TAM: Explaining off-object fixations and central fixation tendencies as effects of population averaging during search

Gregory J. Zelinsky
Department of Psychology, Stony Brook University

Abstract

Understanding how patterns are selected for both recognition and action, in the form of an eye movement, is essential to understanding the mechanisms of visual search. It is argued that selecting a pattern for fixation is time consuming—requiring the pruning of a population of possible saccade vectors to isolate the specific movement to the potential target. To support this position, two experiments are reported showing evidence for off-object fixations, where fixations land between objects rather than directly on objects, and central fixations, where initial saccades land near the center of scenes. Both behaviors were modeled successfully using TAM (Target Acquisition Model; Zelinsky, 2008). TAM interprets these behaviors as expressions of population averaging occurring at different times during saccade target selection. A large population early during search results in the averaging of the entire scene and a central fixation; a smaller population later during search results in averaging between groups of objects and off-object fixations.

Keywords

saccade target selection; global effect; center of gravity fixations; selection for action

Selection for Recognition versus Selection for Action

Selection, the process of choosing one thing out of potentially many, is widely believed to be a core function of attention, and visual search is one of the best behavioral tasks for studying the selection process. In a typical visual search task a person is asked whether a particular target pattern is present or not in a display containing one or more nontarget patterns. Making this determination requires matching features of the target to features present in the search display. A problem exists, however, in the fact that many different locations in the search scene might match the features of a complex target. Two main solutions to this problem have been proposed. One treats this problem as a failure of feature binding, requiring attention to move to locations in the search display so as to bind the features at these locations into unitized perceptual objects (Treisman, 1988; Treisman & Gelade, 1980). Another proposed solution sidesteps the binding problem and assumes that the features from a given location can be preattentively evaluated in terms of spatially local evidence for a target, with the totality of this evidence creating a sort of probability map of potential target locations that can be used to prioritize the allocation of additional processing via selective attention (Wolfe, 1994; Zelinsky, 2008). Although the role of this additional processing is often left unspecified by theories, the assumption is that its function again is to create stable perceptual objects for the purpose of recognizing a target (e.g., Treisman,

Address correspondence to: Gregory J. Zelinsky, Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500, Phone: 631-632-7827, Fax: 631-632-7876, Gregory.Zelinsky@stonybrook.edu.

1988). Both solutions therefore assume that recognition requires that the features of the to-be-recognized pattern be segregated to a specific location in space, with selection being the process through which this feature segregation is accomplished. This process will be referred to as *selection for recognition*.

This selection for recognition view can be contrasted with selections made in the service of action. *Selection for action* theories (Allport, 1987; Neumann, 1987; Norman & Shallice, 1986) contend that many selections occur, not to recognize objects, but to interact with them. When assembling a piece of furniture several tools may be scattered about a personal workspace, all of which are recognized. But because different tools are needed at different times in the task, successful task performance requires treating each component action as a separate sub-task, each with its own selection requirements (Hayhoe & Ballard, 2005; Land & Hayhoe, 2001). Selection in this case involves choosing a particular object to couple with a particular effector. A screwdriver may need to be coupled to the right hand, while a wrench is coupled to the left. These couplings, although they occur in visuo-motor space, are nevertheless selections; out of all the available effectors and all the possible objects, very specific couplings must be chosen.

In the context of a search task the ongoing actions are the eye movements that a person makes as they search.[1] Eye movements are like many other actions in that they reflect a very specific coupling between an effector (the eyes) and the location of a visual object; given that gaze cannot move simultaneously to two spatially separated objects, each object selected for fixation reflects a decision not to look at any other object in the search scene. However, eye movements are particularly interesting cases of action selection because these selections, like the selections underlying recognition, occur very rapidly, are incontrovertibly serial, and typically reflect specific couplings to object locations in a search display. There is even a functional relationship between the selections made in the service of recognition and eye movements; a selected pattern would be most easily recognized if the high-resolution fovea were first brought to that pattern via an eye movement (see Deubel & Schneider, 1996, and Zelinsky, 2008, for additional discussion of this relationship).

One unfortunate consequence of the yoked relationship between the selections for eye movements and those for recognition has been the relative neglect of the former in favor of the latter. For years the visual search literature has embraced the idea that the search process consists of a series of very rapid selections in the service of recognition, as exemplified by the historically popular high-speed serial search models (e.g., Horowitz & Wolfe, 2003; see Ward, Duncan, & Shapiro, 1996, for additional discussion). According to these models, the core search process is fundamentally devoid of any overt action, with selections instead made by a covert process of attention acting as a rapidly moving internal pointer to locations in space. Advocates of this perspective have even assumed that the speed of this covert movement can be estimated by the slope of the function relating manual search times to the number of objects in a search display (the set size effect), often discounting the contributions of actual movements of gaze to the steepness of these functions (e.g., Treisman & Gormican, 1988).

More recent work makes clear that search is undeniably an active process consisting of actual motor behavior (Henderson, 2007; Findlay & Gilchrist, 2003); unless instructed or prevented from doing so, people overwhelmingly elect to move their eyes as they search. And although differences between eyes-fixed and eyes-free search may be small under some laboratory conditions (e.g., Klein & Farrell, 1989), these similarities in eye movement

---

[1]Of course search judgments are also usually indicated by a manual button press, this action occurs after the search task has been completed (Zelinsky & Sheinberg, 1997) and is therefore contingent on the target's recognition.

behavior are unlikely to generalize broadly to search tasks in the real world. Indeed, many everyday search tasks probably require the searcher to move their eyes to the target in order to make a confident search decision (Geisler & Chou, 1995; Zelinsky, 2008), particularly in cases of small and featurally complex targets embedded in cluttered scenes (Henderson, Chanceaux, & Smith, 2009; Henderson, Weeks, & Hollingworth, 1999; Neider & Zelinsky, 2008, in press; Sheinberg & Logothetis, 2001). Given this demonstrable behavior, it is fairer to characterize the search process as an intertwined sequence of recognition and action selections; after the selection of an object for recognition, very often an action must be selected so that gaze can be brought to the to-be-recognized object.

This relationship between recognition and action also complicates the role of object selection during search. To the extent that the objects selected for recognition and action are one and the same, one might argue that choosing to focus on the attention movement or the eye movement becomes a matter of convenience, with no real consequences for understanding visual search. The following section challenges the wisdom of treating these two forms of selection as equivalent. Although the objects selected for fixation and recognition may indeed be one and the same, this shared goal may mask underlying differences in how these selections are implemented. The proposal here is that the shift of attention to an object for the purpose of recognition and the act of physically moving gaze to that same object places very different demands on a selection process, largely due to pre-processing simplifying one task but not the other.

Rank-ordering versus population-pruning: Two modes for selecting saccade targets

Selection for Recognition and the Rank-ordering of Search Targets

Objects are important to search. The clearest example of this is the set size effect, where search efficiency is typically found to decline with each featurally-complex object added to the display (Treisman & Gelade, 1980; Wolfe, 1998). Whether explicitly stated or not, most search theories explain set size effects in terms of selections for recognition. The more objects that are in a scene, the more selections that are needed to recognize these objects.[2] Of course this explanation assumes that search objects cannot be recognized preattentively, an assumption bolstered by the many studies in the early attention literature supporting *early selection theory* (Broadbent, 1958)—the view that selections are based on early visual or auditory features and not on semantic information that implies the prior recognition of objects (see Pashler, 1998, for a comprehensive review of the early versus late selection debate). The products of these early perceptual processes can be usefully described as *proto-objects* (Rensink, 2000; Schneider, 1995; Wischnewski, Belardinelli, Schneider, & Steil, 2010), perceptually-grouped constellations of features that have not yet been recognized as objects. It is these proto-objects that will be assumed to be the candidate patterns for potential consideration as search targets, and that selections occur among these patterns for the purpose of prioritizing their recognition.

