In D. Reitter & F. E. Ritter (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling
(ICCM 2016). University Park, PA: Penn State.

A minimal model of eye movement applied to visual search and change detection

Shane T. Mueller(shanem@mtu.edu) **Yin Yin Tan (yinyint@mtu.edu)**

Hannah North (hbossele@mtu.edu) **Kelly Steelman (steelman@mtu.edu)**
Department of Cognitive and Learning Sciences
Michigan Technological University
Houghton, MI 49931 USA

Abstract

We describe the Eye Movement Minimal Model-Modified (EM4), a lightweight minimally-sufficient model of eye movements that accounts for visual search times in several distinct paradigms. The model allows visual search to be guided by probe-item similarity in different foveal zones, which enables the model to be used as a front-end for various models of visual saliency. We apply the model to four distinct paradigms to demonstrate its flexibility and utility.

Keywords: eye movements; visual search; change detection

Background

In recent years, detailed models of visual processing that represent or are inspired by the human visual system have proliferated, providing many alternate computational approaches to investigating properties of visual attention, saliency, image analysis, and the like (Bruce & Tsotsos, 2009; Itti, Koch, & Niebur, 1998; Wolfe, Cave, & Franzel, 1989). Currently, more than 60 distinct methods of evaluating visual saliency have been compared using on popular benchmark (Bylinskii et al., 2016). In contrast, relatively less attention has been paid to modeling the mechanisms and strategies involved in directing visual attention via eye movements to perform visual search. Although some models of visual saliency have included foveated eye movements (e.g. Itti et al., 1998), models of visual saliency that ignore foveation and eye movement may make either unnecessary or unrealistic assumptions. Hornoff & Halverson (2003; 2004a, 2004b, 2007, 2011), developed and enhanced models of visual search via foveated eye movements using the EPIC computational architecture (Kieras & Meyer, 1997). As part of this effort, they described a "Minimal Model" involving the assumptions they felt most necessary and sufficient for modeling visual search in applied settings. Subsequently, more advances have been made to these models, both at the architectural and strategic level (Kieras, 2011; Kieras, Hornof, & Zhang, 2015), and these developments have been mirrored by a series of models using the ACT-R architecture (e.g. Salvucci, 2001; Nyamsuren & Taatgen, 2013; Choi, Han, Oh, & Myung, 2015). Yet the minimal model is an attractive target for practical simulation modeling outside the context of a cognitive architecture. Its notions have been adopted by several applied models of visual attention (e.g. Teo & John, 2008), but the model was not designed to handle visual search based on saliency and similarity cues, and so its lessons have not been widely as adopted in the broader field of computational vision that has otherwise led to dozens of visual and image-processing models that identify saliency.

In this paper, we describe the Eye Movement Minimal Model-Modified (EM4), which takes the Halverson & Hornoff model as a starting point, implements it as a standalone simulation model. To handle search in more general situations, the model incorporates search based on probe-item similarity, a quantity akin to what many visual saliency models produce naturally as an output, either via an activation or posterior probability distribution. After describing the model, we will show its ability to capture data in several related visual search and flicker change detection paradigms, illustrating its flexibility and utility.

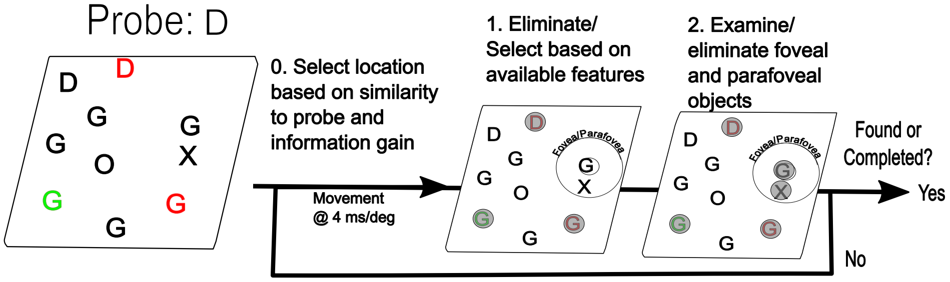
The Eye Movement Minimal Model-Modified

The EM4 is intended as a simple implementation and extension of the core assumptions of the "Minimal Model" proposed by (Halverson & Hornof, 2007), with the goal of accounting for major phenomena in visual search paradigms. The EM4 is implemented as a standalone software routine in the statistical computing language R, so that it can be repurposed and adapted to work with other models of visual processing or human performance, and serve as a lightweight modeling and teaching tool. The source code for the model is available via `https://github.com/stmueller/em4`. The basic stages of the model are shown in Figure . The model operates by simulating the timing of a series of eye movements and other decisions that produce a response in the task.

