display. However, the current code base for the EPIC visual processors are based on the concept of processing each object independently, so implementing the scrambling model would require restructuring the visual architecture code. It is a better tactic to evaluate the prospects of the crowding model before undertaking a careful reprogramming of the EPIC code base. Accordingly, the specific processes used in implementing the simple visual search model in EPIC were reproduced in a stripped-down body of native C++ code. By setting the crowding probability to zero, it was possible to check that the C++ model produced identical RT predictions to the full EPIC version. This was interestingly non-trivial; because the EPIC software directly supports fully parallel processing between and within architectural components, programming an EPIC production-rule strategy is much easier than coding the corresponding behavior in plain C++ code.

Obtaining Reliable Predicted Values

The best-fitting strategy and parameter values were determined by informal iteration. It was necessary to run a very large number of simulated trials to get stable model predictions for the small ERs present in he data. This was achieved by using the C++ "clone" of the actual full EPIC production rule model described above. Because of the very high speed of the clone compared to the full EPIC system (which is coded in C++ as well), it was convenient to obtain *one million* simulated trials for each combination of task condition, set size and trial polarity; all of the reported simulation results use this sample size. The random numbers in the simulation were obtained using the Mersenne Twister generator in the C++ Standard Library.

4.1.3. Metrics for Evaluating Goodness-of-fit

The goodness-of-fit of a model for RT and ER will be reported in terms of three metrics that usefully describe the relationship between the predicted and observed values. This is better than relying on a single metric, because each conveys a different type of useful information, and a different type of misleading information. Thus reporting all three gives a balanced picture of how well or poorly a model fits the data.

Proportion of variance accounted for: r^2

The first metric is the common correlational measure used in modeling work, r^2, the proportion of variance in the observed values accounted for by the predicted values; this basically measures the extent to which predicted and observed values parallel each other. However, it is misleading if the predicted and observed points show parallel trends but differ in magnitude. It is also misleading when there is no observed trend, e.g., flat RTs, because there is no variance to account for even if the predicted and observed values are identical.

Average absolute relative error: aare

Thus it is useful to have a second metric that provides a good measure of how well the values actually match: this is the average absolute relative error, *aare,* the absolute difference between the predicted and observed values relative to the observed values, expressed as a percentage:

$$aare = (\sum_i(|o_i - p_i| / o_i))/n$$

Note how in this measure, under-predictions do not cancel out over-predictions. This gives an easily understood measure of how close in actual values the predictions are to the data, interpretable in terms of rules of thumb customary in engineering practice. However, in this data, some of the ER observed values are extremely small; their presence in the denominator of the relative error tends to "inflate" the *aare* metric, sometimes dramatically, which can be misleading in certain cases.

Average absolute error: aae

A third metric avoids the small-denominator problem; this is the simple average absolute error, *aae*, between the predicted and observed values

$$aae = (\sum_i|o_i - p_i|)/n$$

This measure is in the same units (ms or error proportion) as the data; this can be misleading because it relies on a judgement of how much error is acceptable in terms of those units. For example, an *aae* of 0.1 could be either very good or very bad depending on the units.

Relative reliability of RT and ER data

One additional goodness-of-fit issue stems from the difference in the reliability of the observed RT and ER data. There were about 500 trials per subject per task, set size, and polarity condition. For RT measurements, this sample size can produce very tight estimates for individual subject RTs. However, the variability of proportions, such as ER, is intrinsically higher for the same sample size, and even 500 trials per subject does not constrain the estimate of the individual subject proportion very much. This difference might be why the confidence intervals on the ER data in Figure 2.1 look relatively large when computed in the same way as for the RT data. Thus while the extent to which the predicted values fall within the confidence intervals around the observed data points is an attractive indicator of goodness of fit, the large size of some of the confidence intervals can make it overly generous.

4.2. Explanatory Sequence for the Shape Task

The first explanatory sequence in this report starts with accounting for the Shape task. The reasons for this choice are that of the three tasks in the Wolfe, et al. dataset, this task is the most prototypical visual search task — the task definitely requires visual search; detecting the presence or location of an object is non-trivial, requiring some kind of examination of the display for a few seconds, and there is a large increase in the time required as the number of objects increases. So the key features of this data to be explained are: RTs that steeply increase with set size, almost linearly, with negative trial RT having a greater slope and slightly more curvilinear, than for positive trial RT. The ER for negative trials (False Alarms) is small and almost constant

(as it is in all task conditions), while the ER for positive trials (Misses) increases significantly with set size. The following presents the explanatory sequence to account for these features as a series of numbered steps. The label for each step names the mechanism, parameter, or strategy variation added or modified in that step, and which aspect of the data is addressed in that step.

Step 1. Basic Search strategy and availability for RT slopes. The initial model includes the mechanisms of visual availability, eye and hand movements, and the Basic Search strategy, and attempts to fit the RT effects. ER will be addressed in next step.

Given these mechanisms and the Basic Search strategy, as set size increases, object density increases, and more objects will be covered by each fixation. Thus availability will determine how many eye movements are required before the strategy chooses a response, which then determines the RT. Accordingly, Figure 4.2.1 shows the effects of increasing the threshold

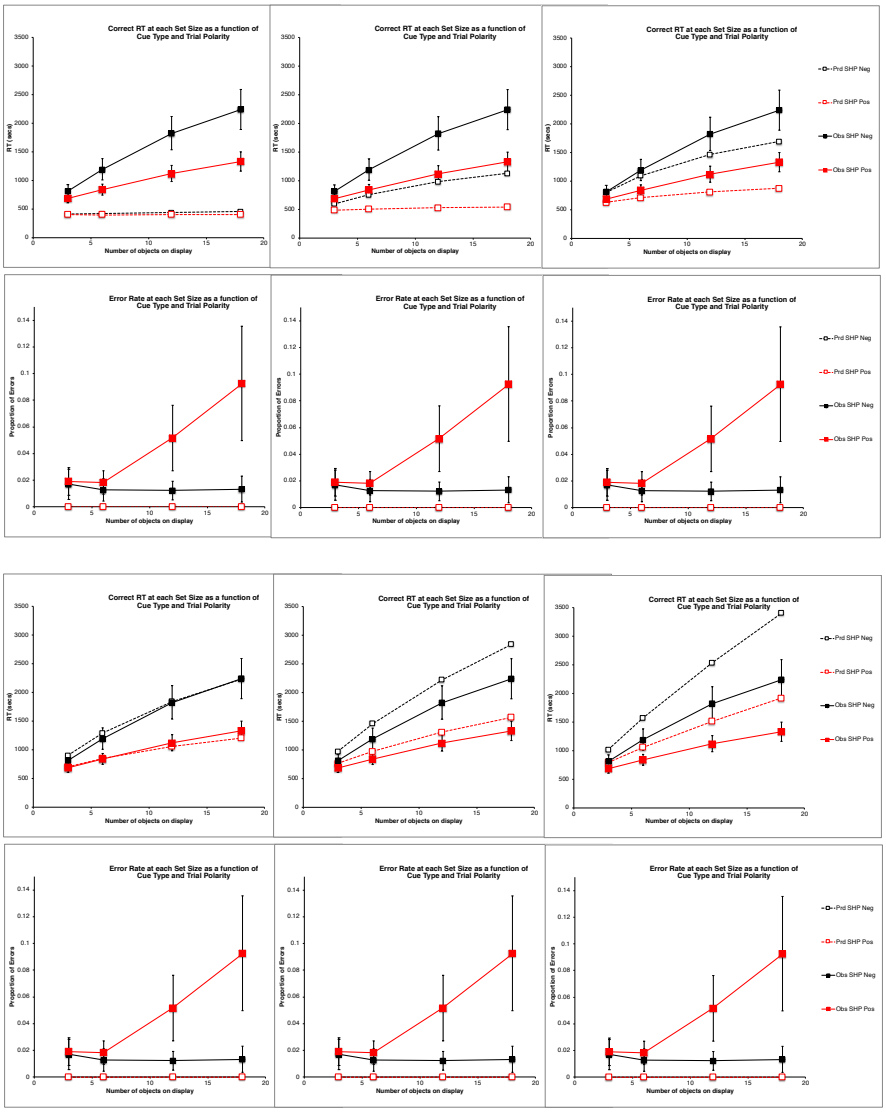


Figure 4.2.1. RT: upper panels in each row, scale 0-3500 ms. ER: Lower panels in each row, scale 0-0.15. Red: positive trial RT and ER; Black: negative trial RT and ER. Observed: Solid points and lines; Predicted: Open points and dotted lines. Top row, left to right: θ_S = {0.1, 0.2, 0.3,} Bottom row, left to right: {0.4, 0.5, 0.6} Sources: SHPAll_VM2dS9b_*_0_0_99_200116

coefficient parameter θ_S from 0.1 on the top row left, where the Shape of almost all objects on the display are available from the initial fixation point, increasing in increments of 0.1 to 0.6 on the bottom row right where very few will be available for any given fixation point. By bracketing the observed RTs, these fits show the effects of availability with the Basic Search strategy. Note that this model predicts no errors, which will be addressed in the next step.

If the Shape property is very widely available (top leftmost panel), the RTs are almost flat and very fast, about 500 ms, because almost no eye movements need to be made. As Shape becomes less available, the Basic Search strategy will make more eye movements, leading to increased RTs with set size. In the bottom rightmost panel where Shape is narrowly available, the RTs are very linear and very steep, especially for negative trial RTs, and the ratio of the negative RT slope to positive RT slope is 2.1. In this case, a fixation usually covers only a single object.

The predicted slopes increase from left to right due to an increase in the number of eye movements generated by the model. Note that in all of the models in this paper, the trial starts with an initial fixation on the fixation point before the display appears. This initial fixation is not counted as one of the eye movements during the trial; only the movements after the display appears are counted. Accordingly, in the bottom rightmost panel, over the range of set size, the mean number of eye movements made by the model goes from 1.9 to 7.6 for positive trials, and 2.9 to 15.0 for negative trials. At intermediate availabilities, the RTs are somewhat negatively accelerated, reflecting how Shape becomes available for more objects in a single fixation because the objects are closer together with greater set size. A fairly good fit to the RTs (r^2 = 0.98) appears in the bottom leftmost panel where $\theta_S = 0.4$, but the predicted RTs are slightly more negatively accelerated than the observed. Here the range of eye movements is 1.5 - 4.0 for positive trials, and 2.4 - 9.1 for negative trials.

In the visual search literature, much is made of the linearity of the RTs with set size and how the ratio between negative and positive RT slopes is often in the vicinity of 2:1, which is consistent with a serial self-terminating search (which is what the Basic Search strategy implements here). However, over the availability parameter range shown in Figure 4.2.1, the predicted linear slope ratios are {594, 9.8, 3.6, 2.7, 2.3, 2.1}. Thus, the RTs are linear and slope ratio is close to 2:1 only to the extent that fixations cover single objects. As pointed out by earlier work with the FVF concept, (see Hulleman & Olivers, 2017, for a review), if the property is available over a wide enough area that more than one object can be recognized at time, the search will be faster because fewer eye movements are needed, and the slope ratio will be larger, mostly because the slope for positive RT becomes flatter.

This set of fits makes it clear that a very simple model of visual search, involving availability, eye movement times, and the Basic Search strategy can account for this family of RT effects: Depending on availability, RTs can range from flat to fairly linear steeply sloping RTs, and negative RTs are generally larger than positive RTs by a ratio that converges on 2:1 as availability decreases. That is, serial processing, as indicated by the 2:1 linear slope ratio, appears when objects have to be separately fixated; when more than one object can be covered in a fixation, negatively accelerated RTs with linear slope ratios greater than 2:1 appear. This range of effects stems simply from how the availability of the relevant object properties governs the need for eye movements.

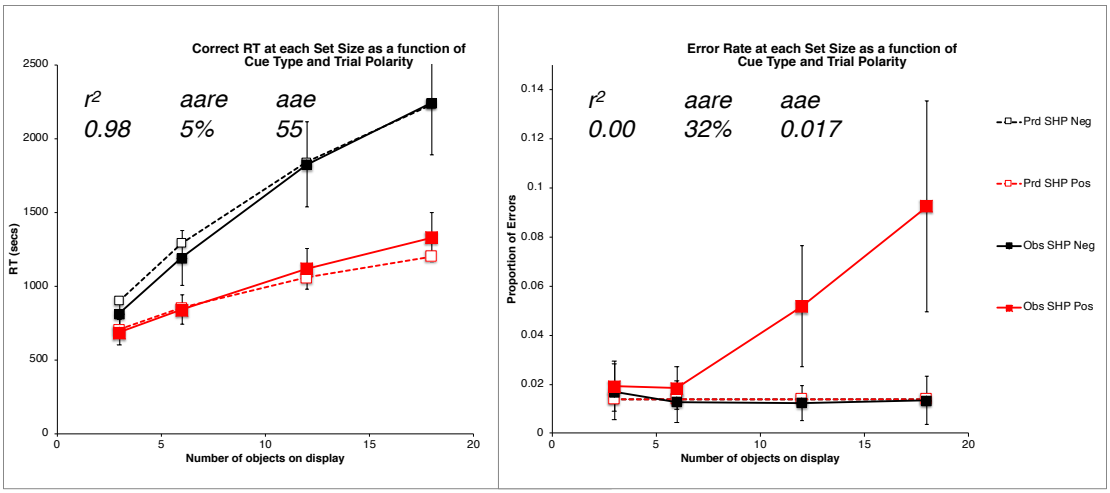


Figure 4.2.2. Effect of adding slip errors. θ_S = 0.4, SlipER = .014.
Source: SHPAll_VM2dS9b_4_0_014_99_200116

Step 2. Action slips for False Alarms. The predicted results in Figure 4.2.1 show zero predicted errors, which is obviously a serious misfit. This model never makes an error because in no case does the visual system produce an *incorrect* representation — the Shape is either veridically available for an object, or it is missing (a *blank* property value), and the search continues until all objects have a known Shape, whereupon the choice rules always apply reliably, and so no errors are made. As pointed out previously, a remarkable feature of these data is that errors on negative trials (False Alarms) are produced at a very low and almost constant rate: the average over all three task conditions is 0.014. In contrast, Misses (errors on positive trials) tend to increase with set size, especially for the Shape condition.

The model was modified to include the action slip mechanism with the *SlipER* parameter set simply to the overall ER for False Alarms of 0.014. After the strategy determines the response, with probability *SlipER*, the *opposite* response is made. Figure 4.2.2 shows the fit of this model with the above good estimate of θ_S = 0.4, and the goodness-of-fit metrics described in Section 4.1.3. While this step improves the ER fit substantially (*aare* goes from 100% to 32%), there is no predicted trend of increasing Miss ER, and so the r^2 for ER remains at 0.0. Note that the average correct trial RT is unaffected by slip errors because they occur independently of the visual processing and strategy execution, and incorrect trial results are not averaged into the correct trial RTs .

Step 3. Crowding for Misses increasing with set size. As described above in Section 3.3.3, under the Basic Search strategy and its variants, the object Shape can be treated as a unitary property. The crowding mechanism scrambles the Shape property of objects that crowd each other, with a blank or unknown property value participating in the scrambling. The strength of the crowding effect is specified by the crowding probability parameter ϕ.

The effect of crowding on RT and ER is somewhat subtle and requires careful explanation. First, note that the crowding scrambling will not produce an illusory target on a negative trial, because there is no unitary target Shape property on the display, so crowding will not result in a False Alarm. But, the scrambling can result in the distractor Shape property being moved to an object whose Shape was not actually available — a form of illusory distractor that will not itself produce a False Alarm on a negative trial.

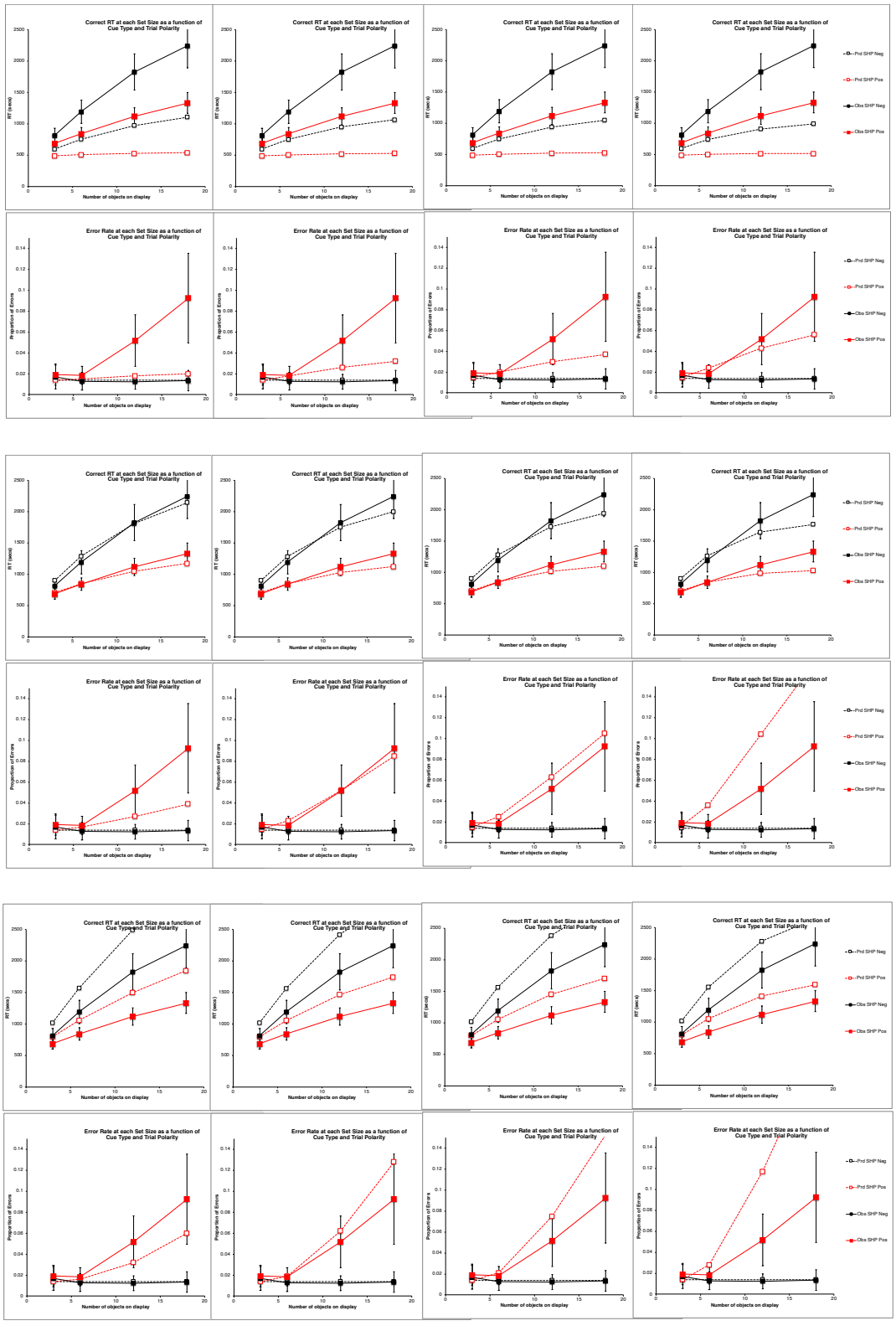


Figure 4.2.3. The effect of varying availability and crowding probability on RT (upper panels in each row, scale 0-2500 ms) and ER (lower panels in each row, scale 0-0.15).
Rows: Top to bottom θ_S = {0.2, 0.4, 0.6}. Columns: Left to right: ϕ_S = {0.025, 0.075, 0.1, 0.2}.