Although proto-objects typically must be recognized before they can be classified as targets or distractors (ones having a distinctive target feature may violate this rule, a phenomenon known as *pop out*; Treisman & Gelade, 1980), the process of selecting proto-objects for recognition is not random. Search theory has long contended that patterns are preattentively prioritized for selection. One of the earliest suggested forms of prioritization proposed that

---

[2]Why an object must be selected to be recognized is a question beyond the scope of this paper, but one plausible speculation is that selection is needed to constrain the routing of feature information to higher visual areas responsible for recognition (Olshausen, Anderson, & Van Essen, 1993).

patterns are rank ordered based on their *salience*, a measure of low-level feature contrast between a pattern and its neighbors (Koch & Ullman, 1985; Itti & Koch, 2000, 2001). According to this view, a map of these saliency values is computed for a search scene, with the most salient point on this map determining the proto-object that will be selected next for recognition. Wolfe and colleagues (Wolfe, 1994; Wolfe, Cave, & Franzel, 1989) elaborated on this idea by suggesting that such bottom-up prioritization should be combined with top-down evidence for a search pattern matching the target. According to this *guided search* model, patterns in a search scene are preattentively analyzed in terms of basic visual features and compared to features of the target to obtain a top-down guidance signal; the better the match to the target, the stronger this signal. These top-down signals are then combined with bottom-up feature contrast signals to form an activation map, with the values on this map ultimately determining the order in which proto-objects will be recognized in a given scene.

Regardless of the name given to the prioritization map, or whether this prioritization is accomplished via bottom-up analysis (e.g., Itti & Koch, 2000), top-down analysis (e.g., Zelinsky, 2008), or some combination of the two (e.g., Wolfe, 1994; Zelinsky, Zhang, Yu, Chen, & Samaras, 2006), critical to the current discussion is the fact that the preattentive construction of these maps provides a built-in prioritization of the selection-for-recognition decisions. The proto-object in the scene corresponding to the most active points on such a map will be the first item to be recognized, with the next most active item selected for recognition if the first is determined not to be the target. This process would continue until the target is detected or each of the proto-objects specified on this prioritized list is recognized and rejected. The recognition of each proto-object might therefore demand attention and consume time, but the process of selecting these patterns does not. According to modern search theory the problem of deciding the order in which proto-objects are recognized is not a problem at all, it has already been solved by the processes used to create these various prioritized maps.[3]

Selection for Eye Movement and Population Codes

Whereas the selection of proto-objects for recognition is minimally demanding, equivalent to shifting an internal pointer down a list of rank-ordered items, the task of selecting an action to couple with an object is probably not. This is because actions are typically coded by motor systems using large distributed populations (Sparks, Kristan, & Shaw, 1997). One of the clearest examples of this is the coding of saccadic eye movements by the superior colliculi. Each colliculus consists of a map of visuo-motor space, with specific cells in this map coding specific saccade vectors. However, moments before an eye movement activity is observed over a very large population of these cells (approximately 25–28% of the collicular map; Anderson, Keller, Gandhi, & Das, 1998), with the direction and amplitude of the eye movement determined by averaging the contributions from the individual motor signals (Lee, Rohrer, & Sparks, 1988; McIlwain, 1975, 1991). This is a population code; the computation of the saccade vector is distributed over a population of cells on the collicular map, with each cell in this population getting a vote in the ultimate motor behavior.

The selection-for-action task in the context of visual search is to send gaze to the proto-object that has already been selected for recognition. To accomplish this goal, the eye movement system must select the one saccade vector that would bring gaze to the location of the suspected target from among all the other locations in the search scene to which gaze might be directed (all nontarget locations). Unlike the prioritized selections made in the service of recognition, what makes this task non-trivial is the fact that the eye movement

---

[3]Although a saliency map or an activation map as a whole might be considered a sort of population code (Zelinsky, 2008), the time needed to find the peak on these maps is not thought to increase with the number of active points (the size of the population).

system must be prepared to position gaze on any thing at any time. Such a "just in time" selection of saccade targets, (e.g., Hayhoe, Shrivastava, Mruczek, 2003; Zelinsky & Todor, 2010) requires that multiple potential actions (saccade vectors) be considered at the start of a search, and more plausibly during each search fixation; there would be no point to process information just in time if it were not possible to act on this processing. It is this need to flexibly react to the moment-by-moment demands of a task that necessitates a readiness to fixate different locations in a scene, and an embodiment of this readiness by the activation of a large population of potential saccade vectors. In the absence of preattentive prioritization, pruning this population to select the one desired vector takes time, making the demands of this selection task far greater than those of selecting a pattern for recognition.[4]

The Target Acquisition Model: Combining selections for recognition and action

The Target Acquisition Model, TAM for short, combines selection for recognition and selection for action in the context of a single theory of eye movements during visual search (Zelinsky, 2008). TAM takes as input an arbitrarily complex visual pattern as a target, and an image of an arbitrarily complex search scene, which it dynamically transforms after each fixation to reflect human retinal acuity limitations. It then attempts to detect the target in this scene, a behavior that is usually accompanied by a series of eye movements that ultimately align a simulated fovea with the target. Driving this behavior is a map representing the visual similarity between the target and the search scene. This *target map* is computed by correlating a vector of color, orientation, and scale-selective filter responses obtained for the target with an array of filter response vectors obtained for the search scene. Each point on this map therefore provides an estimate of visual similarity between the target and that location in the scene, with the point of maximum correlation, referred to as the *hotspot*, being TAM's best guess as to the likely target location. In this sense, TAM's target map is related in principle to other map-based methods of prioritizing search behavior (Itti & Koch, 2000; Wolfe, 1994), with the difference being that this prioritization is derived exclusively from target-scene similarity and not bottom-up feature contrast.

TAM is a selection for recognition theory in the sense that it internally monitors activity on the target map after each fixation to determine if evidence for the target on this map exceeds a high detection threshold. More specifically, if the hotspot correlation is greater than this threshold, TAM responds that the target is present without shifting its simulated fovea, even if it is not yet fixated on the target pattern. However, TAM's selection of the hotspot correlation for comparison to the detection threshold is not assumed to be attentionally demanding, just as the guided search model assumes no demands in finding the highest activation peak and saliency models assume no demands in finding the most salient point. TAM therefore joins the ranks of other search theories in not treating the task of selecting objects for recognition as an attention-demanding problem; in the case of TAM the correlation operation that creates the target map delivers to the search process a prioritized list of target candidates for inspection, making selection from this list trivial. Whereas there is a cost associated with selecting the wrong candidate for recognition (the time needed to reject the object as a distractor), the preattentive prioritization of candidates makes the selection task itself relatively cost free.

