Capture Of Visual Attention By Abrupt Onsets: A Model Of Contingent Orienting

Chris S. Fick (cfick@rice.edu)
Rice University Department of Psychology, 6100 Main Street
Houston, TX 77005 USA

Michael D. Byrne (byrne@rice.edu)
Rice University Department of Psychology, 6100 Main Street
Houston, TX 77005 USA

Abstract

Capture of visual attention has been found to be dependent on the complex interaction of top-down and bottom-up processes. However, the exact nature of this interaction is a subject of current debate. One of the earliest theories of this interaction was the contingent orienting hypothesis, first presented in Folk, Remington and Johnston, 1992 (FRJ). This hypothesis states that attention will only be captured by objects that possess properties for which attention is set. Furthermore, that the demands of the given task determine the properties for which attention is set. FRJ has become a seminal work in the field of visual attention capture, and is often cited without extended discussion concerning the methodology it employed. With the goal of better understanding the processes at work in visual attention capture, an ACT-R/PM model was constructed to simulate experiments 1, 2 and 3 from FRJ. The model produced similar data trends to, and highlighted areas of interest in, FRJ. A comparison of the original and model data are discussed in terms of the original methodology, attentional set and the ACT-R/PM architecture.

Introduction

Understanding how visual attention is captured is crucial in the design of any system with a visual display, from air-traffic control to a word-processing program. Although there has been a vast amount of research conducted in this area there are still uncertainties concerning the conditions in which abrupt stimulus onsets will automatically capture visual attention, (for a summary of the literature see Yantis, 1993). In some situations the appearance of irrelevant onsets have been found to distract subjects, even when instructed to ignore them (Jonides, 1981; Lambert, Spencer & Mohindra, 1987). In other cases subjects are able to ignore onsets (Theeuwes, 1991; Yantis and Jonides, 1990; FRJ).

An explanation for these findings was introduced by Folk, Remington and Johnston, 1992 (which shall be referred to as FRJ). The authors hypothesized that the features of the task determine how the attention of the participant is set, which then mediates if onsets will capture attention. For example, if a target is defined by being a red singleton, then the attention of the subject should be set to discriminate for color, while a non-colored distractor item will not capture attention. This attentional set helps the subject to only attend to the features that may help them optimize task performance. In a search or discrimination task this means keying in on the stimulus properties that define the target.

Subsequent to FRJ, nearly all visual attention capture experiments have discussed their findings in terms of attentional set. The conclusions that have been drawn do not always agree, but the consensus is that the relationship between targets and distractors must be examined. The influential nature of these conclusions within the literature stimulated our decision to model the FRJ experiments.

In FRJ, the experimenters set out to manipulate the features of the cues and targets in a spatial cueing task in order to examine the effects of attentional set. The procedure of Folk, et al 1992 experiment 1, consisted of four abruptly onsetting and offsetting circles, which were presented as a cue around one of four possible target locations (see Figure 1: A2). Subjects had to discriminate whether the target was an = sign or an X, and respond with the appropriate key press. For one group of subjects the target was a single abruptly onset character (Fig 1: A4), but for the other group it was one of four figures, differentiated by color (Fig 1: B4). The authors hypothesized that the features of the target would control the attentional set of the subject, and that the validity of the onset cue would affect response times only in trials where the target was specified by onset. The results were compatible with this hypothesis. Valid cues reduced reaction times for both target types, but invalid cues cost time only in the onset target condition. In a second experiment the two features were reversed, with color acting as a cue (Fig 1: B2). The goal was to see if any feature that the subject was monitoring for would involuntarily attract attention. The results supported this notion. The invalid and valid color cues had the expected cost and benefit effects when the target was distinguished by color and had no effect when the target was a single onset figure. The authors concluded that once the attentional set of the subject had been determined, any stimulus with one of these properties, e.g. color, would automatically capture attention. On the other hand, any stimuli that did not have the properties for which attention was set would fail to capture attention. This was dubbed the contingent involuntary orienting hypothesis.