However, on positive trials the situation is more complex. The target Shape property is visible on the display, but scrambling might move it to a distractor object, producing an illusory target. But because the task requires simply making a present response rather than fixating or identifying the correct target object, the fact that the illusory target object is not the actual target object has little or no effect on RT. But as more fixations are made, it is possible that *all* objects receive a scrambled distractor Shape, including the target object (which then becomes an illusory distractor), before all of the objects have been covered by a fixation close enough to make their actual Shape property available. In this case, all objects appear to be a distractor, and accordingly the Basic Search strategy will immediately halt the trial with an absent response, which is a Miss error on a positive trial. The likelihood that an object's property will be overwritten by a distractor property value depends on both the prevalence of the distractor properties, which depends on the availability and the crowding probability parameters, θ and ϕ. These two parameters can be varied to bracket the effects on RT and ER.

Accordingly, Figure 4.2.3 expands upon Step 1 and Step 2 to show the effects on RTs and ERs of varying both the availability and crowding probability parameters, with *SlipER* remaining at 0.014 and with Unlimited-fixations in the Basic Search strategy. The three rows of RT and ER graphs from top to bottom correspond to decreasing availability, with θ_S going from 0.2 (very available) to 0.4, then to 0.6 (very unavailable). The columns of graphs from left to right correspond to increasing crowding probability, with $\phi_S = \{0.025, 0.075. 0.1, 0.2\}$.

As before, increasing the availability threshold θ_S increases the RT slope, but the extent to which the RTs become more linear also depends on the crowding probability parameter ϕ_S. But the major effect is how Miss ER systematically increases with set size as the ϕ_S increases, but the magnitude of this effect also depends on θ_S. In addition, because Misses result from an early termination of the search when all objects have the distractor property, a higher Miss ER corresponds to more negative acceleration in the RT functions. Thus the availability and crowding mechanisms jointly govern both the RT and ER. A fairly good fit for both RT and ER is obtained with $\theta_S = 0.4$ (middle row) and $\phi_S = 0.075$ (center left).

Figure 4.2.4 is refined from the above, with $\theta_S = 0.425$ and $\phi_S = 0.075$, which exemplifies the preference in this work to emphasize a good fit to ER, since this is a fairly novel goal. As shown in the Figure, the fit to both RT and ER is an extremely good result from only two adjusted parameters, the availability threshold and the crowding probability. Because the crowding can cause the strategy to terminate early, more frequently with increasing set size/density, the RTs are also somewhat more curvilinear.

A step not taken. The above crowding account of Miss ER differs substantially from the common explanation in terms of some kind of time-out process (cf Hulleman & Olivers, 2017), in which if the target has not been detected after some time, or number of fixations (which in the model occur at a roughly constant rate), the subject simply responds absent — this strategy possibility is the Limited-fixations option (Box 3.5 above). However, according to the explanatory priority order, if an effect can be explained with known perceptual mechanisms, i.e., availability and crowding, a strategy variation explanation should not be entertained. Moreover, to acceptably fit the Miss ER data, a different limit is needed for at least three of the set size levels, which requires more parameters than the crowding mechanism, and requires possible additional visual mechanisms and strategy variations to adjust the stopping criterion depending

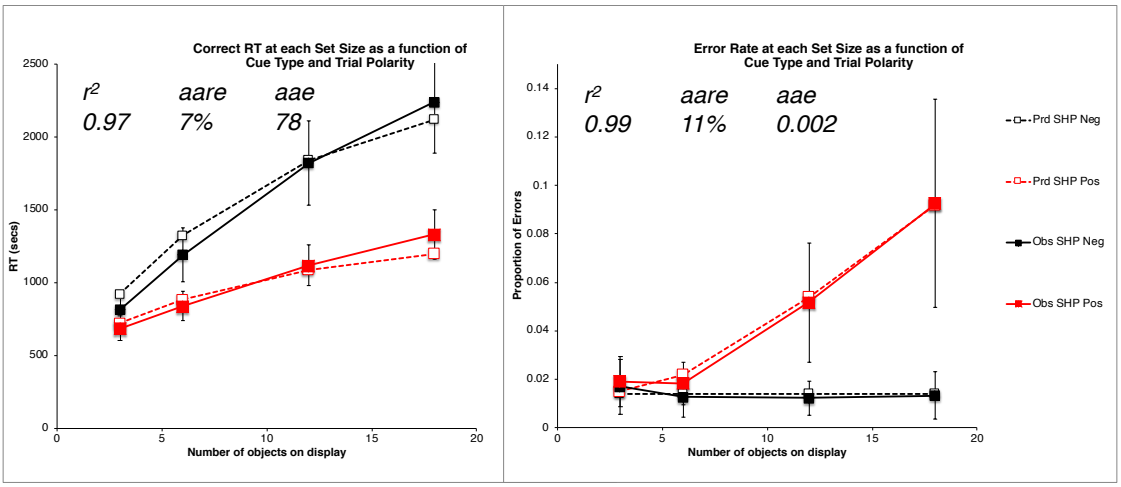


Figure 4.2.4. Good fit with Basic Search strategy, Unlimited-fixations, S_av = 0.425, S_cp = 0.075
Source: SHPAll_VM2eLS9c_425_075_014_99_200304

on the set size. Thus relying on crowding not only conforms to the priority order, but produces a simpler model than a strategy change.

Final model for Shape task. The explanatory sequence started with the basic search strategy, eye movements and manual responses, and found a value for the availability parameter that resulted in an approximate RT fit, and then added slip errors and crowding and suitable parameter values to arrive at a very good fit to both the RT and ER data. The effects in the data reflect only a "fastest reasonable" strategy working with relevant independently documented perceptual-motor limitations; no concept of a perceptual or central "bottleneck" such as covert attention is required to account for this prototypical visual search task.

4.3. Explanatory Sequence for the Color Task

The next explanatory sequence is for the Color task aggregate data. This task is a good next choice for explanation because like Shape, it involves a single property of the objects, but the color property is at the opposite extreme of availability from Shape; there are many results that suggest that object color is very widely available in the visual field (e.g. Williams, 1967) so widely that it is often described as "popping out" in a visual search task. The explanatory sequence here shows that the "pop out" phenomenon is not a special mechanism, but rather is due simply to the single target property being very available over the visual field. As in the Shape sequence, the RT effects in the Color task sequence will be accounted for first, and then the ER effects. However, for brevity, the slip error account of False Alarms developed for Shape will be included in the first step.

Step 1. Basic Search strategy, slip errors, availability for flat RTs. The starting point in the sequence is the same as the initial model for the Shape sequence: Basic Search strategy with Unlimited-fixations, availability, eye and hand movements, with slip errors added. Figure 4.3.1 shows the effects of changing the availability threshold ranging from all objects having available color to only some, with $\theta_{C} = \{0.0, 0.1, 0.15, 0.2\}$, an *SlipER* of 0.014, and no crowding effect (ϕ_C = 0.0).

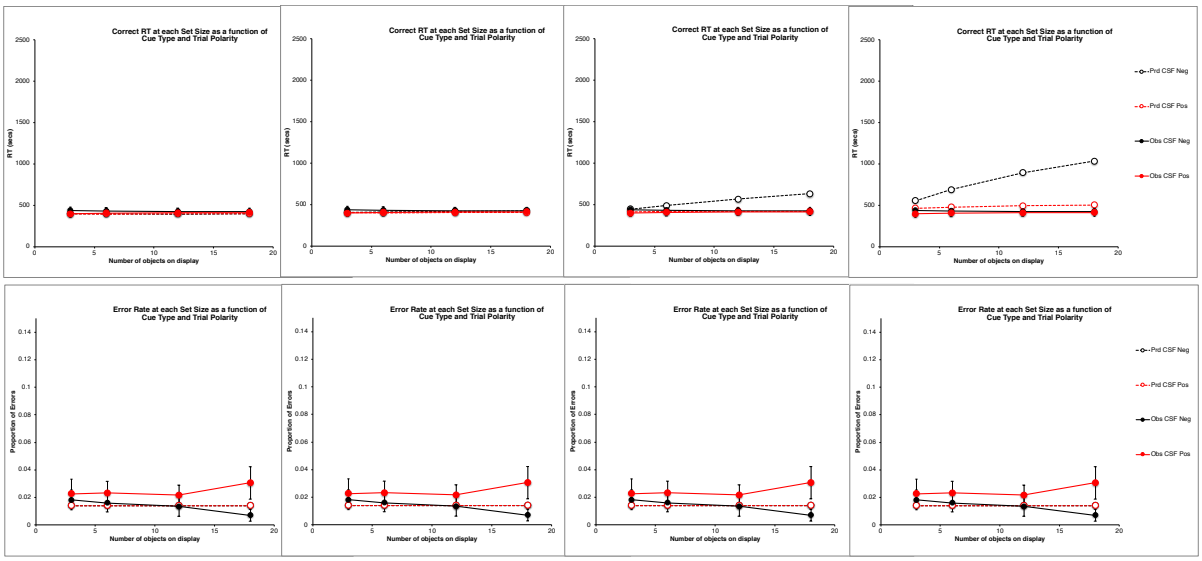


Figure 4.3.1. Effect of availability and Basic Search strategy in the Color task. RT (upper panels, scale 0-2500 ms), ER (lower panels, scale 0-0.15). SlipER = .014. Left-to-right: $\theta_C = \{0.0, 0.1, 0.15, 0.2\}$. $\phi_C = 0.0$.

For the leftmost two fits, which have especially high availability $\theta_C = \{0.0, 0.1\}$, the predicted RTs are very fast and flat and very close to the observed RTs. Increasing RT slopes appear at larger values of θ_C. Contributing to these RTs is the number of fixations/trial produced by the model after the initial fixation. As before, the model counts only eye movements after the initial one that is always made to the fixation point before the trial is started. Thus, any slope in the RT must be produced by subsequent fixations. As expected, for $\theta_C = 0.0$ the model makes zero eye movements — the Color is fully available for all of the objects, so the strategy terminates immediately. For $\theta_C = 0.1$, a few objects initially lack the Color property, but at set size 18, the model makes only 0.022 fixations/trial for positive trials, and 0.128 for negative trials, which produces RTs that are almost flat and almost identical for positive and negative trials, so as to be indistinguishable when plotted on the same scales. This happens because the Color is still almost always available. However for $\theta_C = 0.15$ there is visible slope in the predicted negative RTs resulting from 1.068 fixations/trial at set size 18. This slope is clearly different from the observed flat negative RTs. But the predicted positive RTs are still almost flat due to a tiny .094 fixations/trial at set size 18, and very close to the observed positive RT. In this case, the predicted slope ratio is 48:1! Finally, at even less Color availability, $\theta_C = 0.2$, the predicted positive RTs are now somewhat greater than the observed positive RTs but still fairly flat due to there being only 0.483 fixations/trial at set size 18, while the negative RTs are much more steeply increasing with 2.852 fixations/trial at set size 18. As a result the slope ratio is reduced to only 13 due to the increased positive RT slope.

From these results it is very clear that even as little as an average of one fixation per trial at the largest set size will produce a significantly sloped RT. In turn this means that θ_C needs to be around 0.1 or less to obtain a plausible fit to the fast and flat RTs.

What about the ER effects? This initial Color task model includes *SlipER* = 0.014, the value of the overall observed False Alarm ER. But the observed Miss ER clearly tends to be higher than the False Alarm ER — how is this explained? One possibility is an ad-hoc explanation that

the *SlipER* is simply larger for present responses than for absent responses. But rather than use another degree of freedom for an additional and arbitrary mechanism with a parameter value, it would be better to find an explanation compatible with the priority order for the combination of flat RTs but more Misses than False Alarms.

Step 2. Crowding for Misses. In the Shape task sequence above, crowding produces Misses that increase with set size, and there are values for θ_S and ϕ_S that result in a fit to both RT and ER. Could the same hold for the Color task? Figure 4.3.2 shows the results for θ_C over the same range of values as shown in Figure 4.3.1, and ϕ_C set at 1.0, which is the largest possible crowding effect.

These results are very striking. Higher values of θ_C produce both RTs that increase with set size, and also Miss ERs that increase with set size, similar to the effects for Shape. However, the small values of θ_C that produce extremely few eye movements and flat RTs also produce no apparent increase at all in Miss ER with set size, even when ϕ_C is at its maximum value. Those θ_C values that produce an increase in Miss ER also produce substantial RT slopes. The same pattern emerges for smaller values of ϕ_C.

These predictions result from the following: If the target Color is available, crowding might swap which object has the target Color by producing a pair where the actual target becomes an illusory distractor and an actual distractor becomes an illusory target, but the strategy will still produce a correct present response. It is also possible that scrambling involving multiple objects could result in the target object Color property being lost from overwriting with a blank property, turning the target object into an *illusory blank* as discussed above in Section 3.3.1. The result would be a Miss, even if the target Color was originally available.

So crowding will only produce a Miss error if it scrambles the target object into an illusory blank and produces no illusory target in the process. This scenario is highly unlikely when Color is highly available for the following reasons: (1) The eyes almost always remain on the fixation point whereupon only the initial scrambling would be performed. (2) The probability is low that

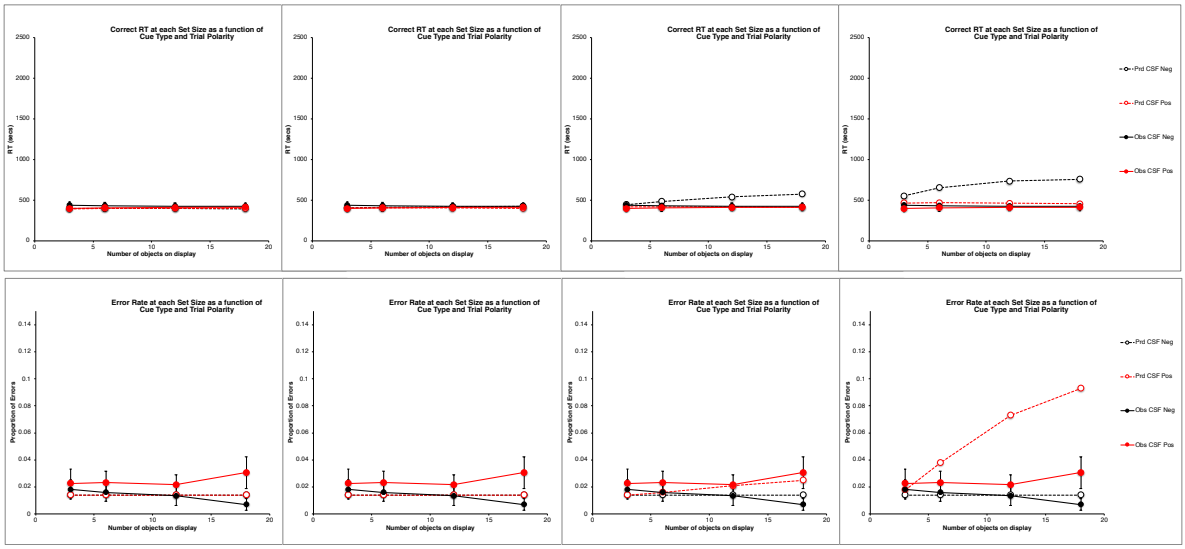


Figure 4.3.2. No effect of crowding. RT (upper panels, scale 0-2500 ms), ER (lower panels, scale 0-0.15). Basic Search strategy, SlipER = .014. Left-to-right: θ_C = {0.0, 0.1, 0.15, 0.2}. ϕ_C = 1.0.

a blank Color will be present, so the chance that crowding will overwrite an available target Color with a blank is low. (3) Since the initial average eccentricity is only 8.8 (Figure 3.4), and the mean number of crowding flankers at this eye position is only 1.0 at set size 18 (Figure 3.5), the likelihood of the single initial scrambling operation having any effect on the target object is low. Thus the correctness of the response depends almost completely on whether the target property is available during the initial fixation, which essentially depends only on θ_C.

Thus, although crowding is at work in the Color task displays, a small θ_C results in a negligible effect of crowding on Miss ER. In contrast, and similar to Shape, if θ_C is large, eye movements produce enough scrambling to produce both RTs and Misses that increase with set size, which contradicts the flat RTs and fairly flat observed Miss ERs in the data.[8]

The conclusion is that if the Basic Search strategy is used, there are no values of θ_C and ϕ_C that will simultaneously produce both the flat and fast RTs and a flat Miss rate higher than the False Alarm rate. According to the priority order, a strategy change would be the next modification to try.

Step 3. Fixed-Eye strategy and availability for flat RTs and Misses. Another way to get a flat RT is the Fixed-Eye strategy (Box 3.2), which after the initial fixation, allows no subsequent eye movements at all, with the response to be made based only on what is available during this initial fixation: On a positive trial, if the target Color is available from the fixation point, the response will be present; if the target Color is not available, the response will be absent, producing some Misses; on a negative trial, all responses will be absent; these responses will be inverted at the constant *SlipER*. Thus, if the Color property is not always available, Miss responses are more frequent than False Alarms, and the RTs are constant, determined only by the fixed time required for perception, decision, and action, while the ERs are determined only by *SlipER* and θ_C.

Figure 4.3.3 shows the Fixed-Eye strategy model predictions for the same set of values for θ_C and *SlipER* as Figure 4.3.1 with $\phi_C = 0.0$. Note that the rightmost ER graph for $\theta_C = 0.2$ has a different scale since the predicted Miss ER is very high (0.276). A fairly good fit is obtained for $\theta_C = 0.1$. Figure 4.3.4 shows a refined fit obtained with $\theta_C = 0.11$. Since the RTs and ERs are flat with set size, the r^2 s are only zero and 0.68 respectively, but the other goodness-of-fit measures are very good.

The Fixed-Eye strategy produces a negligible effect of crowding. This follows from the Step 2 discussion above, where the effect of crowding depended on eye movements. This was verified with a series of Fixed-Eye model fits, not shown, with a very wide range of values of θ_C and ϕ_C. An effect of ϕ_C on predicted Miss ER was detected, but negligible; in the most extreme test cases, crowding added at most 0.001 to the Miss ER.

Final model for the Color task. The Color task is at the opposite extreme from the Shape task, and the models capture the difference. The Color task effects can be accounted for by the extremely simple Fixed-Eye strategy and the availability mechanism. No special "pop-out" mechanism is required; Color is simply visible enough that eye movements would be rarely

[8] As pointed out in Section 2.2 above, and confirmed by a within-subject *t*-test, the apparent slight increase in Miss ER at set size 12 to set size 18 is not statistically significant. This argues against any attempt to account for it in these data.