Primary assumptions

We assume that visual search involves a repeated set of stages in which a target object is selected based on its similarity to the probe and its potential for information gain, following which objects are eliminated or selected based on their similarity to the probe. This repeats (fixation target is selected and foveated, items are eliminated or selected based on similarity) until either the probe object is found or a decision is made to stop search. These probe-item similarity values are *inputs* to the model and we treat them as free parameters. This is a departure from the original description of the minimal model, which used identity-match, and subsequent EPIC visual search models, which have also used feature-level descriptions to represent how different types of information are available at difference eccentricities. Use of similarity provides a useful mid-level representation, such that low-level feature-based visual processing models could produce similarity as an output, perhaps without even requiring those models to be directly embedded within the simulation. The basic

Figure 1: Schematic stages of model. Once fixated (0), target locations are eliminated and selected (1) based on similarity-to-probe. Here, color mismatches may be quickly eliminated in the periphery. Next, (2) targets in the fovea and parafovea are examined, and neighboring targets not eliminated are examined. Once no more targets appear in the parafovea, a new location is selected based on available similarity-to-probe and information gain. The process is repeated until search is complete.



assumptions of the model include:

Information is represented as probe-to-item similarity. The main paradigms investigated by Halverson & Hornoff involved locating a text-based menu item that was always present. Yet even Latin characters have a well-described similarity space (Mueller & Weidemann, 2012), such that similar characters are confusable and take longer to discriminate. In more traditional visual search tasks, targets that are distinguished by a single feature can produce visual 'pop-out', such that the number of distractors does not impact search time. Furthermore, some items in the periphery might be selected or ignored based on similarity to the probe–if the search target is an "O", any "X" in the periphery might be ignored, but a "U" might require more investigation. Consequently, the EM4 represents this information as a similarity score, whose values may differ foveally, parafoveally, and peripherally.

Zones of detection. In reality, the availability of color, shape, size, and location of objects degrade differentially and smoothly as an object's eccentricity increases, with serious degradation starting to occur 30°–45°from the fovea (Boff & Lincoln, 1988). Many recent models have used an eccentricity function (Kieras, 2010; Nyamsuren & Taatgen, 2013), which parametrically defines the availability of different features at different eccentricities, but the EM4 retains just three visual zones: the fovea with radius 1°; the parafovea with typical radius 3.5°; and the periphery which involves the remaining visual field. We assume that the location of visual objects is available everywhere, insofar as any object can be selected as an eye movement destination. Foveated targets can be identified explicitly (with some chance of error), whereas parafoveal targets with high probe-item similarity are more likely than those with low similarity to be selected for subsequent eye movements. In practice, because relevant objects can be detected or rejected parafoveally, the size of the parafovea maps roughly onto the useful field of view (UFOV; Edwards et al., 2006), and the size may depend on properties of the task. In addition, just as peripheral objects that are high in probe-item similarity might direct subsequent eye movements to that location, those high in probe-item *dissimilarity* can be used to eliminate targets from consideration and thus make fast rejection responses (see Chun & Wolfe, 1996). In the fovea, the similarity represents the *probability* that a particular object is identified as the target. For true targets, this maps onto the probability of misdetection used by Halverson & Hornoff, but also permits false alarms if the value is nonzero for foils.

Movement and decision times. We assume that after each fixation, a decision is made about whether the searched-for target has been identified, following which a choice is made about the next eye movement destination. The timing of this decision-action cycle constitutes one of the main free parameters of the model, which we assume is impacted by the nature of the stimuli, as well as dynamic aspects of the environment. Although the original EPIC models attempted to use fixed architectural parameters to determine this timing, different data sets require using some very different decision timing. In addition, this time incorporates all time that is constant with each foveation. A saccadic eye movement is assumed to take place at 4 ms/degree of visual angle.