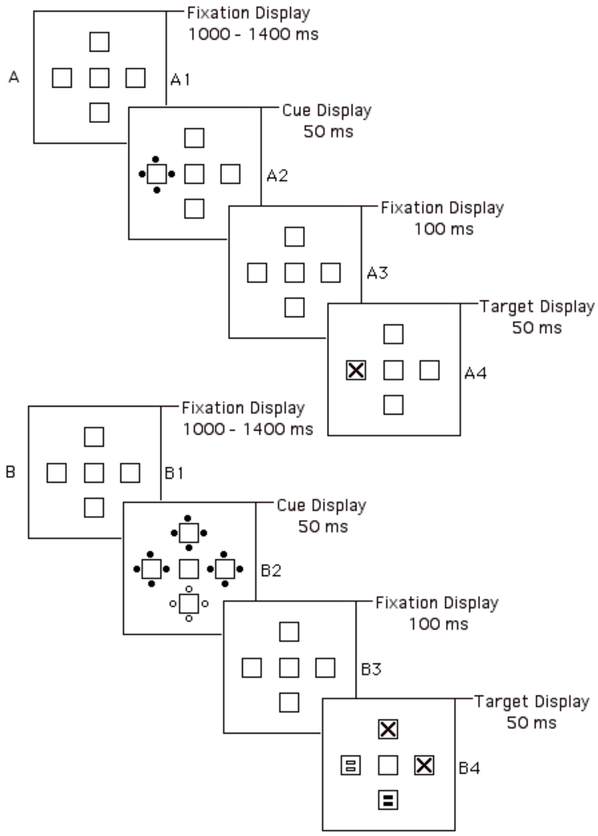

Figure 1. Displays and sequence of events for FRJ Ex. 1 and Ex. 2. (A) Experiment 1, onset cues, is an example of a valid-cue, onset target trial. (B) Experiment 2, color cues, is an example of an invalid-cue, color target trial.

It is important to note the extreme efforts that were used to develop specific attentional sets. The trials were blocked not only by cue and onset type, but also by validity of the cue. In other words, for a given block of trials the cues were either 100% predictive or 100% non-predictive of the target location. In addition, the subjects were informed of this blocking. These factors are relevant to comparisons across experiments that are not designed with such a strong focus on instilling a given attentional set. They also make certain findings from this experiment extremely surprising.

Modeling the Experiment

We used ACT-R/PM (Byrne & Anderson, 1998) to model the experiment. Because the cognitive demands of the search task are minimal, modeling the perceptual-motor processes (e.g., shifting visual attention, discriminating the target, and responding with a keypress) with some fidelity is critical. The ACT-R/PM architecture combines ACT-R's theory of cognition (Anderson & Lebiére, 1998) with modal theories of visual attention (Anderson, Matessa, & Lebiére, 1997) and motor movement (Kieras & Meyer, 1997). ACT-R/PM explicitly specifies timing information for all three processes as well as parallelism between them. It has also been used successfully to model a variety of behavioral phenomena, including visual search tasks. A recent example is Fleetwood and Byrne, 2002, which modeled an icon search task. In addition, there are other features built into the vision module that facilitate the programming of a model for search and spatial cueing tasks. Specifically, the vision system gives special status to objects that have recently onset, such as the cues and targets in this paradigm. Also, the system has the ability to discriminate via color. These "pop-out" features are based on the ways in which the human visual system is pre-attentively sensitive. For a more in depth discussion of ACT-R/PM's vision module, see Byrne, 2001.

The model was based on a rough version of FRJ's contingent orienting hypothesis, ie. the behavior of the model depended on the relationship between the cue and target type. The model began each trial by focusing visual attention on the center box of the display, as subjects were instructed to do. Also, since subjects were informed of which block of trials, cue-type, target-type and cue validity they were about to begin, the model was also provided with this information.

The features of the current trial type determined how the model responded when the cue was presented. When the cue was valid, and shared properties with the target, the model shifted visual attention to the location of the cue. When the cue was dissimilar to the target, the model kept attention fixated on the center of the screen. In the no-cue condition the model picked one of the locations to move to at random, which simulated a guess at the up-coming location of the target.