Although TAM can occasionally make search decisions without making an eye movement, this behavior is the exception rather than the rule. More often the hotspot correlation fails to

[4]Although sequences of eye movements can certainly be pre-programmed (e.g., Zingale & Kowler, 1987), a form of motor prioritization, such pre-programming probably characterizes only a small percentage of eye movement behavior in the real world (with reading being a common exception to this rule; see Schad & Engbert, this issue).

meet the high detection threshold when TAM is not fixated on the hotspot pattern. This might result for one of two reasons: either the hotspot pattern is not the target (in which case its correlation *should* be below threshold), or it is the target but peripheral blurring from simulated retinal acuity limitations has made its correlation too weak to allow its detection. These two possibilities, that the peripherally viewed hotspot pattern is either a distractor or a blurred target, creates uncertainty for TAM, and it attempts to resolve this uncertainty by making an eye movement to the suspected target. If the pattern's correlation still does not meet the detection threshold after its fixation by TAM's high resolution fovea, this object is rejected as a distractor, its location is inhibited on the target map, and TAM sets out to fixate the next most correlated point in its search for the target. The fact that TAM has only one fovea means that these recognition decisions will often unfold serially, similar to the search dynamic described by other approaches (Itti & Koch, 2000; Wolfe, 1994). What distinguishes TAM from these other approaches is the reason why multiple distractors must often be fixated and recognized before the target. Although all distractor recognitions ultimately reflect a failure to correctly prioritize target evidence in a scene, under guided search this faulty prioritization is due to internal noise causing distractors to be confused with targets, under TAM eye movements are directed to proto-objects in order to overcome retina acuity limitations that prevent patterns from being classified as targets or distractors. TAM therefore makes eye movements during search for the same reason that people do— to gather information so as to make better and more confident search decisions.

Although TAM's eye movements ultimately serve the recognition decision, their production means that TAM must also be a model of selection for action. Just because the location of a suspected target is known (the hotspot), the movement needed to bring gaze to that location must still be selected from all possible others. This is because just-in-time processing requires a readiness to make an eye movement to many locations in the visual field, thereby forcing TAM to confront the problem of selecting only one saccade vector from a population of many. This population of competing saccade vectors is visualized by the target map (Figure 1). Each point on this map represents a potential location in the scene to which TAM might shift its gaze, and the value at each point indicates the likelihood of the target being at that location. To select and fixate the hotspot location, TAM prunes the population of potential saccade vectors over time, progressively removing those locations offering the least evidence for the target. If this pruning proceeds until the hotspot is completely isolated on the target map (all less correlated vectors are removed), TAM's gaze would go directly to the suspected target. However, if TAM makes an eye movement before the pruning process has completed, an event determined by the signal and noise properties of the target map, then the spatial average of the remaining target map activity is computed and gaze is sent to this population centroid. TAM therefore captures a key property found in many motor systems, the fact that movements reflect activity averaged over large distributed populations.

EXPERIMENT 1

The fact that TAM can exhibit population averaging behavior means that the location of the to-be-recognized object, indicated by the hotspot on the target map, is partially dissociated from the location in the scene to which gaze will next move. Such a dissociation is unusual for theories of visual search, which more commonly assume that gaze is sent directly to the proto-object offering the most evidence for being the target (the top-ranked item on the prioritized list). These differing degrees of reliance on population coding lead to different predictions regarding how precisely objects should be fixated during search. Specifically, TAM predicts that eye movements should frequently fail to land precisely on objects, what I will refer to here as *off-object fixations*. Indeed, in all cases in which an eye movement is made before pruning has isolated the hotspot, population averaging will likely prevent gaze from landing directly on an object; despite TAM wanting to look at proto-object *h* (the

hotspot), target map activity from proto-objects x, y, and z would pull gaze away from this desired target. In contrast, search theories that do not use a population code predict that gaze should move serially from one proto-object to the next (e.g., Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Itti & Koch, 2000; Kanan, Tong, Zhang, & Cottrell, 2009; Wolfe & Gancarz, 1996), with their order of fixation determined by how the patterns were prioritized by preattentive processes. These theories predict relatively accurate eye movements to objects and few off-object fixations. To test these predictions I used search displays consisting of incontrovertibly distinct and well-separated objects, an O target among Q distractors. To the extent that off-object fixations are observed, this will be consistent with the population coding of eye movements during search, and the competition among multiple proto-objects for accurate fixation.

Methods

Eight undergraduate students from Stony Brook University participated for course credit. All had normal or corrected to normal vision, by self report. Each observer searched for a small O target among 8, 12, or 16 Q distractors. Targets and distractors subtended 0.48° visual angle, and were constrained to appear at 39 locations in the search display. As a result of these placement constraints, neighboring objects were separated by at least 2.1° (center-to-center distance), and no item could appear within 2.1° of the display's center, which corresponded to the starting gaze position. The entire search display subtended 16.8° × 12.6°.

A trial began with the presentation of a central fixation dot, followed by the search display. The observer's task was to press a button upon locating the O target. Because a target was present in each display, a localization procedure was used to confirm that observers were engaged in the task. After the button press signaling that the target was located, the search display was replaced with a localization grid showing a white square frame (0.97°) at the location of each search item. Observers were instructed to fixate in the frame that had contained the target, an event which caused the frame to turn red, then to make a second button press. This second manual response initiated the next trial. No accuracy feedback was provided. There were 180 experimental trials and 30 practice trials. The experimental trials were evenly divided into the three set sizes, leaving 60 trials per cell of the design.

Two different eye trackers were used for data collection. Eye data from six observers were collected using an EyeLink II eye tracker (SR Research). This video-based tracker has a high spatial resolution (0.97° RMS) and allows for 500 Hz sampling. A chin cup was used to minimize head movement and to control viewing distance, and calibrations were not accepted until the average error was less than 0.4° and the maximum error was less than 0.9°. Eye data from two subjects were collected using a Generation 6 dual Purkinje-image eye tracker (Fourward Technologies), interfaced with a 12 channel analog-to-digital converter sampling right eye position at 1000 Hz. This tracker has a spatial resolution of +/− 1 arc minute and has long been used in studies demanding a high degree of precision (Crane & Steele, 1978). A dental impression bite-bar was constructed for each participant, and calibrations were not accepted until the repeated fixation of all nine calibration targets resulted in a maximum eye position error of less than 0.25°. Other than these calibration instructions, and a reminder to look at the fixation cross whenever it appeared on the screen, observers were not told how to direct their gaze. Saccadic eye movements were extracted online using a velocity-based algorithm implementing a roughly 12.5 °/sec detection threshold. Fixation locations were defined as the average eye position between the offset of one saccade and the onset of the next.

Results and Discussion

Observers correctly localized the search targets on more than 99% of the trials; data from the few error trials were discarded. This high level of localization accuracy was accompanied by the accurate fixation of targets prior to the button response terminating the search display. For the EyeLink II observers the average distances between gaze and the target at the time of the initial button press were 0.84°, 0.97°, and 0.88° for the 9, 13, and 17-item conditions, respectively. For the Purkinje observers the corresponding distances were 0.87°, 0.91°, and 0.85°. These gaze-to-target distances did not differ across the three set sizes or between the EyeLink II and Purkinje observers (all p .52). All subsequent analyses will therefore collapse across these manipulations.

Observers clearly chose to fixate targets in this task, but how did this ultimately happen—did search proceed neatly from object to object, or was it more sloppy, with sometimes inaccurate fixations on distractors? To address this question I found the distance between every fixation and the nearest object, excluding initial fixations (which were always at the search display's center). The distribution of these distances is shown by the white bars in Figure 2. There are two notable patterns in these data. First, observers often fixated accurately on objects as they searched. On average, 44% of these fixations were highly accurate, within 0.50° of an object, and another 28% were near to an object, within 1.0°. Collectively, I will refer to these as *on-object fixations*, where fixation on an object is liberally defined as less than 1.0° from center (approximately four times the object's radius). This frequency of on-object fixations is very similar to the 80% rate reported previously using monkey observers (Motter & Belky, 1998). Second, this distribution of gaze-to-object distances could be well fit by an exponential function ($R^2 = 0.96$), and it is the long tail of this function that is of particular interest in the current context. Although the majority of fixations during search were on or near objects, roughly 28% of these fixations were clearly not.