Accepting and rejecting matches. In general, the existence of visual pop-out is taken as evidence that some decisions can be made based on information in the visual periphery. In the tasks described here, detection of high-similarity targets outside of the fovea lead to a subsequent eye movement to the target to confirm and localize the target (partly because most of the tasks we examine require a selection of the target via mouse movement). However, just as a target can be identified in the periphery or parafovea, we also assume that targets can be rejected from consideration based on information in the periphery or parafovea. In the parafovea and periphery, the similarity score represents an activation level, such that values above 0.5 represent greater similarity to a probe, indicating a possible match; values below 0.5 indicate dissimilarity to a probe great enough eliminate from search. Targets with periphery similarity greater than 0.5 each need to be examined and either eliminated or responded to if found to be identical to the target. If all peripheral targets with similarity above 0.5 are examined and eliminated, a negative response can be made. However,for peripheral targets, we assume that target-probe values below 0.5 permit eliminating the target without eye movement (i.e., preattentively), allowing for fast responses. This accounts for findings such as the ability to make a probe-absent decision without examining each target, or (when a probe should produce pop-out) to make a target-absent decision quickly even when nothing is detected (Chun & Wolfe, 1996). Importantly, only probe-item similarity is used directly, and the model is not impacted by target-distractor similarity, which may provide additional gestalt cues for helping to identify and classify oddball search targets or possibly make search less efficient.

Selecting subsequent locations for search. Deciding where to search next (including in cognitive search of memory, physical search of environments, and other domains) involves cost-benefit analysis (Perelman & Mueller, 2015), because the costs of moving must be weighed against the potential gain in information (Drury, 1975; Bruce & Tsotsos, 2009). For search constrained by eye movements, the time needed to move the eye to a new destination is relatively insensitive to the distance moved (only 4 ms/degree), in contrast to the fixed cost of 100 ms or more required to program and execute the movement, and the time required to classify an item once it is foveated. However, deliberate short eye movements help avoid repeated search of a location by making the task of keeping track simpler; this may improve time-to-find, even if a more distant location could offer maximum gain in information, so that a new location with more potential targets may be better than a closer location with only one target.

The present model balances these by first looking for high-similarity unidentified targets in the parafovea; if this fails, it computes a neighborhood activation score for each unvisited target (the sum of the exponentially-discounted similarity of all unvisited nearby targets), and deciding the next eye movement based on a mixture of the normalized inverse neighborhood similarity scores and noisy distance-to-target, so that the next target may (at one extreme) be the next-most-similar, or (at the other extreme), be based purely on the distance selection scheme proposed by Hornoff & Halverson. When target decisions are made based on discounted neighborhood activation, this favors movements to targets in dense regions where a single fixation is able to eliminate several objects.

In summary, the model implements a stand-alone version of the minimal model that operates by repeatedly selecting objects, fixating on them, and eliminating them from contention, until the selected target is found or all targets are eliminated. Next, we will examine how the model fits several related visual search paradigms.

Model Fits to Data

In this section, we will describe the model's fit to several empirical data sets. These include a menu search task, a visual search task, and two flicker-based change detection tasks. The parameter values and goodness-of-fit values (both R and percent deviation, where appropriate) are shown in Table 1. We will examine a menu search task, a feature-search task; and two flicker-change blindness tasks whose performance profiles differ substantially.

Figure 2: Tasks modeled in this paper. A. Menu selection task; B. feature search; C. Dot-flicker change detection; D. Sparse change detection.