When an onset target appeared (experiment 1) knowledge of its location was immediately available to the model. If visual attention was already at this location, the model read the target character and produced the appropriate keystroke response. If the target was at a different location, the model shifted attention to the target and then read it and responded. This additional shift of visual attention was responsible for the greater time costs for attending to invalid cues. When a target defined by color appeared (experiment

2), the model checked to see if the object at the same location as visual attention was red. If so, the model read the target and responded with a simulated key press; if not, it located the red object, shifted attention to that location and then read the target and responded. Requiring the model to verify that it was really looking at a red object, or to search for a red object, yielded greater time costs for color targets than onset targets. The way in which attentional set was controlled is the most abstracted piece, but simulates the subjects' knowledge of what block of trials she was engaged in.

As mentioned, the model was faster to respond to onset targets than those defined by color. For human subjects this cost was likely due to the increased complexity of the visual scene for color targets, since the color target was presented along with 3 foils. In comparison, in the onset target condition the target was the only object to appear on the screen, eliminating the possibility of any distraction from a foil. The model showed the same pattern, and for similar reasons. When a new object appears on the screen its location is automatically stored in ACT-R's visual location buffer. Attention is not automatically drawn to that location, but it can be rapidly shifted to. When more than one object appears simultaneously ACT-R automatically chooses one object and stores its location. If this object did not happen to be the color target than the location of the color target had to be searched for before a shift of attention could be made. Therefore this automatic buffer stuffing yields an advantage for locating a single onset. Importantly, an item of a given color is located without fixating on each possible item, simulating the pop-out, or pre-attentive nature of search for a color singleton.

Results

The model data were produced by running through blocks of 60 trials for each set type, as in FRJ. The model showed similar trends across all of the individual cue-type, target-type and cue validity conditions. The quantitative measures for goodness of fit for each cue type are r-squared, RMS deviation, and percent average absolute error. These results are shown in table 1. The color-cue data was a much closer fit than the onset-cue data, though both had fairly low percentages of average absolute error.

Table 1. Quantitative analysis of model data as compared to FRJ, data.

	Experiment 1	Experiment 2
	Onset Cue	Color Cue
R-squared	0.605	0.884
RMSD	39.3 ms	29.39 ms
% Average Absolute Error	6.54%	3.00%

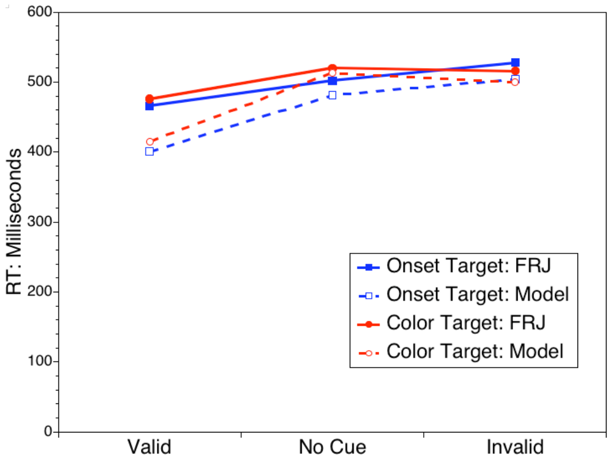

Figure 2. Mean response times for Experiment 1, onset-cues, for original and model data.

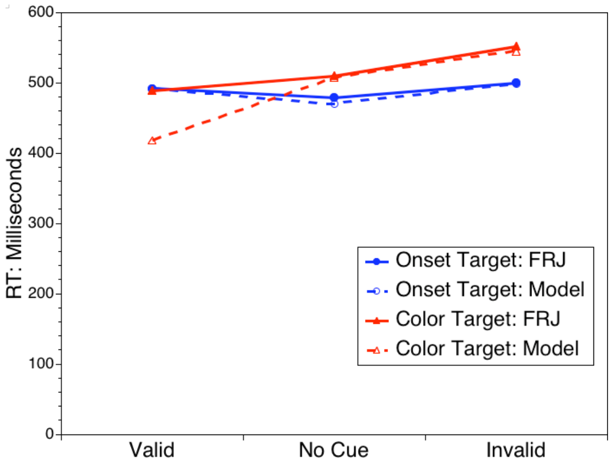

Figure 3. Mean response times for Experiment 2, color-cues, for original and model data.