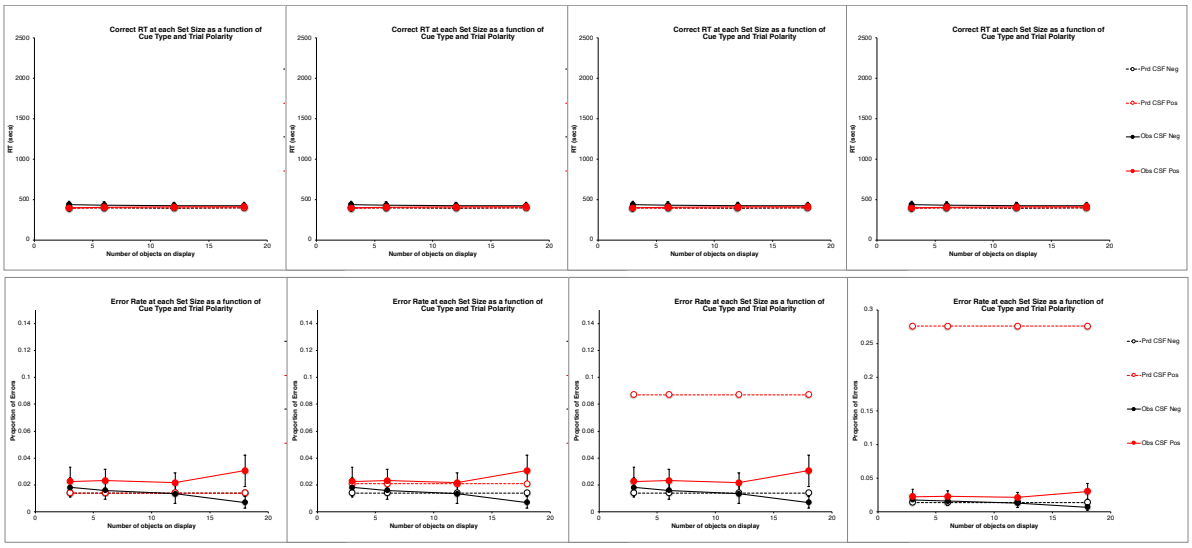


Figure 4.3.3. Effects of availability with Fixed-Eye strategy. RT (upper panels, scale 0-2500 ms), ER (lower panels, scale 0-0.15 except for rightmost which is 0-0.3). SlipER = 0.014.
Left-to-right: θ_C = {0.0, 0.1, 0.15, 0.2}. ϕ_C = 0.0.

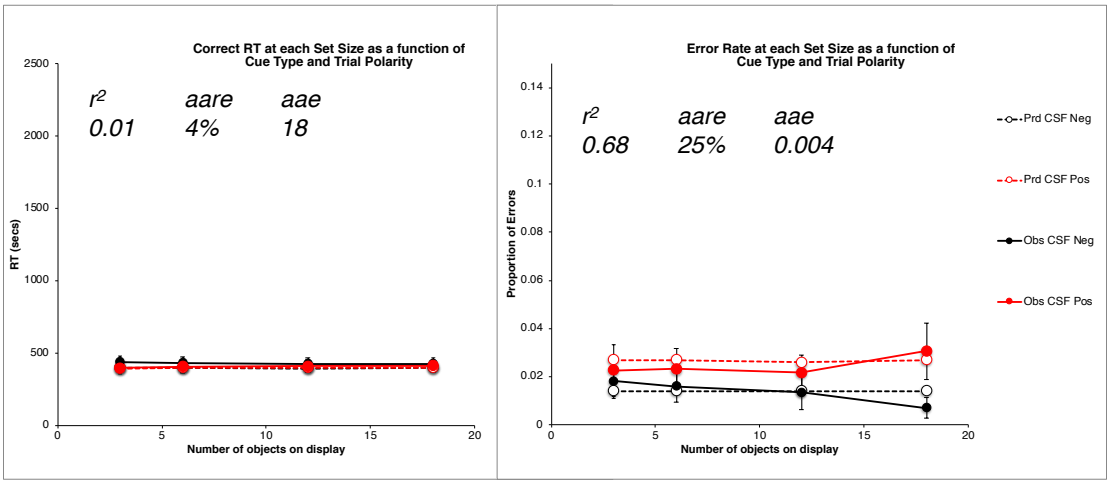


Figure 4.3.4. Good fit with Fixed-Eye strategy, θ_C = 0.11, SlipER = 0.014.
Source:CSFAll_VM2eS9c_11_0_014_0_200225

required, and can simply be eliminated by the Fixed-Eye strategy if some Misses are acceptable. Interestingly, the effect of crowding, while detectable in the model, was not needed to adequately fit the data in this situation of no eye movements and high availability. But a crowding probability parameter value could be assigned to the Color property (e.g. for use in the Conjunction task) without affecting the accuracy of the Color task Model. In fact, the next explanatory sequence shows how crowding plays a powerful role in the Conjunction task, even though both Color and Orientation are highly available.

4.4. Explanatory Sequence for the Conjunction Task

The next explanatory sequence is for the Conjunction task aggregate data, which has a pattern of effects that are different from both the Shape and Color tasks, which represent two extremes of the search task. Unlike the Color task with its flat and fast RTs because no eye

movements were needed, the Conjunction RTs have definite slopes, but much less than for Shape that required many eye movements. Also, unlike the Color task, where availability determined Miss ER and crowding effects played a negligible role due to the lack of eye movements, the Conjunction Miss ER increases substantially with set size, but to a lesser extent than for Shape, where crowding produced increasing Misses. These differences between single-feature and conjunctive search tasks was the key phenomenon motivating the covert attention theories of visual search. So the challenge is satisfactorily explaining these effects with the available EPIC mechanisms. The following explanatory sequence shows how this can be done.

Step 1. Basic Search strategy, slip errors, and availability for RT slopes. The initial model for this task assumes the Basic Search strategy even though Color is a target property in the Conjunction task and the Fixed-Eye strategy produced a good fit for the Color task. However, the Fixed-Eye strategy can be immediately ruled out because the Conjunction RTs are clearly and reliably sloped rather than flat. As pointed out above in Section 3.3.3, the Conjunction task implementation of Basic Search requires more complicated nomination rules for targets and possible targets, because there are different possible combinations of the two properties of Color and Orientation, which are assumed to be independently available.

A reasonable simplifying assumption is that the availability parameter θ_C for Color has the same value in Conjunction as in the best fit for the Color task, namely $\theta_C = 0.11$, and it will be held constant at that value in what follows. A rough conclusion from the available literature is that Orientation is less available than Color, but this data set did not include an Orientation single feature task that could be used to separately estimate θ_O for stimuli of the same size and shape.

Accordingly, Figure 4.4.1 shows a set of bracketing fits with the Basic Search strategy, using the previous best-fit value for Color availability $\theta_C = 0.11$, and with a range of Orientation availability θ_O values from 0.11 (same as Color) to 0.25. Again for brevity, this first model step includes slip error with *SlipER* = 0.014, as in the Shape and Color task models. As shown in Figure 4.4.1, the RTs become definitely sloped for values of θ_O greater than 0.11, and are fit fairly well with $\theta_O = .225$ or .25, consistent with Orientation being less available than Color (middle and middle right panels).

These results make an important point, present but not emphasized, in Step 1 of the Shape and Color explanatory sequences: Eye movements can in fact produce shallow RT slopes; this undermines the basic justification of the covert attention model (c.f. Section 1.3 above). If availability is high, multiple objects can be perceived and judged simultaneously during a single fixation, and so very few eye movements might be required. In the Figure 4.4.1 model results, at set size 18 and $\theta_O = 0.25$, the number of eye movements average only 1.0 for positive trials, and 3.0 for negative trials.

However, because the Basic Search strategy continues until the target is found, as in the corresponding Shape case, there are no Misses except for slip errors. Since crowding could explain the Misses in Shape, adding crowding effects is a good next step to account for Misses in Conjunction.

Step 2. Crowding for Misses. Can crowding account for the Miss ER effects in Conjunction like it did in Shape? Figure 4.4.2 shows a single fit for Basic Search and availability parameters

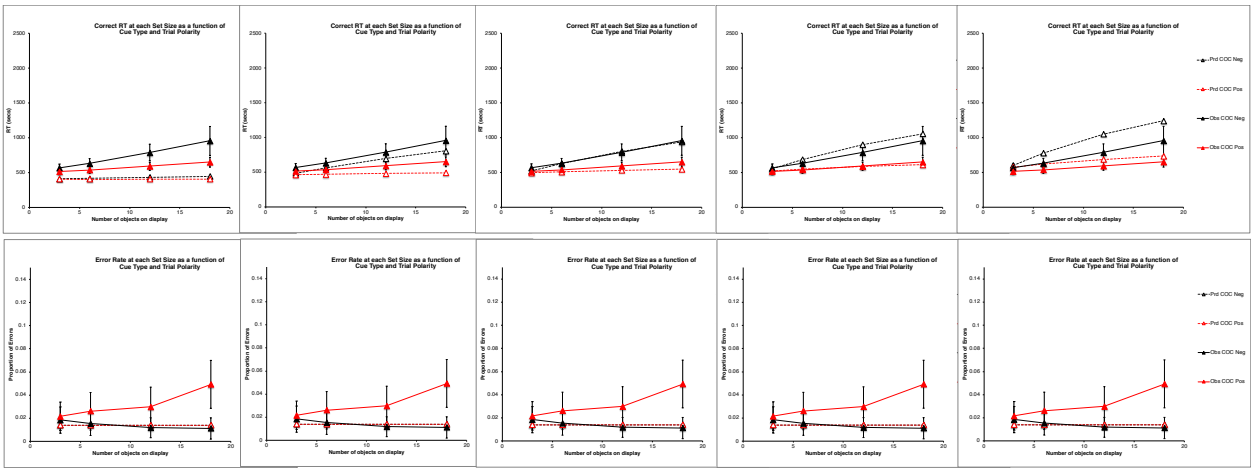


Figure 4.4.1. RT (upper panels, scale 0-2500 ms), ER (lower panels, scale 0-0.15). No Crowding, Unlimited-fixations, SlipER = 0.014, θ_C = 0.11,
Left to right: O_av = {0.11, 0.2, 0.225, 0.25, 0.3}.

based on the best fit from the Step 1 results (θ_C = .11, θ_O = .25) with crowding probability assumed to be equal for Color and Orientation with values $\phi_C = \phi_O = 0.025$.[9]

Only a single fit needs to be shown for this step, because there is a huge problem that makes additional fits irrelevant. Figure 4.4.2 shows that while the RTs are fit fairly well, there is a *massive* predicted False Alarm ER that increases steeply with set size, and shoots completely off the plotted range at set size 12. Any non-zero value for the crowding probabilities ϕ_C and ϕ_O produces this seriously misfitting pattern.

Treating ER as a first-class dependent variable along with RT means that this surprisingly gross misfit cannot be ignored. It results from crowding combining with the Basic Search strategy as follows: The Conjunction task has an almost equal number of Red and Green property values on the display, and also an almost equal number of Horizontal and Vertical property

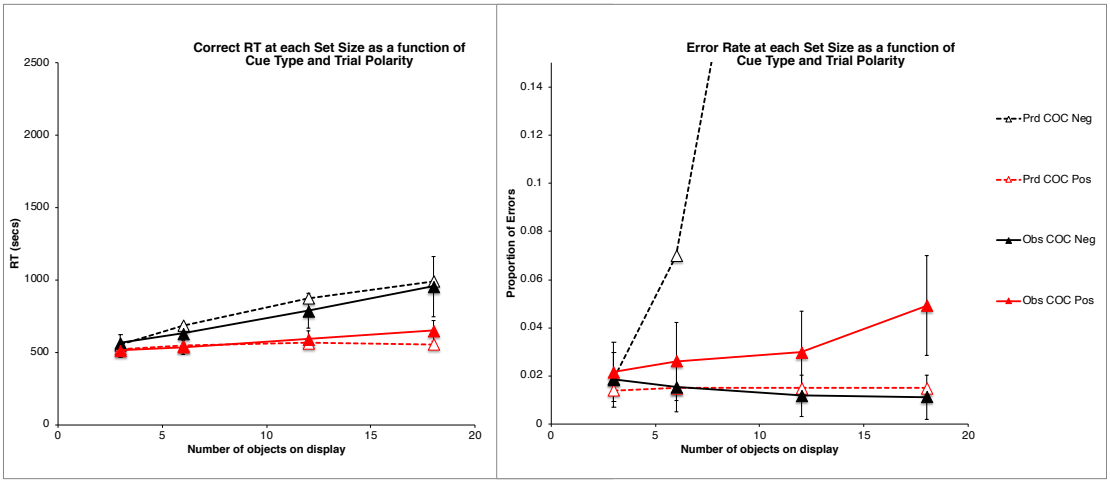


Figure 4.4.2. Scales: 0-2500 ms, 0-0.15. Basic Search unlimited fixations, θ_C = .11, θ_O = .25, $\phi_C = \phi_O$ = 0.025, SlipER = 0.014. Note that the ER scale is the same as in previous graphs, and the False Alarm ER rises to very high off-scale values.

[9] The small value is intuitively reasonable because Color is reported to be relatively insensitive to crowding, and the very distinct values of Orientation might also be similarly resistant to crowding.

values on the display. The target object is the only object that is both Red and Vertical. Since Color is very available, and Orientation fairly available, there are many instances of these values available at a time. As described above in Section 3.3.1, crowding can cause illusory targets and illusory distractors when the Color and Orientation property values get scrambled between the objects. If a Vertical replaces a Horizontal on a Red distractor, or a Red replaces a Green on a Vertical distractor, the result is an illusory target. The Basic Search strategy will terminate as soon as an object appears to be a target regardless of whether it has been fixated. So even at these low crowding probabilities, on negative trials illusory targets are common enough to cause the strategy to frequently terminate early with a False Alarm, and at a rate that increases steeply with set size, as more objects crowd each other on the display.

But if the objects happen to be widely spaced enough on a negative trial that crowding doesn't create an illusory target, a correct absent response will be made, possibly after a few fixations to make the properties all available, resulting in a sloping negative RT. But on positive trials, if the actual target is not visible at first, the frequent illusory targets will cause the strategy to terminate quickly with a present response, which is still the correct response. The result is an almost flat positive RT. Even if illusory targets happen not to be present, the unlimited search will eventually find the actual target, so the only Misses are slip errors, flat with set size.

Thus this model is remarkably incorrect even though it incorporates all of the previously considered visual and motor mechanisms, including crowding. The implication is that performance in the Conjunction task involves something additional and a change to the strategy is the next step to try in the explanatory sequence.

Step 3. Basic Search with Confirm-positive, crowding for RT and ER. The False Alarms can be prevented with the Confirm-positive version of the Basic Search strategy, shown above in Box 3.3, which fixates an apparent target to confirm that it is an actual target before responding present, and continues the search if not.

Figure 4.4.3 shows the results for Basic Search with Confirm-positive strategy using the same availability values as previous, with $\phi_C = 0.025$, and with three different values of the Orientation crowding probability ϕ_O, increasing from left to right from 0.025, then 0.1, to 0.2. Starting at the left panel and going to the right, the positive RTs have a slope similar to the observed slope, but are larger than the observed values, while the negative RTs start much steeper than the observed values, and become extremely steep with increasing ϕ_O. Also going from left to right, the effect of set size on Misses also increases. The problem is that the value of ϕ_O that produces RTs most like the observed (left-most panel, $\phi_O = 0.025$) produces very few Misses, but increasing ϕ_O to a high enough value to produce enough Misses results in huge negative RTs, as well as positive RTs that are substantially larger than the observed.

What is happening on negative trials is that the combination of crowding, Confirm-positive, and Unlimited-fixations means the search continues until all of the objects appear to be distractors, whereupon an absent response is immediately made; but many fixations to resolve illusory targets may be needed before the absent response is triggered. In contrast, no non-slip False Alarms are produced, thanks to the Confirm-positive step. Furthermore, on a positive trial, the only way to make a non-slip Miss error is if the actual target object appears to be an illusory distractor at the same time that all the other objects appear to be distractors. This will not happen very often unless the crowding probability is high, whereupon the RTs are too long.

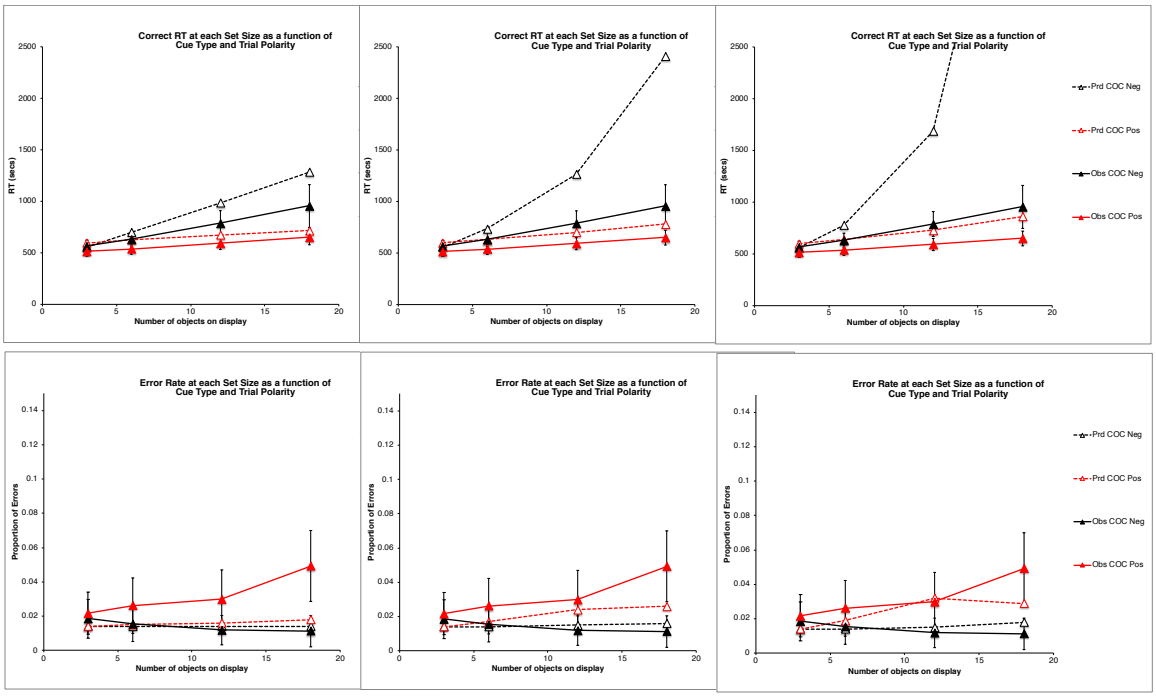


Figure 4.4.3. RT (upper panels, scale 0-2500 ms), ER (lower panels, scale 0-0.15). Basic Search with Confirm-positive and Unlimited-fixations. $\theta_C = 0.11$, $\theta_O = 0.25$, $\phi_C = 0.025$, $\phi_O = \{0.025, 0.100, 0.200\}$, *SlipER = 0.014.*

Thus, even though crowding is incorporated into the model, the problem is that an Unlimited-fixations strategy produces either extremely over-predicted negative RTs or under-predicted Miss ERs. It seems like subjects should be able to produce much faster RTs with an acceptable Miss ER. Perhaps a strategy change can speed up the RTs by limiting the number of fixations.