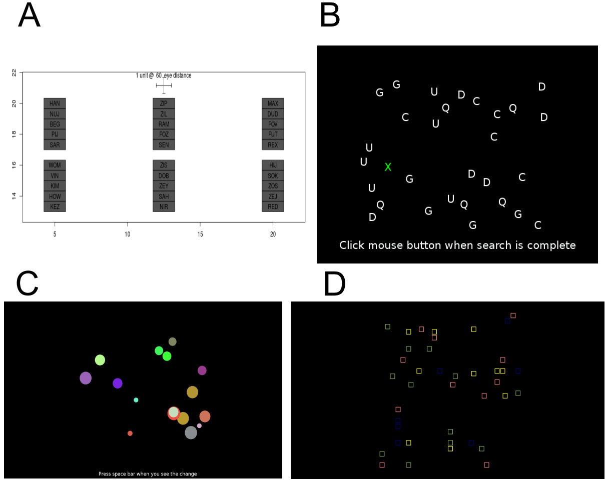
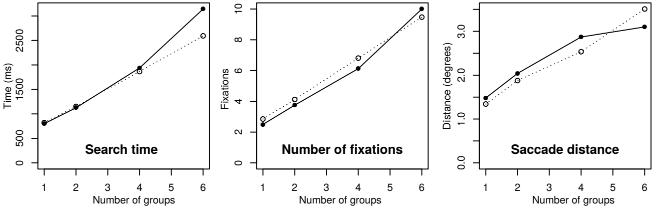


Figure 3: Model and data from menu search task. Left panel shows search time (in ms); center panel shows mean number of fixations; right panel shows mean distance of saccades.



Menu Search Task

The primary task used by Halverson & Hornoff to inform recommendations for a minimal model involved menu search, in which blocks of contiguous text-labeled targets needed to be searched to find a specific target (see Figure 1a). In this task, aside from target location, no information in the parafovea or periphery is useful for localizing the target menu, as the labels were 3-letter strings that could not be easily identified without foveating on or near the menu.

Method This task involved search conditions involving 1, 2, 4, or 6 blocks of menu items, where each block consisted of five items spaced vertically at .66°, arranged in up to three columns, two blocks per column, with a vertical separation of 1.33°and horizontal separation of 7.5°between blocks. Each target had a unique 3-letter label, and on each trial, a participant searched for a specific labeled target. Three critical dependent measures were examined: mean time to find, mean number of fixations, and mean saccade distance. Full parameters are shown in Table 1. Model fits are shown in Figure 3.

Discussion This simulation produced good fits, using parameters and assumptions similar to Halverson and Hornof (2007), the one major exception is the noise parameter used to select subsequent target locations, which was much larger for the present model, primarily because in the present model, eye movements are first made to nearby locations in the parafovea, rather than solely on a noisy-distance scheme. This gives the model a natural preference for nearby objects, and so to counteract this, a larger default noise parameter was required. This illustrates that the EM4 captures the major phonemonon or which the original minimal model was designed to account for.

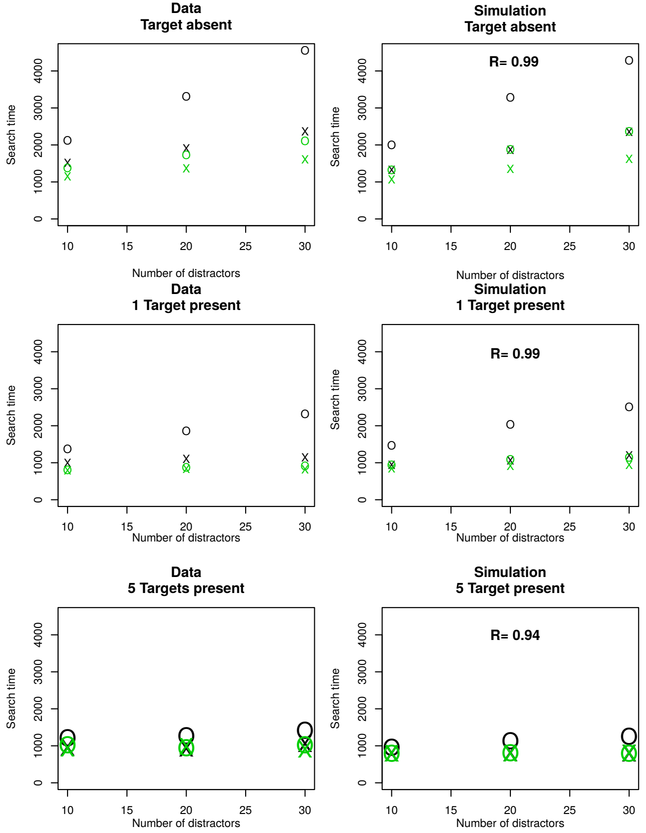
Visual search of simple targets with pop-out