Discussion

The model captured all of the main effects present in the original data. When the cue and target were onsets, there were costs for invalid cues and gains for valid cues, whereas there were no costs for invalid onset cues when the target was distinguished by color. Likewise, there were costs and gains for color cues followed by color targets, but no difference between valid and invalid color cues when the target was an onset. These costs and gains were based on whether the model had shifted visual attention to the cued location. Therefore the model followed the same patterns as the subjects in shifting attention or not, based on the features for which attention was set.

The model was systematically faster than the FRJ subjects when responding to targets in trials with valid cues. This difference can be seen in the valid conditions of experiment 1 but is most noticeable in the valid, color-target condition of experiment 2. One factor that may be partially driving this effect is the speed at which the model is able to determine that it is really looking at the target. In addition, recent eye-tracking data, such as Kramer, Cassavaugh, Irwin, Peterson and Hahn, 2001, have shown that even when onset distractors capture attention, saccades are only made towards the onsets on less than half the trials. Although the onsets in Kramer, et al, 2001 were not predictive of the target location, the principle finding is still relevant. This implies that subjects will occasionally fail to shift attention to a cue, even when they knew it to be valid. The model however never failed to shift attention when it was supposed to. This helps to explain why the FRJ data means are slower than the model data means for valid, target similar cues. This issue may also explain why the costs and benefits for valid and invalid cues were greater in the model than in the FRJ data.

With this in mind the model would be better suited to match human performance if it occasionally failed to shift attention to a valid cued location, and likewise occasionally did shift to invalidly cued locations. This might be accomplished by weighting the productions such that the attentional set was not 100% predictive of the decision of whether or not to shift attention to the cued location.

Although this particular model may have room for improvement, it still points out certain areas of interest in FRJ. For example, it is important to take into consideration the blocking of trials in FRJ. The purpose of the experiment was to examine the effects of different attentional sets, and blocking the trials by cue-type, target-type and validity was an effective way to instill a given attentional set. Surprisingly, even though subjects knew that a given cue was going to be predictive or non-predictive of target location, the match between cue and target type was still a powerful factor. For example, even though subjects knew that the cues were valid, they failed to take advantage of a valid color cue to attend to the location of a subsequent onset target. Subjects were so keyed to attend to onsets that valid color cues were ignored. However, subjects were able to use valid onset cues in reducing response times for color targets. This is yet another example of the uniqueness of onset above other stimulus properties.

This distinction is important to the definition of attentional set. If attentional set is formed based on the goal of optimizing task performance, than subjects in the valid color-cue, onset-target condition, should have been able to form an attentional set for both color and onset. However, rather than setting attention to the properties that would optimize task performance on the whole, subjects' attention seemed to be set for optimizing only the target identification segment. This fact implies either that the subjects were not highly motivated to optimize performance, or that the formation of such a complicated attentional set is difficult or impossible. Therefore, in addition to the relationship between cues and targets, attentional set may need to be examined in terms of task structure. For example, in this task the attentional set formed in the target discrimination segment of the task effected which features captured attention in the cueing segment. Also, this effect was different for color and onset cues.

Another point is that response times for no-cue trials in FRJ were noticeably greater for both target types in experiment 1 than in experiment 2, although there is clearly no cue-type distinction for no-cue trials between experiments. Since the model did not differ for no-cues of a given target type, the model data happened to fall closer to the no-cue data in FRJ, experiment 2. This is a main factor for the poorer fit between the model and experiment 1. Such differences are inherently difficult to model, since they come from identical experimental trials.