On average, about 500 saccades per observer landed more than 1.0° from an object during this search task. Figure 3a shows these off-object fixations for one of the Purkinje observers. The 541 behaviors illustrated in this figure cannot be explained by any theory of visual search proposing direct movements of gaze from one proto-object to another (Ehinger et al., 2009; Itti & Koch, 2000; Kanan et al, 2009; Wolfe & Gancarz, 1996). This is because these theories assume no competition between proto-objects after their rank ordering for selection —the top-ranked object is always selected (not the top two, three, etc.), and gaze is sent directly to that object. Left unexplained, these behaviors have either been hidden, typically by the practice of assigning each fixation to the nearest object regardless of the fixation-to-object distance (e.g., Williams & Reingold, 2001; Zelinsky, 1996), or dismissed as expressions of noise in the eye movement system, if they are acknowledged at all (for other reports that segregated on-object from off-object fixations, see: Findlay & Blythe, 2009; Findlay & Brown, 2006; Motter & Belky, 1998). Note, however, that this explanation is unsatisfying. If this noise can be so great as to produce saccade errors in excess of a degree, how can observers be so successful in ultimately fixating targets with such high precision? [5]

Rather than treating these behaviors as errors, or ignoring them altogether, under TAM they are an expected consequence of the selection of saccade locations from a population code. The black bars in Figure 2 show the distribution of gaze-to-object distances as TAM searched for the identical Q target in the identical 180 search displays shown to observers.

---

[5]Estimates of motor noise in the saccadic eye movement system vary with saccade amplitude (e.g., Abrams, Meyer, & Kornblum, 1989). The mean amplitude of saccades landing more than 1.0° from an object in Experiment 1 was 2.6°. Assuming conservatively that motor noise is 10% of saccade amplitude, this produces a mean noise estimate of only .26°, roughly one quarter of the minimum gaze-to-object distance used to define off-object fixations in this experiment.

TAM's parameters were unchanged from Zelinsky (2008), and that paper should be consulted for a more detailed description of TAM and the setting of its parameters. Turning first to the on-object fixations, TAM fixated objects a bit too accurately. Compared to observers, TAM made more highly accurate fixations ($< 0.5°$) and fewer moderately accurate fixations ($< 1.0°$) on objects. However, TAM's representation of targets relative to a single point in the target image (the target vector point; Zelinsky, 2008) makes this aspect of its behavior difficult to compare to observers. When TAM manages to isolate the hotspot on the target map, its gaze will be directed to this single point, resulting in many highly accurate saccades on objects and an unrealistically low variability in the gaze-to-object distances for these fixations. Human behavior is undoubtedly less constrained in this respect. Even if observers were representing targets relative to a single point, there would likely be some variability in the point that they would choose. This would tend to increase variability in the gaze-to-object distances, producing fewer highly accurate object fixations ($< 0.5°$) and more moderately accurate fixations ($< 1.0°$). Presumably, if variability was also introduced in TAM's representation of targets then it too would exhibit similar behavior, although this conjecture remains untested.

More relevant to the current discussion is TAM's characterization of off-object fixations during search. As illustrated by the four rightmost pairs of bars in Figure 2, TAM's behavior agreed quite well with that of observers. TAM captured the same exponentially decreasing pattern found in the behavioral data; as the distance between gaze and the nearest object increased, the number of off-object fixations decreased. Moreover, the absolute number of these fixations also agreed with human behavior. For each of the four distance bins tested ($1.0–1.5°$, $1.5–2.0°$, $2.0–2.5°$, $> 2.5°$), the number of off-object fixations made by TAM was within the 95% confidence interval surrounding the corresponding behavioral mean. Overall, 31% of TAM's fixations were off-object, compared to an average of 28% from observers, with all 558 of these off-object fixations appearing in Figure 3b. Whereas most theories of visual search discount these off-object fixations as errors, or else ignore them completely, TAM is able to account for these behaviors in addition to the core search behaviors reported elsewhere (Zelinsky, 2008), all without the introduction of any new processes or parameters.

EXPERIMENT 2

Whereas off-object fixations have attracted relatively little attention from the search and scene viewing communities, another oculomotor oddity, the observation that fixations tend to cluster towards the center of natural scenes (Busswell, 1935; Mannan, Ruddock, & Wooding, 1996, 1997; Parkhurst, Law, & Niebur, 2002), has received more thorough study. This central fixation bias is uncommonly robust, and has been reported in both static images and videos (Tatler, Baddeley, & Gilchrist, 2005; Tseng, Carmi, Cameron, Munoz, & Itti, 2009) and for scene inspection and search tasks (Tatler, 2007). Among the many explanations that have been offered for this behavior, several have been ruled out. For example, Tatler (2007) manipulated starting gaze position and found a pronounced tendency for the first eye movement to be directed towards the center of the scene. This is compelling evidence against the central fixation bias being a simple artifact of gaze starting at the center of the scene in most experiments, combined with a predisposition to make relatively small amplitude saccades during scene viewing (Tatler, Baddeley, & Vincent, 2006). In this same study, Tatler (2007; see also Tatler et al., 2005) also obtained local estimates of brightness, chromaticity, contrast, and edge information in images of scenes, and found a clear central fixation bias irrespective of the spatial distribution of these features in the images. This suggests that feature content plays little or no role in one's preference to fixate the center of a scene (but see Tseng et al., 2009). Interestingly, the central fixation bias appears to depend more on time and task; this bias was most pronounced in the first fixation for both search

and scene viewing, but by the second fixation the bias shifted in the direction of image features in the search task. In the scene viewing task the center location remained privileged.

Even after the exclusion of these low-level explanations for the central fixation bias, a host of higher-level potential explanations remain (Tatler, 2007; Tatler & Vincent, 2009; Tseng et al., 2009). Perhaps people have learned that interesting things are most likely to appear at the centers of pictures (photographer bias), and that people tend to look at these objects even when their feature content is low. Alternatively, maybe the center of a scene is in some sense an ideal location from which to collect information (Geisler, Perry, & Najemnik, 2006; Najemnik & Geisler, 2005), making the central fixation bias an intelligent first step in any systematic scanning strategy (Renninger, Vergheese, & Coughlan, 2007). The scene's center may also be the best position from which to extract its gist, or to derive some global feature for the purpose of recognizing the scene and constraining object detection (Torralba, Oliva, Castelhano, & Henderson, 2006).

While all of these explanations are reasonable and even plausible, one should be cautious when concluding for the existence of a high-level cause of a behavior on the basis of excluding low-level alternatives. The logical concern, of course, is that one can never be certain that all potential low-level explanations have been ruled out. Perhaps the central fixation bias *is* due to some low-level factor affecting saccade selection, one that has simply yet to be considered. It is in this role that TAM might be useful in teasing apart low-level from high-level explanations for the central fixation bias. TAM's behavior is completely deterministic; its eye movements depend purely on the target and scene representations, and its own internal dynamics. As such, it is completely devoid of high-level strategic factors or biases. This is not to say that such biases don't exist, only that they don't exist in TAM. Should TAM produce a central fixation bias, this behavior could therefore not be explained as strategic. In this respect, TAM becomes a tool to explore whether other low-level explanations of central fixation behavior have been overlooked.