Although the previous task is a useful starting point, it differs from the most commonly-used visual search paradigms typically used within vision science and psychology. Such search tasks typically differ in three ways from this menu search: (1) they involve haphazard stimulus arrangement, making systematic search more difficult; (2) they often involve search for a specific target character amongst a field of distractors that may be either similar or dissimilar to the target (i.e., searching for a T in a field of Ls or Os); and (3) they are often used to demonstrate visual pop-out or feature search, the finding that the presence or absence of some features can be detected peripherally. Thus, the next step in developing the model was to examine how it can account for a more traditional visual search task.

Method This study involves an implementation of a visual search task with several search targets producing visual popout (Mueller & Piper, 2014, see Figure 1b), in a cross-national study (Tan, 2016) that involved 136 participants. In this task, participants searched for a specified target on each trial (a white or green O or X) in a field of either 10, 20, or 30 round white characters (C, D, G, Q, and U) on a black background that was approximately 15°x 10°of visual angle. On different trials, 0, 1, or five targets were present. These conditions parametrically varied the efficiency of search, the number of distractors, and the number of targets, and provided a systematic data set for modeling search times. On each trial, the field of elements was presented until the participant clicked the mouse button; after which the elements were each replaced by a circle, and the participant was instructed to either indicate the location of an object matching the probe, or a label marked "none" if no objects matched the probe.

Results Accuracy for the task was high (98.5%) and so we will consider only search times, which are shown in Figure 4. The human results (left column) show that response time for rejecting pop-out targets tended to become longer with larger search sets, with a clear ordering from most difficult to least of white O, white X, green O, and green X. The same ordering occurred regardless of whether a target was present or absent, but the times were faster (and the slope with respect to number of distractors was smaller) when a target was present, and this diminished further when multiple targets were present.

Figure 4: Model fits to the visual search task.



The fits to data were good (see Table 1), with several differences in parameter values from the first model being (1) smaller detection/rejection time parameters were used, and (2) eye movements locations were selected based on probe-item similarity, rather than by distance alone; and (3) failure-to-detect was reduced to 0.0. The smaller detection times are reasonable because the current task required detecting a single letter instead of a 3-letter sequence. The use of probe-item similarity (or a similar concept) is necessary to fit these data, and this required making assumptions about probe-item similarity for each target class (green and white Xs, Os, and D/G/U/Q/C) in each zone. The slope of the response time to each target class (with respect to number of distractors) is primarily controlled by the peripheral probe-item similarity. These similarity values were assigned with a uniform distribution having a range of 0.3 units, with the minimum values of .45 (when color and shape match), .3 (when either color or shape match) and .25 (when color and shape mismatch). Thus, the small but positive slope for target-absent responses arises because as the number of distractors increase, the number of distractors with a probe-target similarity above 0.5 increases, requiring additional eye movements to eliminate. Together, these assumptions accounts for search times with a mean proportional absolute deviation of 0.16 and a correlation of .94.

Discussion The visual search data shows that the EM4 can provide a credible account of a more traditional search task, including pop-out effects, target-absent effects, and effects of the number of distractors. Thus, the model is capable of predictions in two major visual search paradigms, including one in which involves parallel feature-based selection and elimination of targets. The model predicts timing of the search task well, but without recording eye movements, there is often a potential for trading off zone size (fovea and parafovea) with dwell time. For example, if the parafovea were twice as large but the dwell time was doubled, a similar fit might be obtained. Along with measuring eye movements directly, another way to constrain the model is with a flicker-based search task, a commonly-used search paradigm that yokes eye movement times to a fixed frequency of presentation. We will next examine two flicker tasks to demonstrate the model's flexibility and help constrain its assumptions about timing of movement. On their surface, the two tasks appear very similar, but produce performance profiles that differ substantially. Thus, it will be important to examine the aspects of the model that change in order to account for these across-task differences.