Overall, ACT-R/PM proved a viable system for modeling this visual attention capture experiment. Even though the vision module was not designed specifically for this paradigm, no major changes had to be made in order for ACT-R/PM to model this task. In fact none of the free parameters that can be adjusted in ACT-R/PM were changed. Effectively the model was a basic version of FRJ's contingent orienting hypothesis applied to an "out-of-the-box" ACT-R/PM framework. Maintaining an unadjusted architecture was also a main reason why we did not apply any of ACT-R/PM's learning functions. Although doing so may have made the model data more representative of human data, learning introduces too many free parameters to justify it's inclusion in this model. Due to the purity of the architecture employed, similarities between the model and human data provided support for both the time measures in ACT-R/PM, and the application of the contingent orienting hypothesis to basic spatial cuing tasks. This speaks well for the generalizablity of the ACT-R/PM system in modeling experiments involving attentional set and spatial cueing.

Experiment 3

In order to examine the effects of blocking trials by validity as well as cue and target features, the authors of FRJ manipulated the validity of the cues in experiment 3. The paradigm was changed so that a cue appeared on every trial, yet it was uncorrelated with the upcoming target location. With four possible locations the cue was predictive on 25% of the trials, negating any benefit of shifting attention to the cue. Any response time differences between predictive and non-predictive cues, therefore, therefore indicate that subjects were unable to withhold an attentional response to the cue. The findings strongly supported the contingent orienting hypothesis. When the cue shared the property of the target, the invalid cues yielded longer response times. Whereas, when the cue and target were dissimilar, there was almost no effect of cue validity. This was expected since the cues were non-predictive in both cases, making target type the only goal-relevant property. However, there were no differences between cue-types for valid cues in both target conditions. This means that valid cues sharing target properties yielded no time gain compared to dissimilar cues. This pattern is different from those in experiments 1 and 2, and suggests that the benefit of the target-similar cues was mediated by the non-predictive nature of the cues.

Modeling Experiment 3

The model that was developed to run FRJ experiments 1 and 2 was adapted to simulate FRJ experiment 3 in two ways. First, the new model relied solely on the match between target and cue type, and not on validity, to determine the cue properties that would trigger an attention shift. Second, for valid, target-similar cues, the model double-checked that it was really at the target location, adding a slight delay in target identification. This delay simulated the reduced benefit for valid cueing due to the non-predictive cue-target relationship.

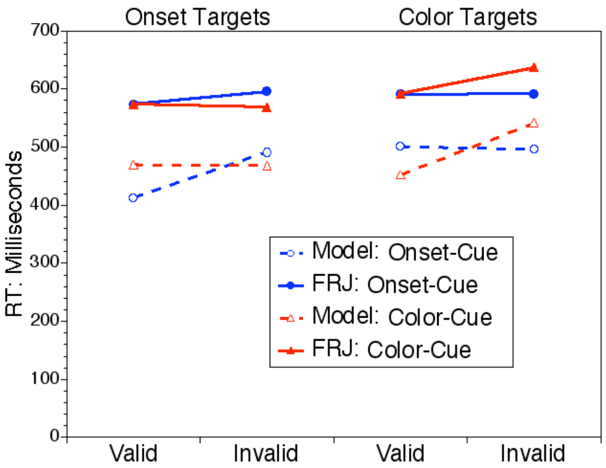

Figure 4. Mean response times for Experiment 3, FRJ and model data.

Results

The main trends in the FRJ data were reproduced in the model data. When the target and cue were dissimilar the validity of the cue had no effect on response times. When the cue and target were similar there were costs for invalid and gains for valid cues. Also, as in the FRJ data, there was a greater difference due to cue validity in the color-cue, color-target condition. The quantitative measures for goodness of fit are shown below in table 2.

Table 2. Quantitative analysis of model data as compared to FRJ, data.

R-Squared	0.608
RMSD	114.14 ms
% Average Absolute Error	18.90%

Discussion

The similarity between the trends in the model and FRJ data suggest that the general effects of attentional set are being simulated. However, the differences highlight three interesting points, which are not discussed in FRJ. Overall, the model data for all three experiments was faster than the subject data. However, this difference was magnified since response times for FRJ experiment 3 are greater than from the first two experiments. Since the only change to the paradigm in FRJ experiment 3 was the blocking of trials by validity, the lack of predictable cueing appears to somehow be causing longer response times.