Methods

Twelve undergraduate students from Stony Brook University participated for course credit, none of whom were observers in Experiment 1. All had normal or corrected to normal vision, by self report. Each observer searched for an O target in 78 images of natural scenes (see Figure 4a, b). The scenes subtended $20° \times 15°$ visual angle (1280×960 pixels), and depicted natural landscapes largely devoid of people, structures, or other artifacts used by people. These 78 scenes were evenly divided into randomly interleaved target present and target absent search trials. On target present trials, a small O (.48°), the same target used in Experiment 1, was digitally inserted into the scene. The target's placement in a given scene was pseudo-random, with the constraints that it would not appear within 2.1° of the image's center, that it (or a part thereof) would not be invisible against the local background, and that targets would be equally likely to appear in each image quadrant over scenes.

Each of the 78 trials began with the presentation of a fixation dot, followed by a search scene. The placement of the fixation dot varied across scenes, with these locations described by the 39 target locations used in Experiment 1 (see Figure 3). Specifically, these locations were constrained to one partial and three full rings surrounding the center of the image. The outermost partial ring contained 10 locations, each at 8.4° eccentricity relative to center, the innermost ring contained four locations, each at 2.1°, and the two middle rings contained 15 and 10 locations at 6.3° and 4.2°, respectively. As a result of these placement constraints, starting gaze positions were distributed evenly over the scenes, and were forced to be at least 2.1° from the scene's center. Each of the 39 starting gaze positions was used twice, once in a target present scene and again in a target absent scene. Within a given target condition (present or absent), each trial therefore had a different starting gaze position. Upon fixating

the dot and pressing a button, the program controlling the experiment checked to determine whether gaze was within .5° of the fixation target. If it was not, a tone sounded signaling the observer to look at the fixation target more carefully. If it was, the fixation dot would be replaced by the search scene. The search scene remained visible until the observer's target present or absent judgment, after which the fixation dot for the next trial was displayed. Eye position was sampled at 500 Hz using an EyeLink II eye tracker, with chin rest. Saccades and fixations were parsed from these samples using the tracker's default settings.

TAM searched for the same O target in the identical 39 target absent images shown to observers. Also consistent with the behavioral conditions, TAM's fovea was pre-positioned to the same 39 starting gaze locations specific to each of these scenes. TAM made three eye movements for each scene, after which its search was terminated. Other than these task-specific differences, this version of TAM was identical in both architecture and parameters to the version reported in Zelinsky (2008).

Results and Discussion

The goal of this experiment was to assess the expression of a central fixation bias for both observers and TAM. Because the actual search behavior is immaterial to this effort, the standard characterizations of search performance will not be reported, other than to say that observers were engaged in the task (misses were 7.2% and false positives were 1.6%). Indeed, and following Tatler (2007), analyses will be confined to the target absent trials so as to exclude influences on gaze resulting from the target's presence in an image.

A central fixation bias predicts that the endpoints of the initial search saccades should cluster near the center of the search scenes. This is precisely the pattern shown in Figure 5a. Of the 468 first fixations made by our 12 observers, 98.9% fell within the 2.1° (4.2° diameter) central window described by the innermost ring of starting gaze positions. The mean distance of these first fixations from center was 0.76°, and their tight clustering resulted in a standard deviation (SD) of only 0.50°. This clear evidence for a central fixation bias replicates the equally clear bias reported by Tatler (2007) in the context of a search task. Adding to the current story is the corresponding behavior from TAM, shown in Figure 5b. Of TAM's 39 first fixations, 97.4% fell within 2.1° of the scenes' centers. This distribution of initial saccade landing positions matched almost perfectly the behavior of the observers, both in terms of the mean distance of first fixations from center (0.58°) as well as the variability in this distribution of distances (SD = 0.57°).

The fact that so many first fixations clustered near the centers of the scenes suggests that these fixations resulted from saccades originating from many of the 39 starting gaze positions. This turned out to be the case. Figure 6 shows the starting position from which each of the center-biased saccades were made, for both TAM (red) and a randomly selected observer (blue). Several patterns are worth noting. First, regardless of the eccentricity or angular location of the starting gaze positions, initial saccades were almost always biased to the scenes' center. This occurred for both TAM and our representative observer. Second, the fact that the central fixation bias was as strong for far starting gaze positions (a) as it was for near (c) can be attributed to a relationship between starting gaze position and initial saccade amplitude; as the distance between starting gaze position and the center increased, so too did the amplitude of the initial saccades. Once again, this pattern was found for both TAM and our observer. Third, failures of a central fixation bias were most common from the starting gaze positions nearest to the center of the images. Of the 78 real and simulated saccades illustrated, only (d) shows two fixations having greater distances from center after the initial saccade compared to before. Interestingly, one of these is from TAM and the other is from our observer. Across participants, the percentage of central bias violations was 2.9% of the total first fixations, similar to the 2.6% of violations from TAM, with all but two of these

originating from the innermost ring of starting gaze positions. Although it is risky to generalize from such few cases, this suggests that a starting gaze position 2.1° from the center of an image is already sufficiently center biased, making a more precise eye movement directly to the center sometimes unnecessary.

How did the central fixation bias change over time? To answer this question we analyzed the expression of this bias in the second and third fixations made during search. The tight clustering of first fixations shown in Figure 5a for observers largely remained in the second fixations (Figure 5c), although there are signs of this cluster starting to relax. The mean distance of second fixations from center was 1.62° (SD = 0.96°), significantly greater than the corresponding distance found for first fixations, $t(11) = 28.6$, $p < .001$ (paired-group). By the third fixations (Figure 5e) the initially tight cluster fell apart. The mean fixation-to-center distance of these third fixations was 2.41° (SD = 1.21°), significantly greater than the mean distance for the second fixations, $t(11) = 8.4$, $p < .001$ (paired-group). TAM showed a similar gradual dissolution of an initially tight cluster of fixations (Figure 5, right). Its second fixations averaged 1.51° (SD = 0.82°) from center, and its third fixations averaged 1.80° (SD = 1.11°) from center. These patterns are consistent with a tendency for gaze to be drawn initially to the center of an image, following by small amplitude saccades away from center as observers, and TAM, begin a more earnest search of the scene.

Figure 7 summarizes many of the above-described data patterns. The mean distance between starting gaze position and the images' center differed depending on the eccentricity of the initial fixation dot, indicating that observers, and of course TAM, were pre-positioning their gaze on this dot as instructed. The large decrease in fixation-to-center distances following the initial eye movement reflects the very pronounced central fixation bias shown in Figure 5 (top) and Figure 6. Regardless of where in the scene gaze was initially positioned, by the first new fixation following search display onset gaze was very near to the center of the image. This occurred for both observers and TAM, whose fixation-to-center distances were all within the 95% confidence intervals surrounding each of the behavioral means. Fixation-to-center distances increased by the second fixations, and even more by the third, as gaze progressively moved away from the center as part of normal search. Perhaps interestingly, these fixations show no effects of different starting gaze positions; once drawn to the center of the scene, search proceeded from that location as if gaze had initially been positioned at the center. Throughout these initial search fixations TAM maintained reasonable agreement with human behavior.

The behavioral central fixation tendencies reported here are unremarkable. Similarly pronounced tendencies have been reported previously (Tatler, 2007; Tatler et al., 2005; Tseng et al., 2009). What is unique to this study is the finding that TAM also directed its initial saccades to the centers of the scenes. The fact that it captured almost perfectly the central fixation tendency of observers in this task is informative, as it points to a previously unconsidered low-level explanation for this behavior. TAM's eye movements cannot be characterized as strategic by any common usage of the term, and they certainly were not affected by any learnt bias stemming from the photographic placement of objects near the centers of scenes. The reason why TAM often looked initially to the scenes' center is because it broadly considered a large population of potential saccade vectors at the start of each search task. However, the size of this population means that time is needed to select the one eye movement to the location in the scene offering the maximum evidence for the target (the hotspot), with the amount of time required by this process ordinarily exceeding the time taken to execute an initial eye movement following search display onset. When TAM makes an eye movement before this selection process is complete, the resulting fixation is based on the averaged location from all the competing saccade vectors remaining at the time of the eye movement. Of course, early during search, and presumably scene viewing, very few of

these candidate saccade vectors would have been excluded from the target map (Figure 4, bottom), resulting in much of the target map contributing to this computation. In this eventuality TAM sends its gaze to the spatial average of this broad population, which quite often will be a location near the center of the scene (Figure 4, top). As locations are pruned from the target map over time, this population will shrink, causing the spatial mean to be less tied to the scene's center. The behavioral expression of this is the progressive movement of fixation away from the center of a scene, the exact pattern illustrated here and reported elsewhere (Tatler, 2007).