Flicker-paradigm change detectione

A commonly used paradigm involving visual search is the flicker change detection task (see Rensink, O'Regan, & Clark, 1997), in which two visual stimuli (either artificial or natural) that differ in some small way are shown repeatedly in succession, with a brief empty 'flash' between them (e.g., 50 ms) that prevents low-level visual change detection and requires deliberate search among the targets to find the difference. Here, even if detection can be done quickly, participants often cannot benefit from more than one eye movement per flash, which constrains the rate of information search. In this experiment, we examined a relatively difficult version of the task that incorporated four different types of change while varying the number of distractors.

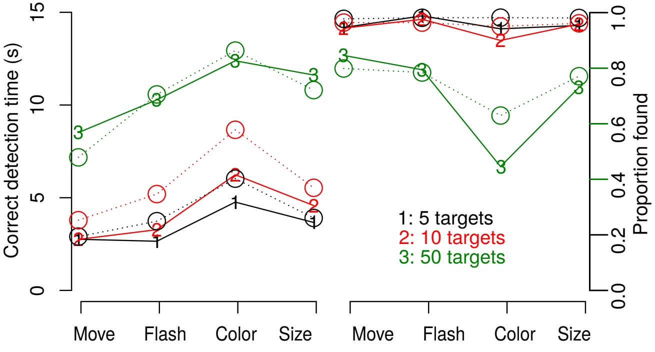
Methods. The present data was collected on the same groups of participants as the visual search task. This task used a modified version of the PEBL dot-flicker change blindness task (changeblindness), and involved three stimulus conditions with 5, 10, or 50 distractors. Each trial involved one of four change types: a color change, a size change, a position change, or a target disappearance. The entire stimulus visual field was approximately 20°x 15°. Each display frame appeared for approximately 450 ms, followed by a 50 ms blank flash, which was sufficient to disrupt low-level visual cues of change. Participants were permitted a maximum of 30s to find the change, which they then indicated by clicking with the mouse at the location of change.

Modeling the task differs from the previous models in that there is no known a priori probe, so the notion probe-item similarity is not applicable. Consequently, for the model, we interpret the zone similarity values to indicate the available evidence for a change across the flicker mask. The model assumes that no real information is available in the periphery, but evidence for a flicker-change will often be detectable in the parafovea (targets have similarity around .8 whereas non-targets have similarity around .6), which can then direct a foveation to confirm and localize the change. Because the rate of search is constrained, the main dependent timing measures are constrained by the effective size of the parafovea, which we adjusted to improve fit to data.

Results. Although small differences were observed in time and accuracy between 5 and 10-target displays, participants were considerably less accurate and slower on the 50-target display. As shown in Figure 5, the flash condition was slower and less accurate than the move condition (a move is essentially a double-flash), and size-change tended to be about as difficult as the flash condition. Color change was by far the most difficult condition. The model produced reasonable fits to the data, although it overpredicted the time needed to find the target on the smaller displays. The model assumes that the difficulty of different conditions arises because of a failure to detect changes of different types once a potential change is foveated, as shown in Table 1. In addition, the parafovea zone had a radius of 5.5°. This indicates that such changes may be available quite far from the fovea, but may often go undetected, which would require frequent revisits to previous locations.

Before discussing the results of the change detection task, we will examine a second study using an alternate version of the task that employs top-down cueing, and thus permits probe-item similarity to play a role in the search task.

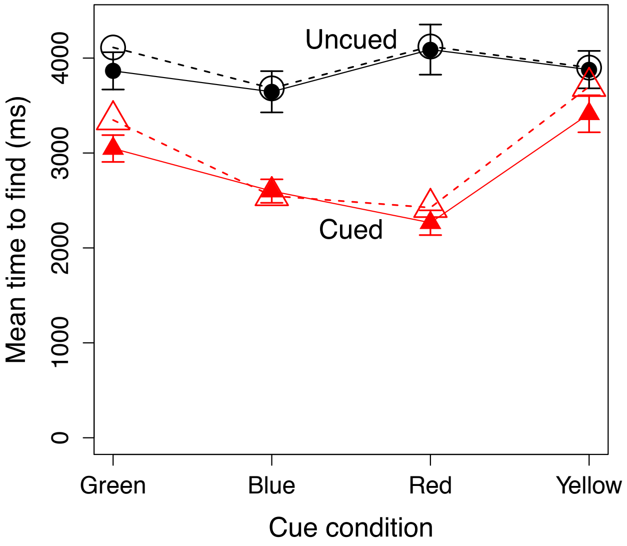
Figure 5: Model fits to dot-flicker task.