Step 4. Basic Search Confirm-positive with the Limited-fixations option for RT and Misses. The strategy was modified to include the option (Box 3.5) of responding absent after a fixed number of fixations *NFix*. Figure 4.4.4 shows a fairly good fit with a pair of the previous availability and crowding parameter values and *NFix* = 3. This strategy still eliminates the excess False Alarms, but also produces fast sloping RTs, and an acceptably low Miss ER that increases with set size. Some preference was given to fitting the ER closely at the higher set size, but the RTs are still fit fairly well.

However, close inspection of the predicted RTs in Figure 4.4.4 reveals that the predicted RT for negative trials is systematically somewhat too fast, especially at the smallest set size where it is predicted to be slightly faster than the positive trial RT, a qualitative misfit. This happens because the Confirm-positive version of Basic Search often requires an extra eye movement before making a present response, but this is not true for an absent response. One more step in the explanatory sequence is required to deal with this small, but systematic, misfit.

Step 5. Confirm-both strategy for positive RTs. The Confirm-positive step is a form of double checking, which is required because there may be many illusory targets in the Conjunction display. But it is also very likely that the same scrambling of Color and Orientation could turn a target into an illusory distractor, which suggests that an additional double-check might be needed to deal with this possibility. But there is no information on which apparent distractor could be the actual target, making this double-check less well-defined than the

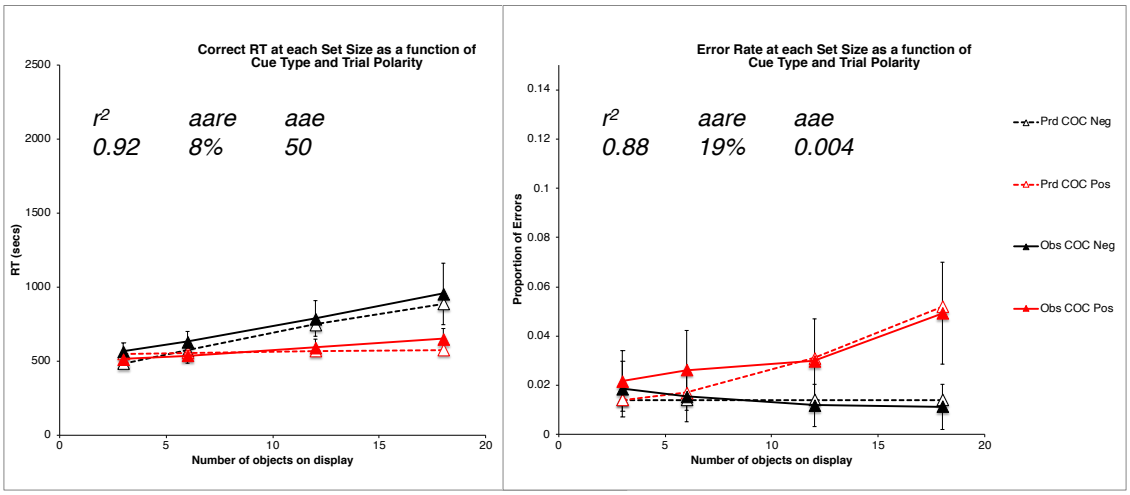


Figure 4.4.4. Confirm-positive with Limited-fixations. θ_C = 0.11, ϕ_C = 0.025, θ_O = 0.2, ϕ_O = 0.025, SlipER = 0.014, NFix = 3.
Source: COCAll_VM2eCS9c_110_025_200_025_014_3_CP_200208

Confirm-positive check. Accordingly, the Confirm-both version of Basic Search (Box 3.4) adds an extremely simple version of an Confirm-negative double-check: if all objects appear to be distractors, and no fixations have yet been made (i.e. the eyes are still on the fixation point), the current most eccentric object is fixated and then checked for being a distractor. If it is a distractor, an absent response is made, if not, the search continues.

As shown in Figure 4.4.5, the fit to the ER is unaffected, but the under-prediction of negative trial RTs has been alleviated, and the r^2, *aare*, and especially the *aae*, are all noticeably better, due to an additional eye movement and property check being made on a subset of the negative trials.

Final model for the Conjunction task. The final model shows that the Conjunction task is very different from the Shape and Color task tasks. The problem is that even with a very low probability of crowding, the scrambling of the highly available Color and Orientation properties

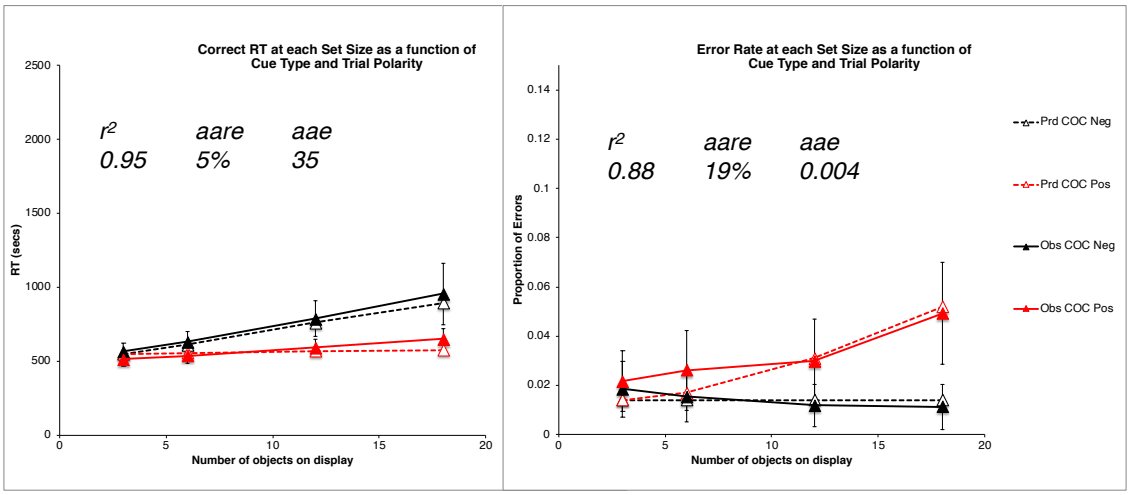


Figure 4.4.5. Confirm-both with Limited-fixations. θ_C = 0.11, ϕ_C = 0.025, θ_O = 0.2, ϕ_O = 0.025, SlipER = 0.014, NFix = 3.
Source: COCAll_VM2eCS9c_110_025_200_025_014_3_CPN_200501

between objects produces a plethora of illusory targets, which would produce massive False Alarms unless the strategy includes some form of double-checking to confirm the presence of the actual target.

This was not at all an issue in the other tasks, where a single and unitary target property means that crowding scrambling could not cause an illusory target to appear on a negative display, and at most would change the apparent, but irrelevant, location of the target in a positive display. However, so prevalent are the illusory targets in Conjunction that the confirm-positive double-checking takes a long time on negative trials, so limiting the fixations is required to shorten the RTs without incurring more than an acceptable number of Misses. The same crowding produces illusory distractors that seem to require an occasional double-check before making an absent response as well.

As discussed above, the contrast between the Conjunction and Color task was a key motivator of the concept of feature binding performed by serial deployment covert attention. This explanatory sequence makes it clear that the Conjunction task is indeed very different from the Color task, but rather than additional cognitive mechanisms of feature binding performed by covert selective attention, the pattern of effects in the conjunction task can be explained much more simply in terms of availability and crowding producing ambiguity in the Conjunction task stimuli, which is not at all present in the Color and Shape task, and a more complex strategy is required to handle this ambiguity to produce reasonably fast and acceptably accurate responses.

4.5. Model Results for All Three Tasks

Using the final models in each explanatory sequence, Figure 4.5.1 shows the predicted RT and ER compared to the data, and Table 4.5.1 shows the strategy and parameter values for the model in each task condition, and the goodness-of-fit metrics for RT and ER in each task condition, and as computed for the whole set of 24 RT and 24 ER data points. The graphs show a very close correspondence between predicted and observed values, indicated quantitatively by the high r^2 values, low *aare,* and low *aae* (in ms and error proportion units). As pointed out above, because in the Color task, well-fitting predicted RTs are flat, the r^2 values for Color task RT have to be essentially zero, and thus are not included in the average of the r^2 values; but when the r^2 is computed for the fit to all 24 RTs from all three tasks, it is quite high because the models correctly predict both flat and sloped RTs.

However, notice that the confidence intervals around the means are quite large for some of the RT data, and also especially for the Miss ER data. The mean RT and ER data for individual subjects were inspected; there were substantial individual differences. For example, in Shape some subjects achieved almost error-free performance, while one subject produced 23% errors in the most difficult condition! Further inspection also showed different patterns in the RT effects for individual subjects in the same condition. The next section applies the models to account for these individual differences.

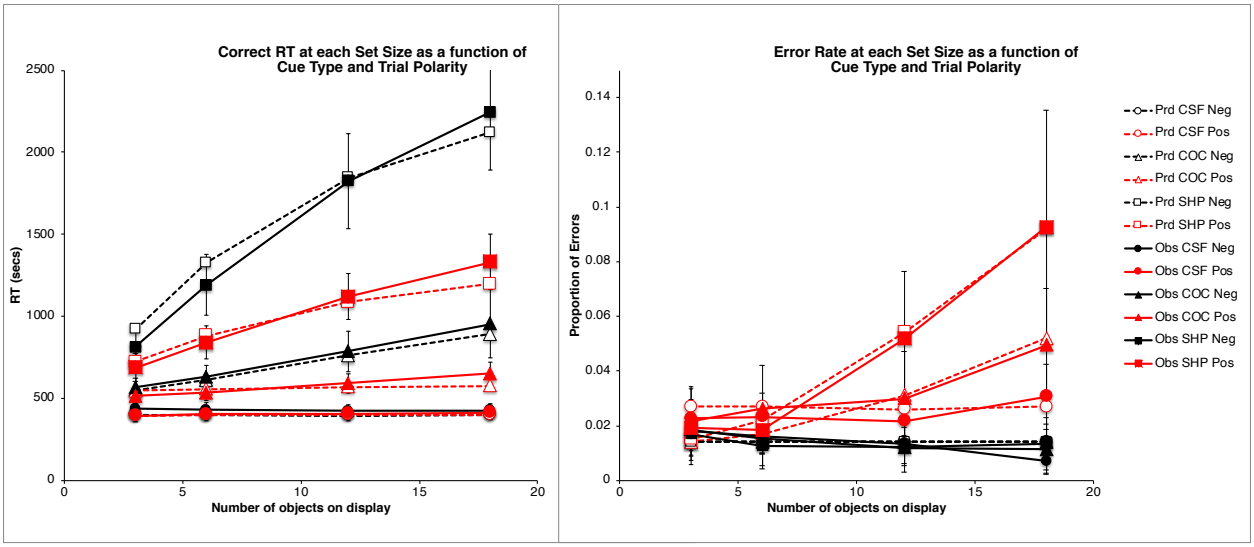


Figure 4.5.1. Predicted (open points, doted lines) and observed (solid points and lines) correct trial RT (left panel) and ER (right panel) for the aggregated data in each task condition. Shape: squares, Color: circles, Conjunction: triangles. positive trials: red, negative trials: black. The 95% confidence intervals are based on the standard error of the mean of the subjects' mean values underlying each data point and thus reflect between-subject variability. Source: AllSubs_VM2eLS9c_110_0_025_200_025_425_075_0_3_99_CPN_200506.

Table 4.5.1 Model Fits for Aggregated Data

		NFix	Color/Shape		Orientation		*SlipER*	GoF: RT			GoF: ER		
Task	Strategy		θ		θ	ϕ		r^2	*aare*	*aae*	r^2	*aare*	*aae*
Shape	Basic Unlimited		0.425	0.075			0.014	0.97	7%	79	0.99	11%	0.002
Color	Eyes Fixed		0.11	0.0			0.014	0.01	4%	18	0.68	25%	0.004
Conjunction	Basic Limited Confirm-both	3	0.11	0.025	0.20	0.025	0.014	0.95	5%	35	0.88	19%	0.004
Average* of fit metrics from each task								0.96	5%	44	0.85	18%	0.003
Fit metrics for all 48 data points								0.98	5%	44	0.95	19%	0.003
* *Averages of r^2 for RT do not include zero value for Color Task.*													

5. Modeling Search Performance for Individual Subjects

As is the case for many topics in human cognition and performance, there has been very little consideration of individual differences in visual search. But in fact, Wolfe, Cave, & Franzel (1989) present some results on individual differences in simple visual search. They provide tables of individual subject positive and negative trial RT slopes, and these slopes differ substantially between subjects in both magnitude and slope ratio. Also, Wolfe, et al. (2010), present individual subject RTs graphically superimposed on the average RTs. Again substantial differences between subjects are visible. Rather than examine why these differences could appear, the individual slopes were averaged together in Wolfe et al (1989), and in Wolfe et al. (2010) individual RT distributions were assumed to be sampled from the same underlying type of distribution. Thus theoretical work on visual search has addressed only the aggregate effects, as is the case in almost all cognitive psychology work, including the above modeling of the aggregate data. However, as noted above, there appear to be substantial individual differences underlying the aggregate data — is modeling the aggregate data misleading the theoretical work?

5.1. Representing Individual Differences in EPIC Models

The EPIC architecture provides a straightforward theoretical framework for characterizing individual differences in terms of two sources: (1) Individual subjects might have different parameters for architectural mechanisms — e.g. people clearly differ in visual acuity. Unfortunately, the Wolfe, et al. (2010) experiment was a between-subjects design, so there is no direct way to isolate this source of individual differences, but it is likely that models for individual subjects might require individual parameter values. (2) Individual subjects might devise different strategies for performing the same task. Earlier EPIC-inspired experimental work showed that individual strategy differences could produce powerful and misleading effects in the aggregate data (e.g. Schumacher, Seymour, Glass, Fencsik, Lauber, Kieras, & Meyer, 2001; Thompson, Iyer, Simpson, Wakefield, & Kieras, 2015).

It is possible that the EPIC aggregate-data models in fact do not account for how *any* individual subjects performed the search tasks, because averaging over subjects who have different architectural parameters and different task strategies could produce average data that actually represents none of the subjects. While this is a long-standing insight in cognitive modeling, as noted long ago by Reitman (1970) and Newell (1973), it is unusual for modelers to collect and model data reflecting this insight, and published data rarely allows the modeler to act on it. Fortunately, the Wolfe, et al. dataset includes all of the individual trials for each subject in each task condition, allowing a subject-by-subject analysis of the underlying individual differences.

More specifically, the seriously misleading effects of assuming individual subjects follow the same strategy was discussed by Reitman (1970) in the context of models of memory and other tasks: the strategies and the underlying memory and processing mechanisms were quite distinct, and effects due to strategy differences were customarily ignored and treated as if they were just "noise" or "error variance" in the data. Later, Newell (1973) followed up his First Injunction (see above) to know the subject's strategy with another:

> *Second Injunction of Psychological Experimentation:* Never average over methods.
>
> To do so conceals, rather than reveals. You get garbage or, even worse, spurious regularity. The classic example of the failure to heed this injunction is the averaging of single-shot learning curves to yield continuous learning curves.… much of the ability of the field continually and forever to dispute and question interpretations arises from the possibility of the subject's having done the task by a not-til-then-thought-of method or by the set of subjects having adopted a mixture of methods so the regularities produced were not what they seemed. (p. 294-296)

The main reason why individual differences in subjects' strategies may have arisen in the Wolfe et al.(2010) data is that like many experimental psychologists, Wolfe, et al. provided ambiguous instructions on speed vs. accuracy — respond "as quickly and accurately as possible". As Pachella (1974) noted, such conjoint instructions are, in essence, logically contradictory; i.e. responses "as accurate as possible" must necessarily lead to extremely long RTs, whereas responses "as quick as possible" must necessarily lead to extremely low accuracy. Faced with this conundrum, subjects necessarily have to choose some intermediate, individually variable, mixture of emphases on response accuracy versus speed. Further complicating the choices, Wolfe et al. (2010) paid subjects an apparently flat rate per hour. No systematic quantitative payoff or incentive scheme was used to influence the interpretation of the instructions in any particular way (cf. Edwards, 1961; Sternberg, 2016). However, Wolfe et al. did provide trial-by-trial accuracy feedback and extensive practice, both of which might induce subjects to adopt stable strategies, but which could still be very different due to how they individually interpreted the ambiguous instructions. For example, personality differences may have played a significant role here (Dickman & Meyer, 1988): because about 4000 trials were required for each subject, relatively impulsive subjects may have chosen to favor speed "to get it over with" and they were free to choose whatever level of errors was personally comfortable. In contrast, relatively reflective subjects may have been more conscientious and favored accuracy even if it took a long time.

It would be better not to leave such choices completely up to the subject's preferences. As discussed by Pachella (1974) and Sternberg (2016), RTs are most interpretable if ERs are reasonably low, so the experimenter should encourage subjects to respond as quickly as they can while making fairly few errors. This can be accomplished with an incentive scheme that heavily penalizes errors, which discourages trying to be too fast, but also penalizes unnecessarily long response times. For example, a good scheme gives the subject on each trial a starting number of points (ideally convertible to money), and deducts a certain amount for each millisecond of response time, but deducts the full starting amount for an error. In conjunction with running feedback on total "winnings," such a scheme can produce much cleaner data because individual subjects will converge to their best possible performance defined in terms of the payoff scheme (c.f. Schumacher, Lauber, Glass, Zurbriggen, Gmeindl, Kieras, & Meyer, 1999; Sternberg, 2016; Thompson, Iyer, Simpson, Wakefield, & Kieras, 2015).

However, at this time the Wolfe, et al. (2010) data are the best dataset available, and the success of the aggregate-data model presented below shows that it is quite systematic. It is worth examining whether systematic individual differences are present in the Wolfe et al. data, and

whether EPIC models can account for individual performance as well as the aggregate data. The above work on modeling aggregate data was not wasted, because at least it provides a repertory of mechanisms and strategies that might be useful in fitting individual data, even if the specific parameters and strategies differ between individuals.

5.2. Levels of Analysis for Modeling Individual Differences

One approach would be to construct a model for each individual subject, each with their own individual parameter values and strategies, yielding in this case 28 separate models. Not only with this be a very labor-intensive effort, but to be intelligible and scientifically useful, some kind of characterization would be needed about how these 28 models were similar and different. That is, the first question about so many models would be whether there were some commonalities, such as similar strategies between individual subjects in each task. Answering such questions would require organizing the models into groups of some sort.

Instead, we followed a more economical approach to arriving at the generalizations: We identified commonalities in the individual subject data by finding subgroups, or *clusters*, of subjects in each task condition whose mean RTs and ERs were similar in magnitude and pattern of effects, and thus likely to have similar parameter values and strategies. We then averaged over the subjects in each cluster; this mean data then represented a particular combination of parameter values and strategy. We then constructed a model for each cluster's mean data. Note this approach conforms to Newell's (1973) injunction to not average over strategies — we averaged over subjects who appear to be following the same strategies, and also appear to have similar parameter values. In what follows, we demonstrate that the parameters and strategies needed for the Wolfe, et al. aggregate data do indeed generalize to subgroups of individual subjects who performed similarly to each other. This satisfying outcome was accomplished as described in what follows.