Although TAM's central fixation tendency does not result from high-level factors, arguably it may be considered strategically optimal. To the extent that there are multiple potential saccade vectors to consider, looking to the spatial average of these locations may be an optimal means of accumulating additional evidence about the target to inform the next eye movement. A similar suggestion has been made by Geisler and colleagues (Geisler et al., 2006; Najemnik & Geisler, 2005), although it is unclear whether their ideal observer model would predict a central fixation bias due to the very limited region over which averaging would be expected (when the centroid location is optimal given limitations on peripheral discrimination). Exploring the relationship between TAM's averaging behavior and ideal observer search theory will be an important direction for future work. More broadly, TAM also informs the relationship between a central fixation bias and a scene's visual features. Whereas previous work has shown that local concentrations of visual features do not correlate well with the central fixation bias (i.e., the bias exists even when the scene's center is not rich in features), TAM suggests that a more global pooling of this feature information might. Any modestly complex scene would have features distributed broadly over its entirety, and statistically the mean spatial location of these features would correspond to the center of the scene. Although the current work was specific to a search task, this underlying principle may apply to a scene inspection task as well. Indeed, upon computing a bottom-up saliency map for the 39 target absent scenes, and obtaining the spatial average for each, I found that these locations produced a qualitatively similar pattern of initial central fixations (Figure 8) as those shown for TAM and human searchers (Figure 5, top). A currently open question is whether the averaging computation underlying a central fixation bias uses raw feature or feature contrast information (as in the case of a saliency map) or evidence that is specific to a search target (as in the case of TAM). Dissociating these possibilities and revealing any task effects on the expression of central fixation tendencies will be another direction for future work.

GENERAL DISCUSSION

Two types of selections should be considered when characterizing eye movements made during the course of search; selections made in the service of recognition, and selections made in the service of action. Selection for recognition is needed to confine the recognition process to a limited set of features. Regardless of whether this is conceptualized as binding features into perceptual objects (Treisman, 1988) or routing them to a limited-capacity recognition system (Olshausen, Anderson, & Van Essen, 1993; Wolfe, 1994) the outcome is the same—a once unidentified proto-object becomes an object. In selection for action the goal is not to recognize something, but rather to do something—to select a particular action from among many potential actions. This requires binding or coupling a specific action with a specific effector, at the exclusion of all other action plans that might have recruited that effector (Allport, 1987; Neumann, 1987; Norman & Shallice, 1986; see also Schneider, 1995). While they differ in kind, these two forms of selection share the same function—just as selection is needed to avoid haphazard combinations of features describing unrecognizable objects, selection is also needed to segregate one action plan from all others,

thereby avoided disorganized movement. More generally, both serve to restrict the input to a system so as to produce a more controlled and better behavioral outcome.

Search theory has long recognized the importance of selection for recognition, but has largely ignored selections made for action, assuming that action selections follow from, and are in a sense equivalent to, recognition selections. While it is true that saccades are often directed to objects that have been selected for recognition (Deubel & Schneider, 1996; Henderson, Pollatsek, & Rayner, 1989; Hoffman & Subramaniam, 1995; Irwin & Zelinsky, 2002; Kowler, Anderson, Dosher, & Blaser, 1995; Shepherd, Findlay, & Hockey, 1986), these two forms of selection should not be treated as equivalent, as each likely imposes different demands on attention. In this paper I have argued that selections for recognition are only minimally demanding of attention, due to the preattentive prioritization of proto-objects into a sort of rank-ordered list, whereas action selections are highly demanding, due to the large population of potential saccade vectors that must be active in order to make possible just-in-time responses during a task. Moreover, to the extent that gaze *movements* are thought to be part and parcel of the search process (Motter & Holsapple, 2007; Findlay & Gilchrist, 2003), with these patterns of eye movements increasingly becoming the explanandum of many search studies, then theories of search must engage both forms of selection. TAM does this.

Under TAM the ordering of recognition selections is expressed as correlations on the target map; there is a most correlated point (the hotspot), a next most correlated point, etc., and all else being equal these correlations will determine the order in which proto-objects are recognized. However, because TAM also seeks to align its simulated fovea with each of these points in an effort to make accurate and confident recognition decisions, it must select a specific saccade vector from among the population of other scene locations to which gaze must be prepared to move. This demanding selection task is described by TAM as a pruning operation; competing saccade vectors, represented as activity distributed broadly over the target map, are progressively pruned until only the desired location is selected. Although TAM very often moves its gaze from one hotspot to the next, meaning its selections for recognition and action are aligned, this need not always be the case. TAM can also make eye movements before the hotspot has been fully selected on the target map, and this dissociation between recognition and action selection has behavioral consequences. When this happens gaze is sent to the centroid of activity remaining on the target map, resulting in the off-object fixations and the centrally biased fixations reported in this paper.

In this sense, TAM is conceptually related to other theories of oculomotor control that assume averaging over a population of neurons in a visuo-motor map (Findlay, 1987; Findlay & Walker, 1999; Godijn & Theeuwes, 2002; Trappenberg, Dorris, Munoz, & Klein, 2001; Van Opstal & Van Gisbergen, 1989; Wilimzig, Schnieder, & Schöner, 2006). Motivating these theories are observations of gaze landing either between two widely separated objects (e.g., Coren Hoenig, 1972; Findlay, 1982, 1997; Ottes, Van Gisbergen, & Eggermont, 1984, 1985), at empty locations near the centers of polygons or angles (e.g., Kaufman & Richards, 1969a; He & Kowler, 1991; Kowler & Blaser, 1995), or at the centers of letter strings (Coëffé & O'Regan, 1987; Jacobs, 1987) or words in the context of reading (Vitu, 1991, 2008). These phenomena have been reported under a variety of names, most typically center-of-gravity averaging (e.g., Kaufman & Richards, 1969b; He & Kowler, 1989; Zelinsky, Rao, Hayhoe, & Ballard, 1997) or the global effect (e.g., Findlay, 1982; Findlay & Gilchrist, 1997), but all share a common theoretical assumption—that multiple saccade vectors are considered in parallel, and that the selected saccade is often the vector average from this population of competing movement signals.

TAM extends this core theoretical principle to visual search, proposing that the same processes underlying these diverse averaging phenomena also exist during the active search of a display. In this context, off-object fixations and central fixation tendencies are just two search-specific expressions of population averaging. However, TAM advances this idea by suggesting that this averaging unfolds over time, and that this temporal dynamic might explain these different behavioral expressions.[6] When an eye movement occurs very early during search, activity over a large population will be averaged, resulting in a fixation near the center of a scene (see also Zelinsky et al., 1997). Later during search, after much of this population has been pruned, effects of averaging become more subtle, expressed now as fixations perhaps landing off objects or between objects. Finally, after search reaches a state in which recognition and action selections become more closely aligned, effects of population averaging may disappear entirely, resulting in gaze moving accurately from one search object to another. TAM therefore explains all three of these behaviors, in addition to the other search behaviors described elsewhere (Zelinsky, 2008), all within the context of a single computational model.