Top-down control in flicker-based change detection

Methods The final data set also used the flicker paradigm, but differed from the previous task in several ways. First, the field of view was larger (28°x 28°), and only one type of change occurred (a single target appeared and disappeared). On each trial 40 symbols appeared, drawn randomly from a set of four colored symbols (red, green, blue, and yellow squares). The larger field of view and more uniform targets made the task substantially easier, perhaps because of crowding and spacing effects (see Pelli, 2008). On half the trials, a color cue was given indicating the color of the change.

Figure 6: Model fits accounting for top-down control in change detection task



Results The single change type, coupled with the more dispersed display, produced a much easier task than the previous one: mean time to find in the 40-target task was under 4 s, comparable to the 10-target condition in the previous experiment. Furthermore, accuracy was close to 100%; in contrast to the 50-80% accuracy produced in the 50-target condition of the previous study. In addition, when cued, the time was reduced further, although the advantage for different colors differed, depending on the color.

To model these data, we assumed that different colors produce a different probability of detecting a change in the fovea, similar to the previous model. However, on cued trials, probe-item similarity is used to eliminate potential targets and constrain search. The best case scenario reduces target locations to around 10, but because of random sampling and layout of points, this does not reduce the time-to-find to 1/4 of the original. The model accurately predicts the that scale of this reduction is about 1/3. Differences in the color cue conditions were modeled by adjusting the peripheral similarity of different colors to the cue, so that in the case of red, typically 10/40 targets needed to be searched, but in the case of yellow and green, closer to 20/40 targets needed to be searched (because of their similarity). In these models, yellow and green often cannot be distinguished rapidly in the periphery, and so more of these targets were foveated to eliminate the foils. Overall, although some of the detection parameters were substantially different from the previous experiment, the model produces reasonable fits for both with interpretable changes in parameters (see Table 1). As in the previous change blindness model, parafovea size was slightly larger than the search task–in this case 4.0°. This is a consequence of the fact that search times on the order of 3-4 s necessarily involve at most 6 to 8 foveations, and the only way to reliably cover the visual field is if each foveation obtains information from this area. Thus, both change detection models suggest that change can be detected at 4 or more degrees from fixation, and that color-changes can be especially easy to miss (with failure rates af around 40% when fixated).

Parameters of models

Table 1 summarizes the main parameters used across tasks. Results are mainly impacted by assumptions about how long each detection/decision phase take, and the probability of detecting information in different visual zones-especially probability of detection failure in the fovea.

Discussion

The EM4 adopts and adapts the minimal model assumptions proposed by Halverson & Hornoff (2002), and extends the model to capture primary effects of several visual search paradigms. We have demonstrated the effectiveness of the model against four data sets, and the parameter settings for these models provide insight into the timing, accuracy, and availability of information. We will conclude by identifying some of the main lessons we have learned from these models.

Lessons of the models

Fovea zones. Accurate prediction of times and accuracies in search tasks require accounting for the information available in different foveal zones, and decisions about both presence and absence of this information.

Peripheral information. Substantial information about both presence and absence of information is available in the visual periphery (i.e., for pop-out tasks) and parafovea (for all tasks), and search is often guided by presence and absence of this information in all three zones.

Target rejection. Rejection of targets frequently occurs without foveation; identification of targets often is coupled with foveation.

Parafoveal preference. Search times can typically be adequately accounted for by a model that attempts to first confirm any high-likelihood targets parafoveally, and then maximize information gained in each subsequent movement.

Probe-item similarity. Probe-item similarity is useful in predicting a number of effects of visual search, sot that models of visual salience may benefit from incorporating probe information.

Conclusions. The EM4 intends to be a simple minimalistic model of foveated eye movement search. It is a standalone model, and so it may be useful for lightweight practical evaluation in human factors domains, as a simulation model education contexts, and as a lightweight front end for visual processing models that produce activation or posterior probability scores that can be interpreted as a probe-item similarity.