Cluster analysis of Wolfe, et al. subjects. To evaluate whether there were meaningful systematic individual differences underlying the effects in the aggregate data, mean RT and ER graphs were plotted for each individual subject. On informal inspection, these graphs seemed to fall into a small set of patterns for each task condition. This impression was confirmed more formally by doing cluster analysis of the subjects' data in each task condition, using the intercept, slope, and best-fit quadratic coefficient for positive and negative RT, and mean ER, Miss ER and Maximum Miss ER at set size 18, as variables to identify the clusters. Of course, cluster analysis is usually done on very large data sets, so its application here is just a way to formalize what would otherwise be a purely intuitive hand-clustering process. However, there were about 500 trials per data point per subject, so the individual subject data were reliable enough to somewhat mitigate the fact that only 9 or 10 subjects were present in each task condition.

The analysis was done with the **R cluster** package (R Core Team, 2017; Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2017) the clustering method was *k*-medoid, and the clusters for k=1 through 5 were determined. The analysis was separate for each task condition, and the variables used in the clustering were made as few as possible consistent with a clear clustering result for k=3. For example, the quadratic coefficient for RT was used as a clustering variable only in the Shape condition, where some of the subjects had strongly negatively accelerated RT functions. In addition, the final clusters were chosen so that averaging the RT and ER data within a cluster preserved the basic qualitative and quantitative trend patterns present for

the individual subjects in the cluster. This required moving a total of two subjects out of the computed clusters into a cluster of one subject each, as will be described below. These single-subject clusters were not included in the model fitting.

The following sections will present, for each task condition, graphs of RT and ER for individual subjects in their clusters, followed by the mean data for each cluster, and finally the results of fitting a model to each cluster.

To anticipate an important conclusion from what follows, the models for the clusters of subjects in each task use the same repertory of visual mechanisms and strategies that accounted for the aggregated data summarized above in Figure 4.5.1 and Table 4.5.1. In fact, in most cases, the only difference between the aggregate and the cluster models are the parameter values. This means that the modeling of the aggregate data was helpful in laying out the possible mechanisms and strategies, even if the cluster-level data has important differences from the aggregate.

The results will be presented for each condition in the same order as the explanatory sequences, starting with the Shape task, next Color, and finally Conjunction. The models were constructed following the principles in the above explanatory sequences, but in the interest of brevity, only the best-fitting final model for each cluster is presented.

5.3. Shape Clusters

For each task, starting here with the Shape task, the following will be presented: (1) the mean data for each subject in each cluster will be graphed together to demonstrate that they actually resemble each other; (2) the individual data in each cluster are averaged together, and shown as a graph, and summary statistics in a table; (3) a model is fit to the mean data for each cluster; (4) a table of parameters and goodness-of-fits statistics for the cluster models is presented along with a summary assessment.

Individual data in each cluster. Figure 5.3.1 shows the four groups of subjects resulting from the cluster analysis of the Shape subjects; each column is a cluster with RT on the top and ER on the bottom; as before, red curves are positive trials, black curves are negative trials. The curves for different subjects are shown by different plotting points (circles, triangles, squares, diamonds). Throughout these presentations, since ER differences tend to easier to describe than RT differences, clusters are shown in order of increasing ER from left to right. Each cluster is labeled in terms of a short description that summarizes its ER and RT characteristics. In Figure 5.3.1, there is a single-subject cluster at the far right that was not included in the modeling. Note that the plotting scales cover the very large RTs and ERs produced in these clusters.

The leftmost cluster *Slow/LowER*, has three subjects with fairly linear RTs with large slopes, but low ER with a small increase for positive ER with set size. The next cluster *VerySlow/ MedER*, has two subjects with much more steeply sloped RTs and a much greater increase in Miss ER along with an extremely low False Alarm ER. The next cluster *NegAcc/HighER*, has relatively fast but negatively accelerated RTs with very much larger ERs, and Misses that greatly increase with set size. The subject in the right-most single-subject cluster was originally grouped with the leftmost cluster, but because the RT slopes for this subject were clearly much lower than the other three subjects and the ER higher, this subject was moved into a single-subject cluster which was not modeled in this report.

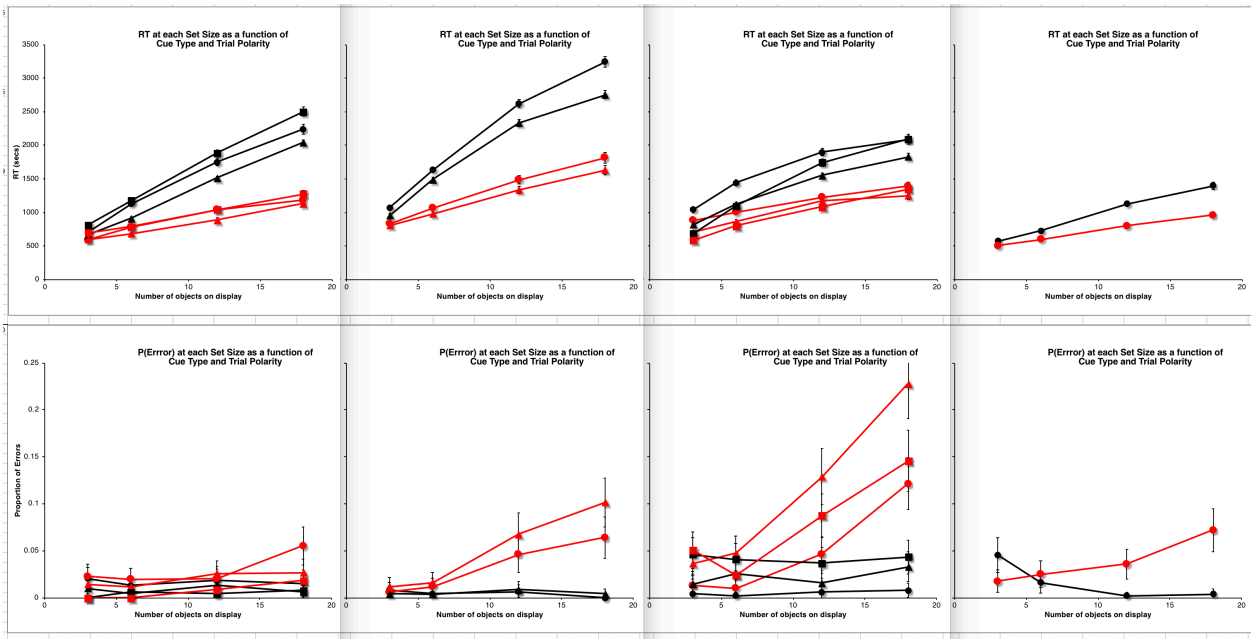


Figure 5.3.1. Individual subject RT and ER in each cluster of the Shape subjects. Scales: 0-3500 ms, 0-0.25. The 95% confidence intervals are calculated for each individual subject, based on the standard error of the subject's mean for RT, and as the binomial confidence interval for ER. There are approximately 500 trials underlying each plotted data point for each individual subject.

Mean RTs and ERs for the clusters. Figure 5.3.2 shows the average RT and ER for the three modeled clusters, and Table 5.3.1 provides their statistics. Note how the basic pattern of the effects for the individual subjects in the cluster is reflected in the means for that cluster, and how the confidence intervals for the data points in a cluster are fairly tight compared to what appeared previously in the overall aggregate data for the Shape task (Figure 2.1). Thus even though there are only a few subjects being averaged together, the subjects in a cluster are similar enough to each other that their average data tends less noisy than in the aggregated data.

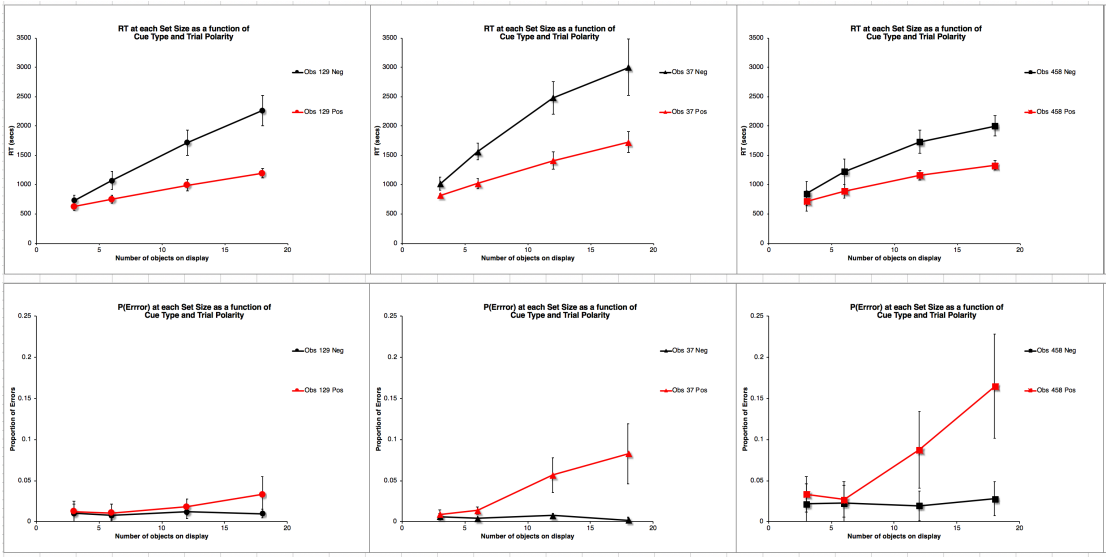


Figure 5.3.2. Mean RT and ER in each cluster of the Shape subjects. Scales: 0-3500 ms, 0-0.25. The 95% confidence intervals on each plotted point are based on the mean RT and ER for that data point for each of the 2-3 subjects included in the cluster.

Table 5.3.1 Observed Shape Cluster Statistics

	Negative				Positive					Slope ratio
Shape Cluster	Intercept	Slope	r^2	ER	Intercept	Slope	r^2	ER	ER Max	
Slow/LowER	446	102	1.00	0.010	521	38	1.00	0.019	0.033	2.70
VerySlow/MedER	718	133	0.98	0.005	651	61	1.00	0.041	0.083	2.19
NegAcc/HighER	707	76	0.96	0.023	628	41	0.98	0.078	0.165	1.87

Model fits for the mean data in each cluster. An EPIC model was fit to each cluster following the same approach described for the aggregate data in Section 4.2 above. In each clusters, *SlipER* was set to the mean False Alarm ER for that cluster. The obtained model fits for the Shape cluster are shown in Figure 5.3.3, and their parameters and goodness-of-fit statistics are listed in Table 3.6. These fits demonstrate strong differences in the visual parameters but the same Basic Search strategy across the clusters of similar subjects. Each cluster is discussed in the following paragraphs.

Cluster Slow/LowER: Basic Search strategy, moderate availability and crowding, low SlipER. The subjects in the leftmost cluster Slow/LowER have steeply sloped RTs and low ER with slightly increasing positive ER. This fit was obtained with the Basic Search strategy with unlimited fixations and somewhat poor availability of the Shape property, but cautiously low *SlipER*. These subjects could be described as needing many fixations because they could not see the object shapes very well in the periphery, but they took their time and achieved fairly high accuracy. Since the fixations are unlimited, the increased Miss ER at greater set sizes is due to crowding effects changing the target into an illusory distractor. As in the overall average data, the model predicts somewhat negatively accelerated RTs on negative trials because more than one object can be perceived in a single fixation as the object density increases with set size. This is a systematic misfit of the model, but as shown in Table 5.3.2, the r^2 for RT is very high nonetheless.

Cluster VerySlow/MedER: Basic Search strategy, very poor availability, moderate crowding, very low SlipER. The subjects in the middle panel of Figure 5.3.3 have very steeply sloped RTs and the Miss ER is medium-high and increases substantially with set size compared to the first cluster. This fit was also obtained with the Basic Search strategy with unlimited fixations, but as shown in Table 5.3.2, with larger availability and crowding parameters than the first cluster. Since the Shape property is quite a bit less available than in the first cluster, the predicted RTs are more linear and match the observed RTs quite well. The larger crowding parameter also produces the higher Miss ER. The goodness of fit metrics are extremely good. These subjects apparently had a great deal of trouble detecting the Shape property and so had to make more fixations than in the first cluster, but also were more subject to Misses from crowding effects. The large value for ER *aare* is a good example of the difficulty posed by very low values in the observed data; since these values appear in the denominator of the relative error calculation, even a small absolute deviation of the predicted from the observed value will produce a large relative error. In contrast, the r^2 and *aae* show a fairly close fit.

Cluster NegAcc/HighER: Basic Search strategy, moderate availability, high crowding, moderate SlipER. As shown in the rightmost panel of Figure 5.3.3, these subjects produced RTs that are fairly fast, but strongly negatively accelerated, and the ERs are higher than in the first two Shape clusters and especially, the Miss ER is much higher and increases with set size more

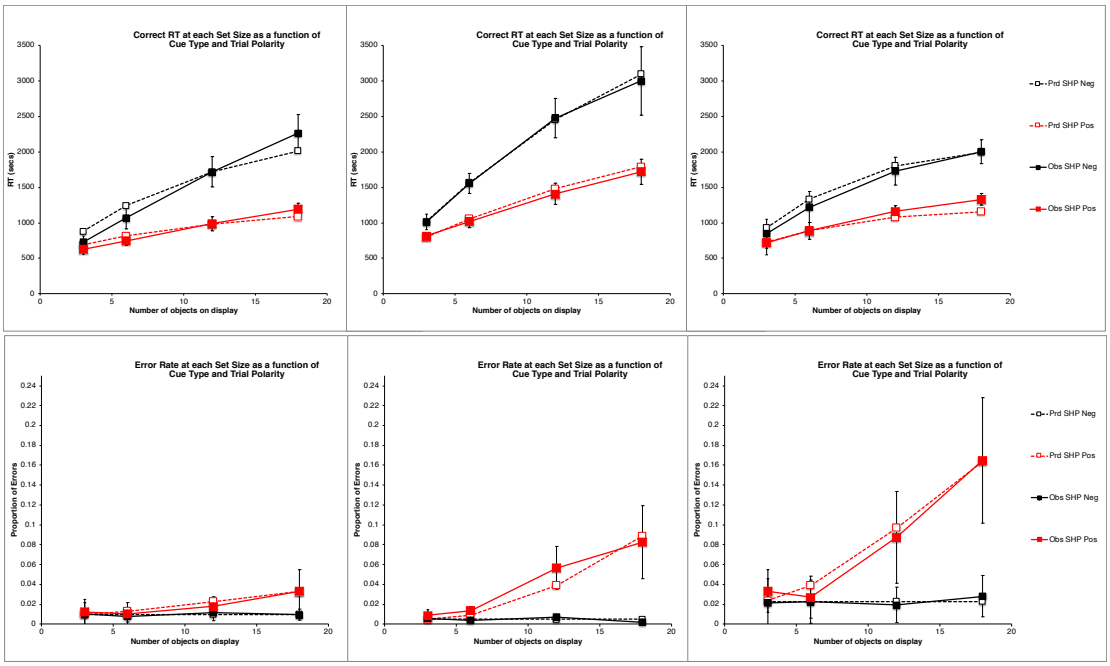


Figure 5.3.3. Observed (solid lines and points) and Predicted (dotted lines, open points) for the Shape Cluster Means. Upper panels: RT, scale: 0-3500 ms. Lower panels: ER, scale 0-0.25. Left panel: Slow/LowER. Middle panel: VerySlow/MedER. Right panel: NegAcc/HighER.

Sources: SHP129_VM2eLS9c_375_025_01_99_NC_200309, SHP37_VM2eL9c_6_05_005_99_NC_200309, SHP458_VM2eLS9c_435_15_023_99_NC_200604.

than any other cluster in any other condition. The fit captures these effects with substantial crowding effects with some help from a somewhat elevated *SlipER*. As shown in Table 3.6, this fit was obtained with moderate poor availability, a large crowding probability, and high *SlipER*. This combination produces the powerful trend in Miss ER, as well as the negatively accelerated RTs; the crowding produces many illusory distractors which truncate the search and produce many Misses.

A possible alternative model for this cluster is one similar to the *path not taken* model in the Shape explanatory sequence, namely a strategy with fixations limited to 8 and smaller θs and ϕs, which produces very similar predictions and goodness of fit; thus there is more than one way to get negatively accelerated RTs. But as noted above, introducing a strategy change when visual parameters produce a good fit violates the priority order in the explanatory sequence. In fact,

Table 5.3.2

	Shape				*SlipER*	GoF: RT			ER		
Shape Cluster **Basic Search strategy**	θ	ϕ				r^2	*aare*	*aae*	r^2	*aare*	*aae*
Slow/LowER	0.375	0.025			0.01	0.96	9%	101	0.92	15%	0.002
VerySlow/MedER	0.6	0.05			0.005	1.00	3%	41	0.94	43%	0.005
NegAcc/HighER	0.435	0.150			0.023	0.96	6%	67	0.98	16%	0.005
Average fit metrics						0.97	6%	70	0.94	25%	0.004

Figure 4.2.3 above shows that parameter values in this range produce such RT and ER patterns, which did not appear in the aggregate data.

Summary assessment of models for the Shape clusters. As shown by the average fit metrics in Table 5.3.2, on the whole, the model goodness-of-fit metrics are extremely good for the Shape clusters. The model accounts for the clusters in a straightforward way: All three clusters differ in terms of visual parameters and *SlipER* but use the same Basic Search strategy as in the aggregate data. Thus the individual differences in task performance were due to parameter differences only. Despite this similarity of the models, it is clear that the aggregate data obscured the rather large differences in the individual parameter values that cause the individual cluster RT and ER functions to have very different characteristics.

5.4. Color Task Clusters

Individual data in each cluster. Figure 5.4.1 shows the three groups of subjects resulting from the cluster analysis of the Color task subjects. As before, each column is a cluster with RT on the top and ER on the bottom; red curves are positive trials, black curves are negative trials. The curves for different subjects are shown by different plotting points (circles, triangles, squares, diamonds). The clusters are shown in order of increasing ER from left to right. In what follows, each cluster is labeled in terms of a short description. The plotting scales cover the range for RT and ER used for most graphs in this report.