In conclusion, population averaging, a computation intrinsic to selection for action, is able to explain two very different eye movement behaviors occurring during search, off-object fixations and central fixation tendencies. These behaviors, when compared to the many other oculomotor behaviors accompanying search, might be easily dismissed as being relatively unimportant. Neither seems closely connected to core search processes, or clearly implicated in benchmark search patterns. However, dismissing these behaviors would be a mistake. They provide us with clues as to the different selection operations underlying search. They tell us that search, a behavior that has long been associated with serial item-by-item processing, might in fact use parallel processing and as part of a population code. TAM provides a framework for understanding this code, and for describing the temporal dynamic of selections made in the course of search.

Acknowledgments

References

Abrams RA, Meyer DE, Kornblum S. Speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system. Journal of Experimental Psychology: Human Perception and Performance. 1989; 15:529–543. [PubMed: 2527960]

Anderson RW, Keller EL, Gandhi NJ, Das S. Two-dimensional saccade-related population activity in the superior colliculus in monkey. Journal of Neurophysiology. 1998; 80:798–817. [PubMed: 9705470]

Allport, DA. Selection for action: Some behavioural and neuro-physiological considerations of attention and action. In: Heuer, H.; Sanders, AF., editors. Perspectives on perception and action. Hillsdale, NJ: Lawrence Erlbaum Associates Inc; 1987. p. 395-419.

Bonitz VS, Gordon RD. Attention to smoking-related and incongruous objects during scene viewing. Acta Psychologica. 2008; 129:255–263. [PubMed: 18804752]

Broadbent, DE. Perception and Communication. London: Pergamon Press; 1958.

---

[6]Each change in the threshold used to prune points from the target map is assumed by TAM to take one unit of time. However, the current version of TAM does not specify the mapping between this unit and saccade latencies; it may be the case that each change in the threshold consumes a constant amount of time, but more likely this relationship is nonlinear. Until this mapping is specified in future work, TAM will lack the ability to make the sorts of detailed predictions of fixation duration found in many current models of reading and scene perception (e.g., Nuthmann & Henderson, this issue).

Buswell, GT. How people look at pictures: A study of the psychology of perception in art. Chicago: University of Chicago Press; 1935.

Coëffé C, O'Regan JK. Reducing the influence of nontarget stimuli on saccadic accuracy: Predictability and latency effects. Vision Research. 1987; 27:227–240. [PubMed: 3576983]

Coren S, Hoenig P. Effect of non-target stimuli upon length of voluntary saccades. Perceptual and Motor Skills. 1972; 34:499–508. [PubMed: 5063190]

Crane HD, Steele CS. Accurate three-dimensional eye-tracker. Applied Optics. 1978; 17:691–705. [PubMed: 20197858]

Deubel H, Schneider WX. Saccade target selection and object recognition: Evidence for a common attentional mechanism. Vision Research. 1996; 36:1827–1837. [PubMed: 8759451]

Ehinger KA, Hidalgo-Sotelo B, Torralba A, Oliva A. Modelling search for people in 900 scenes: A combined source model of eye guidance. Visual Cognition. 2009; 17(6):945–978. [PubMed: 20011676]

Findlay JM. Global visual processing for saccadic eye movements. Vision Research. 1982; 22:1033–1045. [PubMed: 7135840]

Findlay JM. Visual computation and saccadic eye movements: A theoretical perspective. Spatial Vision. 1987; 2:175–189. [PubMed: 3154944]

Findlay JM. Saccade target selection in visual search. Vision Research. 1997; 37:617–631. [PubMed: 9156206]

Findlay JM, Blythe HI. Saccade target selection: Do distractors affect saccade accuracy? Vision Research. 2009; 49:1267–1274. [PubMed: 18691610]

Findlay JM, Brown V. Eye scanning of multi-element displays: II. Saccade planning. Vision Research. 2006; 46:216–227. [PubMed: 16169036]

Findlay JM, Gilchrist ID. Visual spatial scale and saccade programming. Perception. 1997; 26:1159–1167. [PubMed: 9509150]

Findlay, JM.; Gilchrist, ID. Active vision. New York: Oxford University Press; 2003.

Findlay JM, Walker R. A model of saccadic eye movement generation based on parallel processing and competitive inhibition. Behavioral and Brain Sciences. 1999; 22:661–721. [PubMed: 11301526]

Geisler WS, Chou K. Separation of low-level and high-level factors in complex tasks: Visual search. Psychological Review. 1995; 102:356–378. [PubMed: 7740093]

Geisler WS, Perry JS, Najemnik J. Visual search: The role of peripheral information measured using gaze-contingent displays. Journal of Vision. 2006; 6:858–873. [PubMed: 17083280]

Godijn R, Theeuwes J. Programming of endogenous and exogenous saccades: Evidence for a competitive integration model. Journal of Experimental Psychology: Human Perception and Performance. 2002; 28:1039–1054. [PubMed: 12421054]

Hayhoe MM, Ballard D. Eye movements in natural behavior. Trends in Cognitive Sciences. 2005; 9(4):188–194. [PubMed: 15808501]

Hayhoe MM, Shrivastava A, Mruczek R, Pelz JB. Visual memory and motor planning in a natural task. Journal of Vision. 2003; 3(1):49–63. [PubMed: 12678625]

He P, Kowler E. The role of location probability in the programming of saccades: Implications for 'center-of-gravity' tendencies. Vision Research. 1989; 29:1165–1181. [PubMed: 2617863]

He P, Kowler E. Saccadic localization of eccentric forms. Journal of the Optical Society of America A. 1991; 8:440–449.

Henderson JM. Regarding scenes. Current Directions in Psychological Science. 2007; 16:219–222.

Henderson JM, Chanceaux M, Smith TJ. The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. Journal of Vision. 2009; 9(1):1–8.

Henderson JM, Pollatsek A, Rayner K. Covert visual attention and extrafoveal information use during object identification. Perception & Psychophysics. 1989; 45:196–208. [PubMed: 2710617]

Henderson JM, Weeks P, Hollingworth A. The effects of semantic consistency on eye movements during complex scene viewing. Journal of Experimental Psychology: Human Perception and Performance. 1999; 25:210–228.

Hoffman JE, Subramaniam B. The role of visual attention in saccadic eye movements. Perception & Psychophysics. 1995; 57:787–795. [PubMed: 7651803]

Horowitz TS, Wolfe JM. Memory for rejected distractors in visual search? Visual Cognition. 2003; 10(3):257–287.

Irwin DE, Zelinsky G. Eye movements and scene perception: Memory for things observed. Perception & Psychophysics. 2002; 64:882–895. [PubMed: 12269296]

Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research. 2000; 40:1489–1506. [PubMed: 10788654]

Itti L, Koch C. Computational modeling of visual attention. Nature Reviews Neuroscience. 2001; 2(3): 194–203.

Jacobs A. On localization and saccade programming. Vision Research. 1987; 27(11):1953–1966. [PubMed: 3447349]

Kanan C, Tong MH, Zhang L, Cottrell GW. SUN: Top-down saliency using natural statistics. Visual Cognition. 2009; 17(6):979–1003. [PubMed: 21052485]

Kaufman L, Richards W. Spontaneous fixation tendencies for visual forms. Perception & Psychophysics. 1969a; 5:85–88.

Kaufman L, Richards W. "Center-of-gravity" tendencies for fixations and flow patterns. Perception & Psychophysics. 1969b; 5:81–85.