References

Boff, K. R., & Lincoln, J. E. (1988). *Engineering data compendium. human perception and performance* (Tech.

Table 1: Distinct parameters settings used in different paradigms.

Model	Fixation Time	Response Time	Parafov. radius	Detection failure	Parafov. Similarity	Peripheral Similarity	Scaled Deviation	*R*
Menu search	250	100	3.5	.09	.7	.6	.08/.065/.091†	
Visual search	200	500	3.5	0	.7	.45,.3,.3,.25	.16	.94
Change Detection I	500	100	5.5	.05,.05,.4,.09	.775-.825	.6	.16/.063‡	.97
Change Detection II	500	100	4.0	.04,.02,.05,.04	.7/.85	.48,.46,.4,.5	.045	.98

Note: Scaled deviation is mean of absolute error divided by observed value; *R* is Pearson's correlation. Multiple values in each parameter cell indicate values for distinct conditions. †Values for response time, fixations, and saccade distance, respectively. ‡Values for response time and accuracy.

Rep.). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Bruce, N. D., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of vision*, *9*, 1–24.

Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., et al. (2016). *MIT saliency benchmark.* http://saliency.mit.edu/.

Choi, Y., Han, J., Oh, H., & Myung, R. (2015). Cognitive model of human visual search with saliency and scene context for real-world images. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 59, pp. 706–710).

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive psychology*, *30*(1), 39–78.

Drury, C. G. (1975). Inspection of sheet materialsmodel and data. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *17*(3), 257–265.

Edwards, J. D., Ross, L. A., Wadley, V. G., Clay, O. J., Crowe, M., Roenker, D. L., et al. (2006). The useful field of view test: normative data for older adults. *Archives of Clinical Neuropsychology*, *21*(4), 275–286.

Halverson, T., & Hornof, A. J. (2004a). Local density guides visual search: Sparse groups are first and faster. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 48, pp. 1860–1864).

Halverson, T., & Hornof, A. J. (2004b). Strategy shifts in mixed-density search. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 529–534).

Halverson, T., & Hornof, A. J. (2007). A minimal model for predicting visual search in human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 431–434). ACM.

Halverson, T., & Hornof, A. J. (2011). A computational model of "active vision" for visual search in humancomputer interaction. *Human-Computer Interaction*, *26*(4), 285–314.

Hornof, A. J., & Halverson, T. (2003). Cognitive strategies and eye movements for searching hierarchical computer displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 249–256). ACM.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis & Mach. Intelligence*(11), 1254–1259.

Kieras, D. E. (2010). Modeling visual search of displays of many objects: The role of differential acuity and fixation memory. In *Proceedings of the 10th international conference on cognitive modeling* (pp. 127–132).

Kieras, D. E. (2011). The persistent visual store as the locus of fixation memory in visual search tasks. *Cognitive Systems Research*, *12*(2), 102–112.

Kieras, D. E., Hornof, A. J., & Zhang, Y. (2015). Visual search of displays of many objects: Modeling detailed eye movement effects with improved epic. In *Proceedings of the 13th international conference on cognitive modeling.*

Kieras, D. E., & Meyer, D. E. (1997). An overview of the epic architecture for cognition and performance with application to human-computer interaction. *Human-computer interaction*, *12*(4), 391–438.

Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of neuroscience methods*, *222*, 250–259.

Mueller, S. T., & Weidemann, C. T. (2012). Alphabetic letter identification: Effects of perceivability, similarity, and bias. *Acta Psychologica*, *139*(1), 19–37.

Nyamsuren, E., & Taatgen, N. (2013). Pre-attentive and attentive vision module. *Cogn. Sys. Research*, *24*, 62–71.

Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current opin. in neurobiology*, *18*, 445–451.

Perelman, B., & Mueller, S. T. (2015). Identifying mental models of search in a simulated flight task using a pathmapping approach. In *Proc. of the 18th International Symposium on aviation psychology (ISAP 2015)* (pp. 398–403).

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological science*, *8*(5), 368–373.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research*, *1*(4), 201–220.

Teo, L., & John, B. E. (2008). Towards a tool for predicting goal-directed exploratory behavior. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 52, pp. 950–954).

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. (01755)