All of the clusters have fast RTs (about 500 ms) that are quite flat with set size and very similar for positive and negative trials. Reading from the left, the first cluster *FlatSlow/LowER* contains two subjects who have the slowest RT and very low ER that is essentially unaffected by set size. The second cluster *FlatFast/MedER* of three subjects has fast RTs and low and flat negative ER, and positive ER that tends somewhat to increase with set size. The rightmost

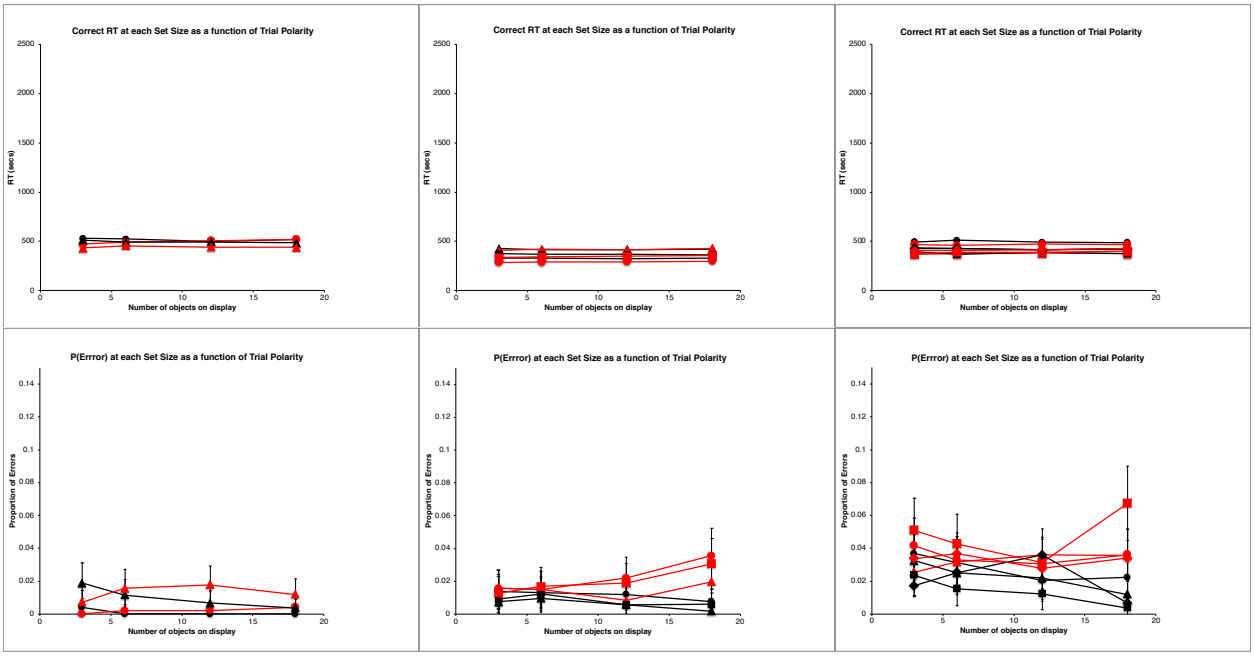


Figure 5.4.1. Individual subject RT and ER in each cluster of the Color task subjects. Upper panels: RT, scale: 0-2500 ms. Lower panels: ER, scale: 0-0.15. The 95% confidence intervals are calculated as described for Figure 5.3.1.

cluster *FlatFast/HighER* of four subjects shows much higher and varied ERs but show little overall effect of set size.

Mean RTs and ERs for the clusters. Figure 5.4.2 shows the average RT and ER for each of these three clusters; and Table 5.4.1 shows the statistics for each cluster based on those means. The slope ratios are not meaningful because the positive and negative RT slopes are very small and noisy. As before, note how the basic pattern of the effects for the individual subjects in the cluster is reflected in the means for that cluster, and how the confidence intervals for the data points in a cluster are fairly tight.

Model fits: All Color task clusters fit with Fixed-eye strategy, different but high availabilities, different SlipERs. As shown in Figure 5.4.3, for all three clusters, the Fixed-eye strategy provided a good fit, with only the Color availability parameter and the *SlipER* adjusted to fit each cluster. Table 5.4.2 provides the parameter values. As discussed above for the overall average data, the Color crowding probability parameter ϕ_C makes a negligible effect; it was set to a place-holder value of 0.025 to be consistent with the Conjunction models.

As in the overall Color task model, once *SlipER* was set to the False Alarm ER, the Miss ERs were accounted for by adjusting the availability parameter. Note that because the predicted and observed RTs are flat and similar for positive and negative trials, the r^2 for RT fits is constrained to be very small — when there is no variability in the observed data, then there is no variance for

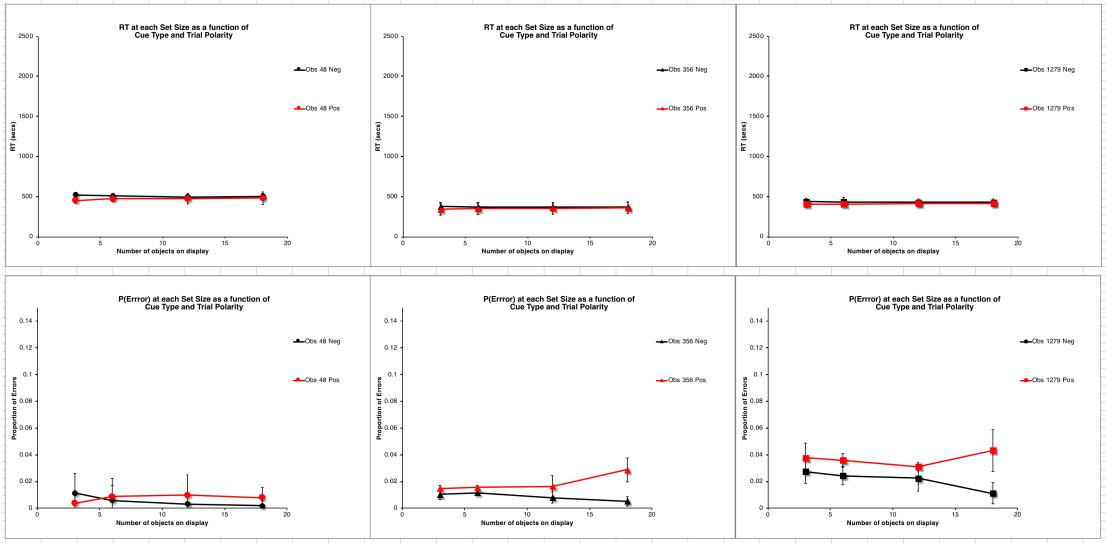


Figure 5.4.2. Mean RT and ER in each cluster of the Color task subjects. The 95% confidence intervals on each plotted point are based on the mean RT and ER for that data point for each of the 2-4 subjects included in the cluster.

Table 5.4.1 Observed Color Cluster Statistics

	Negative				Positive					Slope ratio
Color Task Cluster	Intercept	Slope	r^2	ER	Intercept	Slope	r^2	ER	ER Max	
FlatSlow/LowER	520	-1.30	0.58	0.01	456	1.49	0.70	0.01	0.01	-0.87
FlatFast/MedER	375	-0.24	-0.32	0.01	341	1.08	0.89	0.02	0.03	-0.23
FlatFast/HighER	440	-0.78	-0.86	0.01	341	1.08	0.89	0.02	0.043	-1.00

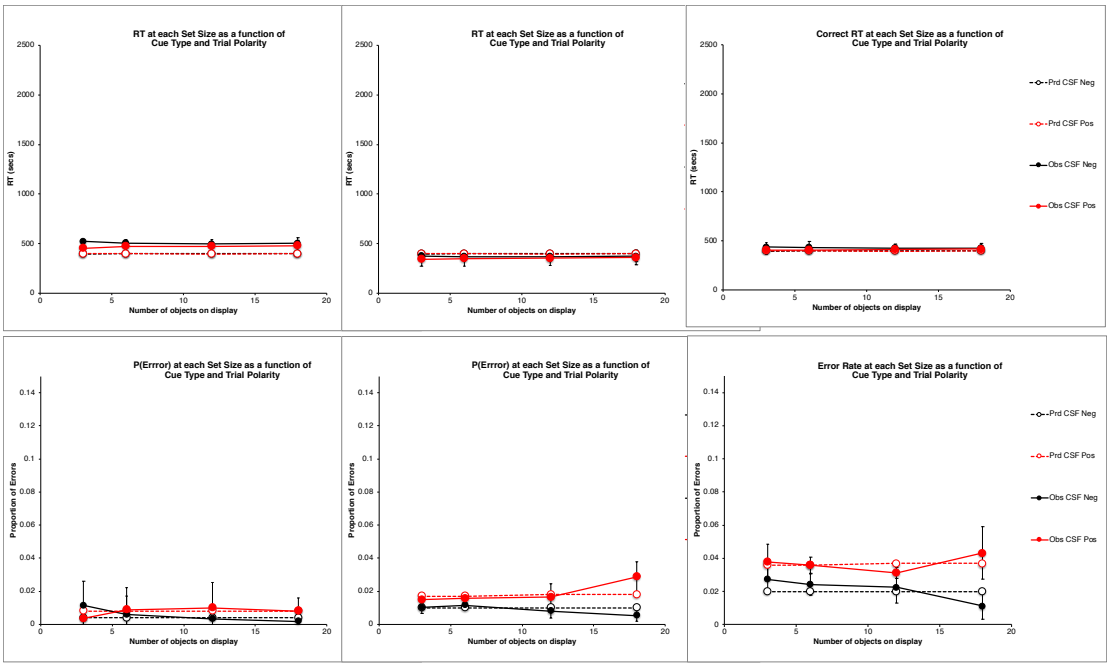


Figure 5.4.3. Observed (solid lines and points) and Predicted (dotted lines, open points) for the Color task Cluster Means. Upper panels: RT, Scale: 0-2500 ms. Lower panels: ER. Scale: 0-0.15. Left panel: FlatSlow/LowER. Middle panel: FlatFast/MedER. Right panel: FlatFast/HighER.

Sources: CSF_48_Vm2eLS9c_09_025_004_0_NC_200409, CSF_356_VM2eLS9c_1_025_01_0_NC_200416, CSF_1279_VM2eLS9c_115_025_02_0_NC_200416

the prediction to account for! The RT fit for the first cluster FlatSlow/LowER could be improved by setting the *VDelay* parameter to a higher value instead of the assumed value of 100 ms in all tasks and clusters. The slight upward trend in Misses at set size 18 for the second and third clusters is similar to the aggregate data, and there is no straightforward way for the models to account for this small discrepancy.

Summary assessment of models for the Color task clusters. As shown in Table 5.4.2, the fits for the Color task clusters are overall very good. The account for the Color task clusters is very simple: As in the aggregate data, all subjects followed the Fixed-Eye strategy, which made the RTs flat and fast, but differences in *SlipER* and Color availability produced distinct levels of Miss ER. As with the aggregate data model, ϕ_C had a negligible effect. But θ_C varied over a small range, implying that Color availability was very high for individual subjects.

Table 5.4.2

	Color				*SlipER*	GoF: RT			ER		
Color task Cluster Fixed-eye Strategy	θ	ϕ				r^2	*aare*	*aae*	r^2	*aare*	*aae*
FlatSlow/LowER	0.09	0.025			0.004	0.05	18%	89	0.10	48%	0.002
FlatFast/MedER	0.10	0.025			0.010	0.08	11%	38	0.63	25%	0.003
FlatFast/HighER	0.115	0.025			0.020	0.10	5%	22	0.69	22%	0.005
Average fit metrics						0.08	11%	50	0.47	31%	0.003

5.5. Conjunction Task Clusters

Individual data in each cluster. Figure 5.5.1 shows the groups of subjects resulting from the cluster analysis of the Conjunction subjects; again, each cluster is shown in a column with RT above and ER below, and are shown in order of increasing ER from left to right, with the exception of the single-subject cluster at the far right.

The first cluster *Sloped/LowER* contains two subjects with very low ERs and RTs that have substantially greater slope than the aggregate data. The second cluster, *AlmostFlat/MedER*, contains four subjects with almost flat RTs and Miss ER increasing with set size. The third cluster *AlmostFlat/HighER* has three subjects who also have almost flat RTs but with very high ERs, both False Alarms and Misses, with a tendency for Miss ER to increase with set size. This cluster is clearly more heterogenous than the first two in both RTs and ERs. The right-most single-subject cluster has RTs very similar to the first cluster and was grouped in that cluster by the analysis; however, the ERs are much higher and trended more with set size than the first cluster. To meet the criterion of not producing misleading mean values for a cluster, this subject was moved out of the first cluster to become a single-subject cluster which was not modeled for this report.

Mean RTs and ERs for the clusters. Figure 5.5.2 shows the average RT and ER for the first three clusters, and Table 5.5.1 shows the statistics for these average data. The first cluster, Sloped/LowER is very different from the second and third in that it has steeply sloped RTs, especially for negative trials, and extremely low ER. The second and third clusters have very similar and almost flat RT slopes, but very different ERs. Note that because the third cluster AlmostFlat/HighER is more heterogenous than the other clusters in this data set, the confidence intervals around this average data are relatively large. Clearly the aggregate data misrepresented the individual subjects in the Conjunction condition; the RT slopes and ER are very different for the first cluster compared to the other two, which have similar RTs but very different ERs.

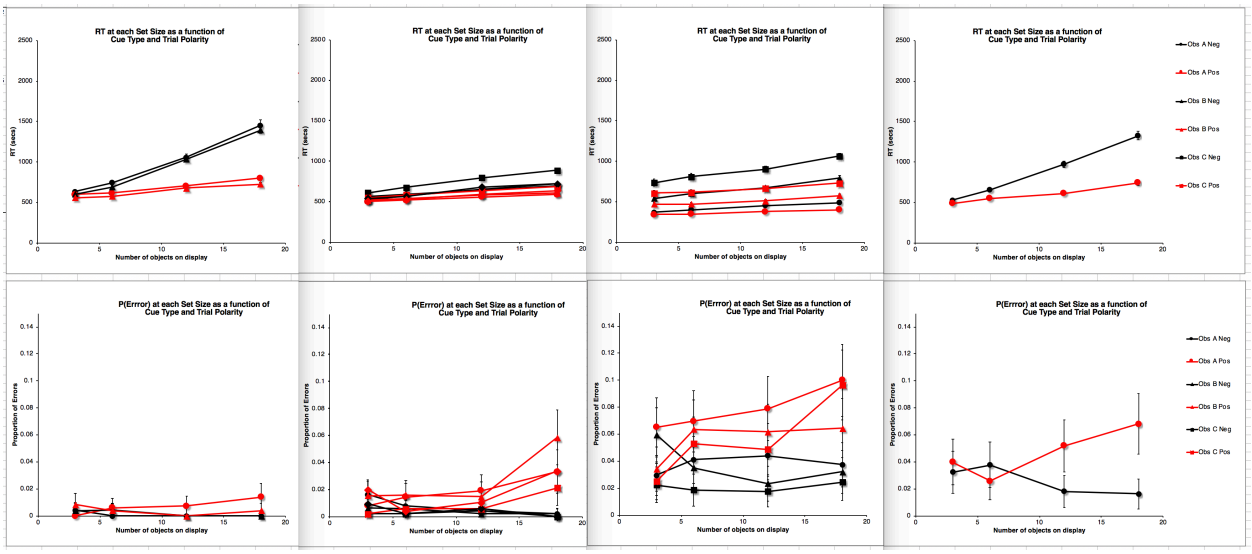


Figure 5.5.1. Individual subject RT (upper panels), scale 0-2500 ms, and ER (lower panels), scale 0-0.15, in each cluster of the Conjunction subjects. The 95% confidence intervals are calculated as described for Figure 5.3.1.

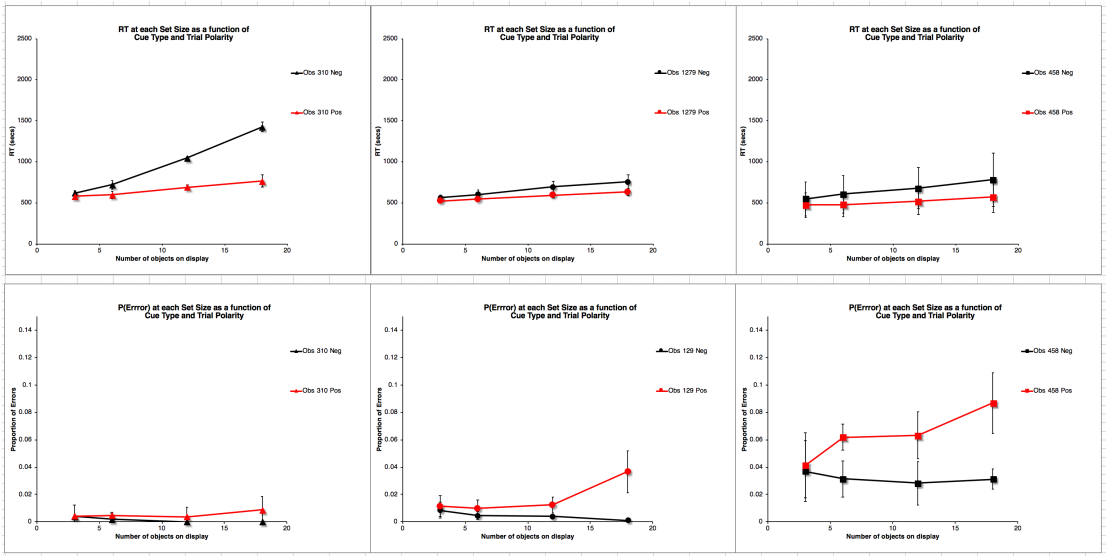


Figure 5.5.2. Mean RT and ER in each modeled cluster of the Conjunction subjects. RT (upper panels), scale 0-2500 ms, and ER (lower panels), scale 0-0.15. Left panel: Sloped/LowER. Middle panel: AlmostFlat/MedER. Right panel: AlmostFlat/HighER. The 95% confidence intervals on each plotted point are based on the mean RT and ER for that data point for each of the 2-4 subjects included in the cluster.

Table 5.5.1 Observed Conjunction Cluster Statistics

	negative				positive					Slope ratio
Conjunction Cluster	Intercept	Slope	r^2	ER	Intercept	Slope	r^2	ER	ER Max	
Sloped/LowER	420	54	0.99	0.001	531	13	0.99	0.005	0.009	4.21
AlmostFlat/MedER	524	13	0.99	0.004	499	7	1.00	0.018	0.037	1.78
AlmostFlat/HighER	505	15	0.99	0.03	446	7	0.97	0.063	0.086	2.31

Model fits for the mean data in each cluster. The model fits for the Conjunction clusters are shown in Figure 5.5.3, and the parameters and goodness-of-fit statistics are listed in Table 5.5.2. The plotting scales are the same as those used for the Color task clusters and most of the other graphs.

Cluster Sloped/LowER: Basic Search with Confirm-both and unlimited fixations, poor availability, low crowding, very low SlipER. The leftmost panel in Figure 5.5.3 shows the predicted and observed RTs and ERs for the Sloped/LowER cluster. The model for this cluster fits extremely well. As noted above, a very low observed ER can lead to a very large *aare* value, in this case, infinitely large since a couple of the observed ERs are actually zero; accordingly, this cell in Table 5.5.2 includes only the cases where the observed ER is greater than zero. The *aae* provides a good indication that the predicted ERs are very close to the observed.

The data and the strategy choice for this cluster are very different from the aggregate Conjunction data and model. Apparently these subjects had some difficulty seeing the Color and Orientation combinations, as shown by the parameter values, and chose to minimize errors with a very methodical search and great care to prevent slip errors.