It is also probable that the non-predictive nature of the cues weakened their power in capturing attention, even when they shared properties with the target. Since validity was shown to have an effect in how attention was set in experiments 1 and 2, it is fair to assume that it had an effect in experiment 3. This means that target-similar cues in experiment 3 elicited shifts of attention on fewer trials than in the valid-cue conditions of experiment's 1 and 2. Such a trend would yield data with a shallower slope between the valid and invalid cues. This trend can be seen when comparing the model and FRJ data. The model for experiment 3 did not lessen the percentage of cues that elicited attention shifts. Therefore the model data yielded a greater difference between target-similar valid and invalid cues.

The final difference between the model and FRJ data is due to the relative response times between valid, target-similar cues and target-dissimilar cues. The lack of difference between valid and invalid target-dissimilar cues implies that they are not eliciting attention shifts. Whereas the difference between valid and invalid target-similar cues implies that they are eliciting attention shifts. With this in mind, one would expect there to be gains for valid, target-similar cues. This is the case in the model data, but is not represented in the FRJ data, where valid, target-similar cues yield response times equal to target-dissimilar cues. Some factor must be driving up the response times of trials with a valid, task-similar cue. This effect was addressed in the model with the addition of the double-take production, which simulated a moment of surprise, which may occur when the cue is actually predictive of the target. Since this occurs on only 25% of the trials it may cause the subject to re-verify that they are indeed focused on the target. However, while the addition of the double-take production did raise response times in relevant conditions, the delay was not great enough to match the FRJ data.

The results from the model of FRJ experiment 3 highlight points in the subject data that are not explained solely by the contingent orienting hypothesis. In particular, they suggest that in addition to attentional set created by the properties of the target, predictable validity also determines the capturing power of a cue. Shifts of visual attention are an all or none phenomenon, so we would expect that a cue either captured attention or did not. Therefore reduced cue validity should be reducing the number of trials in which the cue captures attention.

One possibility for future research would be to replicate FRJ with modern eye-tracking equipment. This would provide insight into how often saccades were really being made to the cued locations in each experimental condition, and how response times were subsequently effected. Such data would help in the future development of cognitive and perceptual models such as ACT-R/PM.

The model may also be altered to simulate other visual search and spatial cueing tasks. The adaptations required and subsequent comparison to the experimental data may provide insight into why these two paradigm types often yield different findings. This could help produce a clearer picture concerning the interplay between top-down and bottom-up attention capture.

References

Anderson, J. R., and Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation of visual attention. *Human-Computer Interaction*, 12, 439-462.

Byrne, M. D. (2001). ACT-R/PM and menu selection: applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies.*

Byrne, M.D. & Anderson, J.R. (1998). Perception and action. In J.R. Anderson & C. Lebiere (Eds.), *The Atomic Components of Thought*, 167-200. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fleetwood, M. D. & Byrne, M. D. (2002) Modeling icon search in ACT-R/PM. *Cognitive Systems Research*, *3*, 25-33

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1030-1044.

Jonides, J. (1981). Voluntary verses automatic control over the mind’s eye’s movement.In J. B. Long & A. D. Baddeley (Eds.), *Attention & performance IX* (187-203). Hillsdale, NJ: Erlbaum.

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

Kramer, A., Cassavaugh, N., Irwin, E., Peterson, M., & Hahn, S. (2001). Influence of single and multiple onset distractors on visual search for singleton targets. Perception and Psychophysics. 6, 952-968.

Lambert, A., Spencer, E., & Mohindra, N. (1987). Automaticity and the capture of attention by aperipheral display change*. Current Psychological* Research and Reviews, 6, 136-147.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. Perception & Psychophysics, 49, 83-90.

Yantis, S. (1993). Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology: Human Perception and* Performance, 19, 676-681.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary verses automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 121-134.