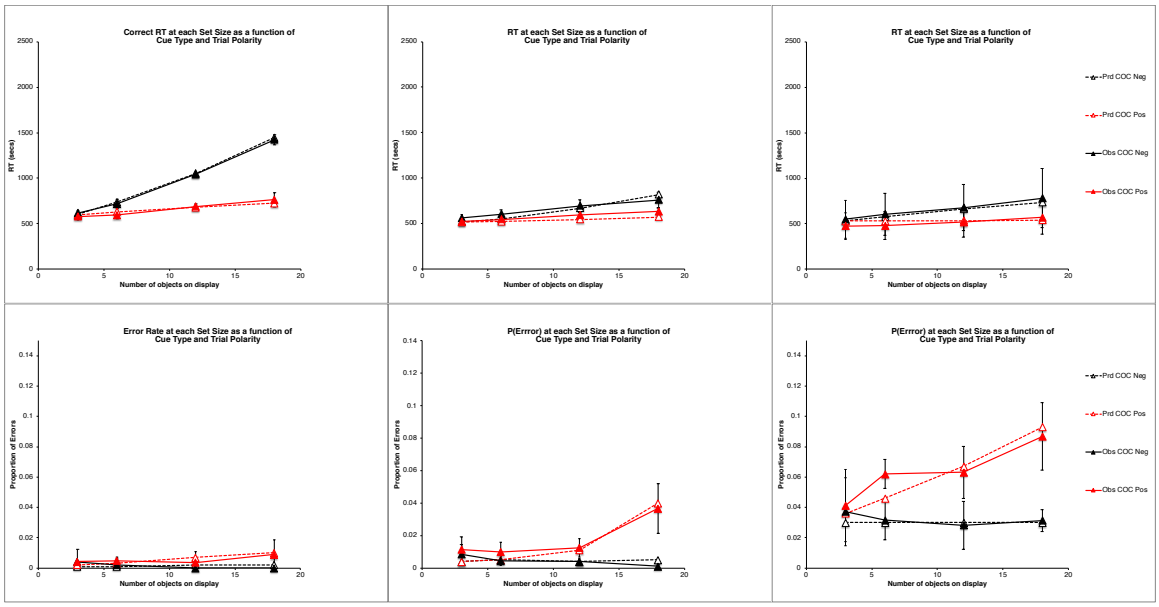


Figure 5.5.3. Observed (solid lines and points) and Predicted (dotted lines, open points) for the Conjunction Cluster Means. RT (upper panels), scale 0-2500 ms, and ER (lower panels), scale 0-0.15. Left panel: Sloped/LowER. Middle panel: AlmostFlat/MedER. Right panel: AlmostFlat/HighER.
Sources: COC_310_VM2eLS9c_15_025_25_05_0015_99_CPN_200309, COC_1279_VM2eLS9c_1_075_15_075_0045_3_1_1_99_CPN_200512, COC_458_VM2eLS9c_11_025_18_025_03_2_1_1_99_CPN_200512.

AlmostFlat/MedER: Basic Search with Confirm-both with limit of three fixations, high availability, moderate crowding, low SlipER. The middle panel shows the fit to cluster AlmostFlat/MedER. This cluster used the same strategy as the model for the aggregate data in which the fixations were limited to three to produce both fast and fairly flat RTs and Miss rates that increase with set size. These subjects had very good Color and Orientation availability, leading to fairly fast reaction times overall, and moderate crowding probabilities. Apparently these subjects could resolve the stimuli better than the first Conjunction cluster, and so opted to maximize speed by limiting the number of fixations, and because they were careful to avoid slip errors, they were able to perform quickly with fairly few Misses.

AlmostFlat/HighER: Basic Search with Confirm-both with limit of only two fixations, moderate availability, low crowding, high SlipER. The rightmost panel shows cluster AlmostFlat/HighER. In this model, the number of fixations limit is two, which together with somewhat lower availabilities and the high *SlipER*, produces a higher Miss ER compared to the second cluster fit. These subjects also apparently sought speed at the expense of accuracy, more so than the second cluster, and allowed even fewer fixations and took much less care in avoiding slip errors.

Summary assessment of models for the Conjunction clusters. All of the cluster models use the Basic Search with Confirm-both strategy presented in the explanatory sequence for the aggregate Conjunction data; without the Confirm option, models for these clusters suffer from the same massive False Alarm rates as discussed in the explanatory sequence. But the cluster models differ in the availability, crowding, and *SlipER* parameters, but especially on the extent to which the number of fixations is limited. In particular, the model for the first cluster, Sloped/LowER, has unlimited fixations, very different from the aggregate data model. But like the

Table 5.5.2

Conjunction Cluster and Strategy	Color		Orientation		*NFix*	*SlipER*	GoF:RT			GoF:ER		
	θ	ϕ	θ	ϕ			r^2	*aare*	*aae*	r^2	*aare*	*aae*
Sloped/LowER Basic Confirm-both Unlimited-fixations	0.15	0.025	0.25	0.05		0.0015	0.99	3%	21	0.51	53%	0.002
AlmostFlat/MedER Basic Confirm-both Limited-fixations	0.10	0.075	0.15	0.075	3	0.0045	0.90	6%	40	0.90	76%	0.003
AlmostFlat/HighER Basic Confirm-both Limited-fixations	0.11	0.025	0.18	0.025	2	0.03	0.91	6%	32	0.91	11%	0.005
Average fit metrics							0.93	5%	31	0.77	47%	0.003

model for the aggregate Conjunction data, the models for the second and third clusters, AlmostFlat/MedER, and AlmostFlat/HighER, limit fixations, and thus have almost flat RTs and ERs that increase with set size. The huge difference between the first cluster and the other other two is a strong argument that the aggregate data was hiding a substantial difference between subject strategies. The average goodness-of-fit metrics in Table 5.5.2 for these three clusters on the whole are fairly good, taking into account that a couple of the ER *aare* values are high due to very small observed ERs.

As discussed above with the aggregate Conjunction modeling, the Conjunction task is inherently more complicated in its strategy requirements which are dictated by how crowding effects render the two-feature stimuli ambiguous compared to the single-feature tasks. Subjects responded to this complexity in different ways. While one cluster was very slow and highly accurate, the other two tried to be very fast, as shown by the small number of allowed fixations, but with poor accuracy.

The Conjunction clusters provide a good demonstration of the perils of aggregating data and attempting to model it in the aggregate, especially when the lack of clear instructions and incentives encourages subjects to choose their strategy rather freely. Apparently, the AlmostFlat subjects, faced with 4000 trials of this fairly uninspiring task, opted to "get it done" even at the expense of more errors, while the Sloped/LowER subjects rose to the challenge of minimizing errors, even if it took much longer. Although the Conjunction aggregate model resembles the model for the second and third Conjunction cluster data, it completely misrepresents the data for the first cluster.

5.6. Summary of model fits for aggregated and clustered data

Table 5.6.1 shows a summary of the goodness of fit metrics for all of the presented models. The first two rows corresponded to the last two row in Table 4.5.1 above, showing the goodness of fit metrics averaged over the three final models of the aggregated subject data in the three task conditions, and the fit to all 48 points of the aggregated data. The third row shows the metrics averaged over the presented models for the three clusters of subjects in each task condition, corresponding to the averages of the last rows in Table 5.2.2, Table 5.4.2, and Table 5.5.2. As

Table 5.6.1 Summary of Goodness-of-Fit for Aggregated and Cluster Data

Source of model fit metrics	GoF: RT			ER		
	r^2	*aare*	*aae*	r^2	*aare*	*aae*
Average* of separate fits of each task with aggregated data	0.96	5%	44	0.85	18%	0.003
Overall fit of all aggregated data (all 48 mean data points)	0.98	5%	44	0.95	19%	0.003
Average* of separate cluster fits for all three tasks	0.95	7%	50	0.73	34%	0.004
** Averages of r^2 for RT do not include zero value for Color task tasks.*						

mentioned above, not included in these averages are the zero r^2 values for the flat RT curves in the Color Task. The average fit is extremely good for RT, with very high r^2 values of 0.95 and above, *aare* under 10%, and *aae* on the order of only 50 ms. ER, being intrinsically less stable in these data, was not fit as well, especially when the different clusters differed in their *SlipER*. For ER, the r^2 values are reasonably high at 0.73 and above, but the ER *aare* values are inflated by some very small observed ERs; more encouragingly, for the average absolute error *aae*, the average predicted values were off by rather less than half a percentage point.

According to the EPIC architecture, individual subjects could differ in architectural parameters and task strategy. In the presented fits, there were two perceptual parameters for each property: the availability threshold coefficient θ and the crowding probability ϕ, and a single adjusted motor parameter, the *SlipER*, always estimated directly from the False Alarm ER. There were three fundamentally different task strategies, the Basic Search strategy for Shape, the Fixed-eye strategy, required for the Color task condition, and Basic Search with a double-check confirmation step, and usually a limit on the number of fixations, for Conjunction.

Thus the same set of strategy variations and perceptual/motor mechanisms that could account for the aggregate data also could account for the subject clusters in all three tasks. With the single exception of one of the Conjunction clusters, the strategy variation that fit the aggregate data also fit the clustered data. Thus parameter differences accounted for most of the differences between the clusters and the aggregate data for that task, but also for most of the differences between clusters within a task, showing that parameter differences can produce very different RT and ER effects in the data. The choice of strategies and parameters produced a good set of fits for the overall average data in the three task conditions, and the three clusters of subjects that appeared in each of those three conditions.

6. Conclusions and Implications

The present research has yielded conceptual insights, theoretical implications, and valuable lessons regarding the essential nature of simple visual search and how to model it. This work strongly reinforces previous claims that in order to understand, explain, and predict basic high-speed human performance, the framework provided by a principled, parsimonious, realistic, empirically-based cognitive architecture like EPIC can be extremely useful and instructive. In the case of simple visual search, such an architecture also leads to important implications for experimental methodology, not only for empirical investigations of simple visual search, but also more broadly to experimental studies regarding other types of human performance.

6.1. Overall results

The EPIC model fits are extremely good, both qualitatively and quantitatively — the models predict actual quantitative values, not just trends. They match the observed data very closely, and do so using remarkably simple empirically-supported perceptual/motor mechanisms that interact in a straight-forward manner with plausible, rationally-justified task strategies.

The major difference between the modeled search task conditions is the strategy required to perform the task given the perceptual properties of the task displays: A slow and systematic search with lots of fixations for the Shape task, where the relevant property is not very available; a fast no-fixation strategy for the Color task, where the relevant property is extremely available; a double-checked and time-out strategy for the Conjunction task to deal with how crowding produces many illusory combinations of properties.

It should be noted that the model strategies are all rational and simple ways for subjects to achieve performance at a desired speed or accuracy given the task and displays confronting them; there are no "fiddly" or arbitrary "kludges" in these strategies to force them to fit the data. Because architectures like EPIC enable explicit and programmable task strategies, different strategies are easy to explore, implement, evaluate, explain, and justify.

Future work with cognitive modeling should take these lessons to heart, and step up to a higher level of rigor, simplicity, and accuracy in theory and modeling.

6.2. Commonality of mechanisms, parameters, and strategies in the models

The aggregate average data and subject-cluster data in all three search task conditions can be fit using the same visual availability and crowding mechanisms with plausible parameter adjustments, and the same mechanisms and parameters for eye and hand movements with only *SlipER* adjustments. Likewise, the models fit both the aggregate and cluster data using the same family of strategies: Basic Search, Fixed-eye, and Basic Search with Confirm-both and Limited-fixations. At the subject cluster level, in each task the same strategies were followed by the cluster models as in the aggregate data model, with one exception: in the Conjunction task, one of the subject clusters was much slower and more accurate than the other clusters; this was captured simply by using the Basic Search strategy in the model for that cluster.

6.3. Theoretical Implications

The basic goal of work with the present EPIC architecture is to determine how much can be understood, explained, and predicted on the basis of models of human cognition and

performance that perform strategy execution with no inherent central bottleneck or other limitations, such as covert visual attention. This a priori theoretical approach forces models of performance to include perceptual/motor characteristics and task strategies as the principal explanatory constructs instead. The approach succeeded extremely well, both quantitatively and qualitatively, in accounting for the observed effects on RT and ER of task type, stimulus properties, and set size, at both at the aggregate and individual subject level.

Implications for covert attention theory. The success of the present models demonstrates that other popular theoretical constructs, such as covert attention shifting and perceptual feature binding (Treisman & Gelated, 1980; Wolfe et al., 1989) — which have been invoked previously to account for data from simple visual search — are probably superfluous and irrelevant to a correct account. Explanations based on these constructs would attribute the effects to these hypothetical limitations without any reference to perceptual/motor or strategy mechanisms. In contrast, the more fundamental and empirically-supported perceptual/motor mechanisms, together with a remarkably simple cognitive strategy mechanism, completely suffices to explain the effects in detail, with no such hypothetical attentional limit being required. It is a question for historical research Why "*attention" has received so much attention* in past visual search research, while more concrete and essential mechanisms were ignored, is a question for historical research.

Alternative crowding mechanisms. Replacing the hypothetical covert-attention concept with more justifiable mechanisms and processes raises other interesting issues. For example, a feature of the present RT fits — as can be seen in results from the Shape task — is that the predicted RTs tend to be more curvilinear than the data, especially on negative trials. This somewhat exaggerated curvilinearity stems from the fact that, for higher set sizes, more objects are covered by a single fixation because of the greater average object density. (The Shape task models do not use the Limited-fixations option, which can also produce negatively accelerated RTs.) An issue thus arises about whether there might be a plausible modification to the model that would reduce the exaggerated curvilinearity in the predicted negative RTs.

A possible way to eliminate this curvilinearity would be for greater crowding to increase the visual availability threshold as well as causing increased scrambling (cf. Section 3.3.1 above). This elaboration of the crowding mechanism was briefly explored as part of the present project, but not pursued extensively for three reasons: (1) additional parameter estimates from the data would be required; (2) the very large confidence intervals around the negative RTs at larger set sizes make the necessary reduction in RT curvilinearity uncertain; (3) the excellent goodness-of-fit obtained with the simple crowding mechanism seems adequate for present purposes. In combination, these reasons suggest that pursuing this approach could amount to overfitting the data.

Implications for models of both RT and ER. The models presented here account precisely for both RT and ER data. They do so in a way that sharply contrasts with conventional stochastic information-processing accounts of speed-accuracy tradeoffs, especially those in which information accumulation is modeled as a discrete random walk or a continuous diffusion process that eventually crosses one or the other of two (*yes* or *no*) decision boundaries, yielding a correct or incorrect response and the corresponding RT (Ratcliff, Smith, & McKoon, 2015). In contrast to the present model, the stochastic processing approach was taken in Wolfe's (2007) Guided Search 4.0 model.

Such models like Guided Search 4.0 suffer from Newell's (1973) criticism in that they lack an explicit control process. Also, even more critically, they produce only the superficial statistics of speed-accuracy tradeoff phenomena, and do not describe *architectural mechanisms* underlying them. That is, they do not provide answers to questions like: Exactly what information is being accumulated? Where is this information stored? What examines it? How is it decided which response to make and when?

In contrast, for the EPIC models presented here, these processes are represented in terms of the computational mechanisms of the architecture. The information being accumulated consists of the visual properties of the display objects that are detected by the early visual system. It accumulates in the visual perceptual store as eye movements are made; the strategy rules examine the information and choose additional eye movements. Ultimately the rules decide when enough information has accumulated to justify a present or absent response.[10]

In some sense, the information processing by the present EPIC models corresponds to a random walk during the course of a trial. However, the characteristics of the search are determined by the interaction of fixed and general architectural mechanisms rather than simply summarized in terms of a basic stochastic process like discrete random walks or continuous diffusion. Thus the explanatory possibilities are far richer. In the EPIC models it is clear how stochasticity of processing may occur; for example the random positions of the objects on the display contribute substantially to the variability in predicted RT for a given search task and set size. By taking advantage of this richer set of explanatory approaches, perhaps explicit cognitive architecture models will provide a more principled approach to jointly accounting for RT and ER in the future.

What determines SlipER? Related to errors and speed-accuracy tradeoffs in the EPIC models, there is another specific issue involving the action slip mechanism and its possible "controllability." We generally believe that a strategy is something that can be controlled; presumably with experience people acquire and refine their procedure for performing a task. In contrast, perceptual-motor parameters would not be under similar voluntary control; they are somehow "built-in" to the perceptual/motor "hardware". However, the *SlipER* parameter seems problematic for this distinction between controllable and fixed components of the architecture.

Specifically, the *SlipER* parameter was remarkably constant over search task conditions in the aggregate data, but was strikingly different at the individual subject cluster level. It was often very small when RTs were longer and more sloped, but often higher when the RTs are nearly or essentially flat. How such differences in *SlipER* might arise is unclear. According to the models, action slips occur at the final response-specification stage of manual movements. It is therefore striking that visual factors associated with large sloped RTs were often accompanied by relatively low slip error rates, and vice-versa, whereas higher slip error rates appeared in the context of visual factors that led to shallow-sloped RTs. Furthermore, in the aggregate data, the False Alarm rate tended to decrease slightly with increasing set size, which suggests that subjects might have been reducing their *SlipER* somewhat when the trial was more difficult. Both of these patterns suggest that action slip error rates might be either under cognitive control to some extent, or

[10] Note that the present EPIC models in fact produce predicted RTs for incorrect responses, and could be modified to contain additional error-producing mechanisms (such as fast guesses), but the Wolfe et al. (2010) data simply do not contain enough error RTs across conditions to support rigorous testing.

reflect more general individual characteristics such as impulsivity, as discussed by Dickman & Meyer (1988).

6.4. Implications for Experimental Methodology

Individual differences in strategy choice and perceptual/motor parameters. As noted above, there is a crucial, frequently-ignored problem in human performance experiments: if the methodology does not clearly and strongly incentivize subjects to approach the task in any particular way, they are likely to devise their own strategies for performing the task. As a result their performance will reflect some combination of their individual architectural parameters and their individually chosen strategies, which may vary haphazardly from one individual to the next. It is unclear how to allocate aspects of individual performance between individual architectural parameters and individual strategy choice. The present explanatory sequence methodology, which gives priority to architectural-parameter-fitting before strategy-fitting, provides such an allocation. Even so, as discussed more below, there are cases where a strategy-fit seems just as good as a parameter-fit. How can we be more confident about what aspects of individual performance are due to architecture parameters versus strategy choices?

There are only two ways to resolve this conundrum: First, measurements of individual subject parameters with the same stimuli in relevant psychophysical tasks could provide a valuable clarification of how visual factors affect visual search performance at the individual level. Yet such attempts have been extremely rare in visual search work, even though they would not be very difficult to accomplish. Second, using performance incentives to stabilize or influence strategy choice would also clarify individual performance, especially for a task like the Conjunction task, where the three subject clusters revealed by the present analysis differed greatly in their apparent preferences for speed and tolerance of errors. Without the needed clarification, multiple questions remain to be answered: Why would some subjects take their time and be extremely accurate, while others in the very same task "bailed out" after only two fixations? The best RT studies use an incentive scheme that sharply penalizes errors while somewhat rewarding speed. Would using this scheme in visual search experiments result in individuals producing performance more similar to each other?

Too often, the issue of individual difference has been swept under the rug. Cavalier researchers have neither controlled for strategy differences nor examined whether individual subjects' performance was consistent with their theories. Often they have failed to even consider different strategies or how different underlying abilities would affect strategy choice. The present results thus not only contribute to understanding how the visual/motor architecture and strategy choice play roles in visual search tasks, but also suggest that future experimental work should attempt to directly measure architectural parameters, adopt incentive methods to stabilize subject strategies, and examine individual subject data for theoretically significant differences.

Eye tracking is the way forward. A final and paramount methodological recommendation also follows from the present research. The EPIC models show that early visual processes and eye movements controlled by task strategies account extremely well for the results from simple visual search, contrary to the earlier belief that these aspects could be ignored. The obvious way to support or refute either of these claims is for future studies of simple visual search to use eye-tracking in addition to RT and ER measurements. Now that eye-tracking has become much less

expensive and much easier to conduct, there is no reason not to use it, given the theoretical stakes at hand

6.5. Implications for Modeling Methodology

Constructing explanatory sequences worked well. Early in the development of these models, the EPIC Philosophy (Section 3.2.1 above) provided valuable general guidance — e.g. no models using a new covert attention mechanism were attempted even when the temptation was strong. However, initially the model development often seemed like a random walk through the space of possible models consistent with this general guidance. The last stages of model development were greatly accelerated when we introduced and systematically applied the approach of constructing explanatory sequences with a rigorous priority order of adding mechanisms.

In retrospect, as summarized above, the development of EPIC models has always followed an explanatory sequence methodology with this same priority order, but did so implicitly and not as systematically. The present paper has presented a systematic and well-defined explanatory sequence process for model development. It seems quite likely that the same methodology would be equally useful with other theoretical approaches. For example, this methodology would help distinguish the roles and contributions of assumptions that are based more on empirical phenomena versus those that are more hypothetical.

Priority of strategy modification versus parameter adjustment. The priority ordering for explanatory sequences might need some modification. There were a few places in the model development where a strategy modification and a parameter adjustment could account more-or-less equally well for the data. These cases include are the above-mentioned ones in the aggregate data for the Shape explanatory sequence, and in fitting the data for the third Shape cluster. In each of these cases, either crowding effects, or the effects due to limiting the number of fixations, could produce Miss errors that increase with set size and faster or negatively accelerated RTs. Thus either a parameter modification to θ and ϕ, or a strategy change with a specific value of *Nfix*, can produce similar predicted results. However, this is not a general result; usually a particular strategy cannot match the data regardless of the parameter values (and vice-versa) — other explanatory sequences reported above make that clear.

The priority ordering imposed by the EPIC Philosophy for modifying a model in an explanatory sequence to improve its fit has strict constraints. There are strong reasons for this. These constraints require first trying perceptual parameter adjustments because they are more grounded in the empirical literature and thus less speculative than modifications to the task strategy. Furthermore, we definitely know that individuals differ in fundamental perceptual parameters such as visual acuity.

Nonetheless, it is also likely that people can devise highly adaptable strategies; it could be argued that subjects' deciding to limit their fixations is a likely adaptation in the Shape task, especially for large set sizes. So assuming they made such a choice might be more plausible than the particular values of θ and ϕ that would produce the same effects.

The problem is that we do not truly know whether particular values of θ or ϕ are plausible. As already mentioned, visual search researchers almost always use different stimuli in every laboratory or experiment, and they very rarely measure the perceptual parameters of their stimuli

in an independent psychophysical task. Thus we have no way of knowing whether the values for θ and ϕ assumed in the present models to produce the observed effects are more, or less, plausible than a strategy modification that produces the same effects. The bottom line is that until more complete parametric data on perceptual mechanisms are available, the stated perceptual-first/strategy-second priority order is the best we can do for now.

7. References

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Abrams, R.A., Meyer, D.E., & Kornblum, S. (1989). Speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system. *Journal of Experimental Psychology: Human Perception and Performance*. 15 (3), 529-543.

Anstis, S.M. (1974). A chart demonstrating variations in acuity with retinal position. *Vision research*, 14, 589-592.

Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In R. W. Spence, & J. T. Spence (Eds.), *Advances in the psychology of learning and motivation, 2*. New York: Academic Press. 90-197.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, 9(12):13, 1–18, http://journalofvision.org/9/12/13/, doi:10.1167/9.12.13.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature,* **226**, 177–78.

Broadbent, D.E.(1958). *Perception and Communication*. London: Pergamon Press.

Buetti, S., Cronin, D.A., Madison, A.M., Wang, Z., & Lleras, A. (2016). Towards a better understanding of parallel Visual Processing human vision: Evidence for exhaustive analysis of visual information. *Journal of Experimental Psychology: General*. 145 (6), 672-707. http://dx.doi.org/10.1037/xge0000163

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Carpenter, R.H.S. (1988). *Movements of the eyes* (2nd ed). London: Pion.

Carrasco, M., & Frieder, K.S. (1996). Cortical magnification neutralizes the eccentricity effect in visual search. *Vision Research*, 37, 63-82.

Chang, H. & Rosenholtz, R. (2016). Search performance is better predicted by tileability than presence of a unique basic feature. *Journal of Vision*, 16(10):13, 1–18, doi:10.1167/16.10.13.

Dickman, S.J., & Meyer, D.E. (1988). Impulsivity and speed-accuracy tradeoffs in information processing. Journal of Personality and Social Psychology, 54, 274-290.

Douven, I. (2017). Abduction. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), URL = <https://plato.stanford.edu/archives/sum2017/entries/abduction/>.

Edwards, W. (1961). Costs and payoffs are instructions. Psychological Review, 68(4), 275-284.

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

Gordon, J., & Abramov, I. (1977). Color vision in the peripheral retina. II. Hue and saturation. *Journal of the Opthalmological Society of America*, 67(2), 202-207.

Halvorson, T. & Hornof, A.J. (2011). A computational model of "active vision" for visual search in human-computer interaction. *Human-Computer Interaction*, 26, 285-314.

Harris, C.M. (1995). Does saccadic undershoot minimize saccadic flight-time? A Monte-Carlo study. *Vision Research*, 35, 691-701.

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.

Hornof, A.J., & Halverson, T. (2003). Cognitive strategies and eye movements for searching hierarchical computer displays. *Proceedings of ACM CHI 2003*, New York: ACM, 107-114.

Hornof, A. J., & Kieras, D. E. (1997). Cognitive modeling reveals menu search is both random and systematic. *Proceedings of ACM CHI 97: Conference on Human Factors in Computing Systems*, New York: ACM, 107-114.

Hornof, A. J., & Kieras, D. E. (1999). Cognitive modeling demonstrates how people use anticipated location knowledge of menu items. *Proceedings of ACM CHI 99: Conference on Human Factors in Computing Systems*, New York: ACM, 410-417.

Hulleman, J., & Olivers, C.N.L.(2017). The impending demise the item in visual search. *Behavior & Brain Sciences*, (40). Cambridge University Press. doi:10.1017/S0140525X15002794, e132

Keshvari, S., & Rosenholtz, R. (2016). Pooling of continuous features provides a unifying account of crowding. *Journal of Vision*, 16(3):39, 1–15, doi:10.1167/16.3.39.

Kieras, D. E. (1981). Knowledge representations in cognitive psychology. In L. Cobb & R. M. Thrall (Eds.), *Mathematical Frontiers of the Social and Policy Sciences, AAAS Selected Symposium 54,* Boulder, CO: Westview Press. 5-36.

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. (2009). Why EPIC was Wrong about Motor Feature Programming. In A. Howes, D. Peebles, R. Cooper (Eds.), *9th International Conference on Cognitive Modeling – ICCM2009*, Manchester, UK.

Kieras, D. (2010). Modeling Visual Search of Displays of Many Objects: The Role of Differential Acuity and Fixation Memory. *The 10th International Conference on Cognitive Modeling – ICCM2010*, August 6-8, 2010, Philadelphia, PA.

Kieras, D.E. (2016). A summary of the EPIC Cognitive Architecture. In S. Chipman (Ed.), *The Oxford Handbook of Cognitive Science*, Volume 1. Oxford University Press. 24 pages. DOI: 10.1093/oxfordhb/9780199842193.013.003

Kieras, D.E. (2018). Visual search without selective attention: A cognitive architecture account. *Topics in Cognitive Science*, ISSN: 1756-8765 online, 1-18. https://rdcu.be/bfiJ7 DOI: 10.1111/tops.12406

Kieras, D.E., & Bovair, S. (1986). The acquisition of procedures from text: A production-system analysis of transfer of training. *Journal of Memory and Language*, 25, 507-524

Kieras, D.E & Hornof, A.J. (2014). Towards accurate and practical predictive models for active-vision-based visual search. In *Proceedings of CHI 2014: Human Factors in Computing Systems. New York: ACM, Inc*.

Kieras, D.E., & Hornof, A. (2017). Cognitive architecture enables comprehensive predictive models of visual search: Commentary on Hulleman & Olivers. *Behavioral & Brain Sciences,* 40, 29-30. doi:10.1017/S0140525X16000121, e142

Kieras, D.E, Hornof, A., & Zhang, Y. (2015). Visual search of displays of many objects: Modeling detailed eye movement effects with improved EPIC. Poster in *Proceedings of the 13th International Conference on Cognitive Modeling (ICCM 2015)*, Groningen, The Netherlands, April 9-11, 2015.

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.E., & Meyer, D.E. (1995). Predicting performance in dual-task tracking and decision making with EPIC computational models. *Proceedings of the First International Symposium on Command and Control Research and Technology*, National Defense University, Washington, D.C., June 19-22. 314-325.

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.

Kieras, D. E., & Meyer, D. E. (2000). The role of cognitive task analysis in the application of predictive models of human performance. In J. M. C. Schraagen, S. E. Chipman, & V. L. Shalin (Eds.), *Cognitive task analysis*. Mahwah, NJ: Lawrence Erlbaum, 2000. 237-260.

Kieras, D., Meyer, D., & Ballas, J. (2001). Towards demystification of direct manipulation: Cognitive modeling charts the gulf of execution. In M. Beaudouin-Lafon & R.J.K. Jacob (Eds.), *Proceedings of the CHI 2001 Conference on Human Factors in Computing Systems*. New York, ACM. Pp. 128 – 135.

Kieras, D. E., Meyer, D. E., Ballas, J. A., & Lauber, E. J. (2000). Modern computational perspectives on executive mental control: Where to from here? In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 681-712). Cambridge, MA: M.I.T. Press.

Kieras, D.E., Meyer, D.E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In A. Miyake and P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control.* New York: Cambridge University Press. 183-223.

Kieras, D.E, Wakefield, G.H., Thompson, E.R., Iyer, N., Simpson, B.D. (2016). Modeling two-channel speech processing with the EPIC cognitive architecture. *Topics in Cognitive Science*, 8, 291–304. DOI: 10.1111/tops.12180.

Klein, R. & Farrell, M. (1989). Search performance without eye movements. *Perception & Psychophysics*, 46(5), 476-482.

Laird, J. E., Newell, A., and Rosenbloom, P.S. (1987) Soar: An architecture for general intelligence. *Artificial Intelligence*, 33, 1-64.

Lawrence, M.A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0. https://CRAN.R-project.org/package=ez

Levi, D.M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, 48, 635-654. doi:10.1016/j.visres.2007.12.009

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K.(2017). cluster: Cluster Analysis Basics and Extensions. R package version 2.0.6.

Martinez-Conde, S., Macknik, S. L., Hubel, D.H. (2004). The role of fixation eye movements in visual perception. *Nature Reviews Neuroscience*, 5, 229-240.

McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. Journal of Experimental Psychology. Human Perception and Performance, 25, 1517–1539.

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

Meyer, D. E., & Kieras, D. E. (1997b). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of Psychology Refractory-Period Phenomena. *Psychological Review*, 104, 749-791

Meyer, D. E., & Kieras, D. E. (1999). Precis to a practical unified theory of cognition and action: Some lessons from computational modeling of human multiple-task performance. In D. Gopher & A. Koriat (Eds.), *Attention and Performance XVII. Cognitive regulation of performance: Integration of theory and application* (pp. 17 -88). Cambridge, MA: M.I.T. Press.

Minsky, M. (1961). Steps towards artificial intelligence. *Proceedings of the Institute of Radio Engineers*, 49, 8-30.

Morris, A., & Horne, E.P. (1960).*Visual Search Techniques: Proceedings of a Symposium Sponsored by the Armed Forces - NRC Committee on Vision. Publication 712.* Washington, D.C.: National Academy of Sciences - National Research Council.

Motter, B.C, & Simoni, D.A. (2008). Changes in the functional visual field during search with and without eye movements. *Vision Research*, 48(22), 2382-2393.

Neisser, U. (1963). Decision-time without reaction-time: Experiments in visual scanning. *The American Journal of Psychology*, 76(3), 376-385.

Newell, A. (1973). You can't play 20 questions with nature and win. In W. G. Chase (Ed.)V*isual Information Processing*, New York: Academic Press, 283-308.

Norman, D.A. (1981). Categorization of action slips. *Psychological Review*, 88(1). 1-15.

van Opstal, A.J, & van Gisbergen, J.A.M. (1989). Scatter in the metrics of saccades and properties of the collicular motor map. *Vision Research*, 29(9), 1183-1196.

Pachella, R.G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition.* Potomac, Md: Erlbaum Associates, 1974.

Põder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision*, 7(2):23, 1–12, doi:10.1167/7.2.23.

Pelli, D.G., Palomares, M., & Majaj, N.J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. Journal of Vision, 4, 1136-1169. doi:https://doi.org/10.1167/4.12.12

Pelli, D.G., & Tillman, K.A. (2008a). The uncrowded window of object recognition. *Nature Neuroscience*, 11(10), 1129-1135. doi:10.1038/nn.2187.

Pelli, D.G., & Tillman, K.A. (2008b). The uncrowded window of object recognition. *Nature Neuroscience Online Supplement*, 11(10), 1129-1135. doi:10.1038/nn.2187

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ratcliff, R., Smith, P.L., & McKoon, G. (2015). Modeling regularities in response time and accuracy data with the diffusion model. *Current Directions in Psychological Science* 24(6): 458–470. doi:10.1177/0963721415596228.

Reitman, W. (1970). What does it take to remember? In D.A. Norman (Ed.), *Models of Human Memory*. New York: Academic Press. 469-509.

Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Review of Vision Science*, 2, 437–57. doi: 10.1146/annurev-vision-082114-035733

Rosenholtz, R., Huang, J., & Ehinger, K.A. (2012). Rethinking the role of top-down attention in vision: effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology*, 3:13. https://doi.org/10.3389/fpsyg.2012.00013

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, 12(4):14, 1–17, http://www.journalofvision.org/content/12/4/14, doi:10.1167/12.4.14.

Rosenholtz, R., Yu, D., & Keshvari, S. (2019). Challenges to pooling models of crowding: Implications for visual mechanisms. *Journal of Vision*, 19(7):15, 1–25, https://doi.org/10.1167/19.7.15.

Sanders, M.S. & McCormick, E.J. (1987). *Human Factors in Engineering and Design*. New York: McGraw-Hill.

Schumacher, E. H., Lauber, E. J., Glass, J. M., Zurbriggen, E. L., Gmeindl, L., Kieras, D. E., & Meyer, D. E. (1999). Concurrent response-selection processes in dual-task performance: Evidence for adaptive executive control of task scheduling. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 791–814.

Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D., Lauber, E. J., Kieras, D. E., & Meyer, D. E. (2001).Virtually perfect time-sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science*, 2001, 12, 101-108.

Scialfa, C.T., & Joffe, K.M. (1998). Response times and eye movements in feature and conjunction search as a function of target eccentricity. *Perception & Psychophysics*, 60(6), 1067-1082.

Smith, S.L. & Thomas, D.W. (1964). Color versus shape coding in information displays. *Journal of Applied Psychology*, 48, 137-146.

Strasburger, H. (2020). Seven myths on crowding and peripheral vision. *i-Perception*, 11(3), 1-46. DOI: 10.1177/2041669520913052

Sternberg, S.(1969). The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), Attention and performance II. *Acta Psychologica*, 30, 276-315.

Sternberg, S. (2016). In defense of high-speed memory scanning. The Quarterly Journal of Experimental Psychology, 69(10), 2020-2075. DOI 10.1080/17470218.2016.1198820

Thompson, E.R., Iyer, N., Simpson, B.D., Wakefield, G.H., Kieras, D.E., & Brungart, D.S. (2015). Enhancing listener strategies using a payoff matrix in speech-on-speech masking experiments. *Journal of the Acoustical Society of America*. 138(3), 1297-1304.

Townsend, J.T., & Wenger, M.J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review*, 11(3), 391-418.

Treisman, A.M. (1969). Strategies and models of selective attention. *Psychological Review*, 76(3), 282-299.

Treisman, A. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology,* 12, 97-136.

Treisman, A. M. and Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology* 14, 107–141.

Triesman, A. M., Sykes, M., & Gelade. G. (1977). Selective attention and stimulus integration. In S. Dornic (Ed.), *Attention and Performance VI.* Potomac, Md: Erlbaurn, 1977. 333-381.

Virsu, V. & Rovamo, J. (1979) Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research,* 37:475–94.

Wertheim, A. H., Hooge, I. T. C., Krikke, K., & Johnson, A. (2006). How important is lateral masking in visual search? *Experimental Brain Research*, 170, 387-402. DOI 10.1007/s00221-005-0221-9

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.

Wolfe, J.M. (2007). Guided Search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems*. (pp. 99 - 119). Oxford University Press.

Wolfe, J. M. (2014). Approaches to Visual Search: Feature Integration Theory and Guided Search. In A.C. Nobre & S. Kastner (Eds), T*he Oxford Handbook of Attention*. Retrieved from DOI: 10.1093/oxfordhb/9780199675111.013.002.

Wolfe, J.M., Cave, K.R., & Franzel, S.L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 419-433.

Wolfe, J.M., Palmer, E.M., Horowitz, T.S. (2010). Reaction time distributions constrain models of visual search. *Vision Research*, 50, 1304-1311.

Wright, C.E., Marino. V.F., Chubb, C., & Mann, D. (2019). A model of the uncertainty effects in choice reaction time that Includes a major contribution from effector selection. *Psychological Review*, 126 (4), 550-577. http://dx.doi.org/10.1037/rev0000146

Yashar, A., Xiuyun, W., Jiageng, C., & Carrasco, M. (2019). Crowding and binding: Not all feature-dimensions behave the same way. *Psychological Science,* September 2019. DOI: 10.1177/0956797619870779

Zelinsky, G., & Sheinberg, D. (1995). Why some search tasks take longer than others: Using eye movements to redefine reaction times. In J.M. Findlay, R. Walker, & R.W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications* (pp. 325-336). North-Holland: Elsevier Science Publishers.

Zelinsky, G.j. & Sheinberg, D.L. (1997). Eye movements during parallel-serial visual search. Journal of Experimental Psychology: Human Perception and Performance 25(1), 244-262.

Technical Report Distribution List

Thomas McKenna
ONR HUMAN & BIOENGINEERED SYSTEMS
875 N. Randolph Street
Arlington VA 22203-1995

Defense Technical Information Center
8725 John J Kingman Road Ste 0944
Fort Belvoir, VA 22060-6218

Naval Research Laboratory
ATTN: CODE 5596
4555 Overlook Avenue SW
Washington, DC 20375-5320