Klein RM, Farrell M. Search performance without eye movements. Perception & Psychophysics. 1989; 46:476–482. [PubMed: 2813033]

Koch C, Ullman S. Shifts in selective visual attention: Towards the underlying neural circuitry. Human Neurobiology. 1985; 4:219–227. [PubMed: 3836989]

Kowler E, Anderson E, Dosher B, Blaser E. The role of attention in the programming of saccades. Vision Research. 1995; 35:1897–1916. [PubMed: 7660596]

Kowler E, Blaser E. The accuracy and precision of saccades to small and large targets. Vision Research. 1995; 35:1741–1754. [PubMed: 7660582]

Land MF, Hayhoe M. In what ways do eye movements contribute to everyday activities? Vision Research. 2001; 41:3559–3565. [PubMed: 11718795]

Lee C, Rohrer W, Sparks D. Population coding of saccadic eye movements by neurons in the superior colliculus. Nature. 1988; 332:357–360. [PubMed: 3352733]

Mannan SK, Ruddock KH, Wooding DS. The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. Spatial Vision. 1996; 10:165–188. [PubMed: 9061830]

Mannan SK, Ruddock KH, Wooding DS. Fixation sequences made during visual examination of briefly presented 2D images. Spatial Vision. 1997; 11:157–178. [PubMed: 9428094]

McIlwain JT. Visual receptive fields and their images in superior colliculus of the cat. Journal of Neurophysiology. 1975; 38:219–230. [PubMed: 1092813]

McIlwain JT. Distributed spatial coding in the superior colliculus: A review. Visual Neuroscience. 1991; 6:3–13. [PubMed: 2025610]

Motter BC, Belky EJ. The guidance of eye movements during active visual search. Vision Research. 1998; 38:1805–1815. [PubMed: 9797959]

Motter BC, Holsapple J. Saccades and covert shifts of attention during active visual search: Spatial distributions, memory, and items per fixation. Vision Research. 2007; 47:1261–1281. [PubMed: 17418364]

Najemnik J, Geisler WS. Optimal eye movement strategies in visual search. Nature. 2005; 434:387–391. [PubMed: 15772663]

Neider MB, Zelinsky GJ. Exploring set size effects in realistic scenes: Identifying the objects of search. Visual Cognition. 2008; 16(1):1–10.

Neider MB, Zelinsky GJ. Cutting through the clutter: Searching for targets in evolving complex scenes. Journal of Vision. in press.

Neumann, O. Beyond capacity: A functional view of attention. In: Heuer, H.; Sanders, AF., editors. Perspectives on perception and action. Hillsdale, NJ: Lawrence Erlbaum Associates Inc; 1987. p. 361-394.

Norman, D.; Shallice, T. Attention to action: Willed and automatic control of behaviour. In: Davison, R.; Shwartz, G.; Shapiro, D., editors. Consciousness and self regulation: Advances in research and theory. New York: Plenum; 1986. p. 1-18.

Nuthmann A, Henderson JM. Visual Cognition. XXXX

Olshausen B, Anderson C, Van Essen D. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. Journal of Neuroscience. 1993; 13:4700–4719. [PubMed: 8229193]

Ottes FP, Van Gisbergen JAM, Eggermont JJ. Metrics of saccade responses to visual double stimuli: Two different modes. Vision Research. 1984; 24(9):1169–1179. [PubMed: 6523740]

Ottes FP, Van Gisbergen JAM, Eggermont JJ. Latency dependence of colour-based target vs nontarget discrimination by the saccadic system. Vision Research. 1985; 25:849–862. [PubMed: 4024483]

Parkhurst DJ, Law K, Niebur E. Modeling the role of salience in the allocation of overt visual selective attention. Vision Research. 2002; 42:107–123. [PubMed: 11804636]

Pashler, H. The Psychology of Attention. Cambridge, MA: MIT Press; 1998.

Van Opstal AJ, Van Gisbergen JAM. A nonlinear model for collicular spatial interactions underlying the metrical properties of electrically elicited saccades. Biological Cybernetics. 1989; 60:171–183. [PubMed: 2923922]

Renninger LW, Vergheese P, Coughlan J. Where to look next? Eye movements reduce local uncertainty. Journal of Vision. 2007; 7(3):1–17.

Rensink RA. The dynamic representation of scenes. Visual Cognition. 2000; 7:17–42.

Schad DJ, Engbert R. The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. Visual Cognition. XXXX

Schneider WX. VAM: A neuro-cognitive model for visual attention control of segmentation, object recognition, and space based motor action. Visual Cognition. 1995; 2:331–376.

Sheinberg DL, Logothetis NK. Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. Journal of Neuroscience. 2001; 21:1340–1350. [PubMed: 11160405]

Shepherd M, Findlay J, Hockey R. The relationship between eye movements and spatial attention. Quarterly Journal of Experimental Psychology. 1986; 38A:475–491. [PubMed: 3763952]

Sparks, DL.; Kristan, WB.; Shaw, BK. The role of population coding in the control of movement. In: Stein; Grillner; Selverston; Stuart, editors. Neurons, Networks, and Motor Behavior. MIT Press; 1997. p. 21-32.

Tatler BW. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. Journal of Vision. 2007; 7(14):1– 17. [PubMed: 18217799]

Tatler BW, Baddeley RJ, Gilchrist ID. Visual correlates of fixation selection: Effects of scale and time. Vision Research. 2005; 45:643–659. [PubMed: 15621181]

Tatler BW, Baddeley RJ, Vincent BT. The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. Vision Research. 2006; 46:1857–1862. [PubMed: 16469349]

Tatler BW, Vincent BT. The prominence of behavioural biases in eye guidance. Visual Cognition. 2009; 17:1029–1054.

Torralba A, Oliva A, Castelhano M, Henderson JM. Contextual guidance of attention in natural scenes: The role of global features on object search. Psychological Review. 2006; 113:766–786. [PubMed: 17014302]

Trappenberg TP, Dorris MC, Munoz DP, Klein RM. A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. Journal of Cognitive Neuroscience. 2001; 13:256–271. [PubMed: 11244550]

Treisman AM. Features and objects: The Fourteenth Bartlett Memorial Lecture. Quarterly Journal of Experimental Psychology. 1988; 40A:201–237. [PubMed: 3406448]

Treisman AM, Gelade G. A feature integration theory of attention. Cognitive Psychology. 1980; 12:97–136. [PubMed: 7351125]

Treisman AM, Gormican S. Feature analysis in early vision: Evidence from search asymmetries. Psychological Review. 1988; 95:14–48.

Tseng PH, Carmi R, Cameron I, Munoz DP, Itti L. Quantifying center bias of observers in free viewing of dynamic natural scenes. Journal of Vision. 2009; 9(7):1–16.

Vitu F. The existence of a center of gravity effect during reading. Vision Research. 1991; 31:1289–1313. [PubMed: 1891819]

Vitu F. About the global effect and the critical role of retinal eccentricity: Implications for eye movements in reading. Journal of Eye Movement Research. 2008; 2(3):1–18.

Ward R, Duncan J, Shapiro K. The slow time-course of visual attention. Cognitive Psychology. 1996; 30:79–109. [PubMed: 8660782]

Wilimzig C, Schnieder S, Schöner G. The time course of saccadic decision making: Dynamic field theory. Neural Networks. 2006; 19:1059–1074. [PubMed: 16942860]

Williams DE, Reingold EM. Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. Psychonomic Bulletin & Review. 2001; 8:476–488. [PubMed: 11700898]

Wischnewski M, Belardinelli A, Schneider WX, Steil JJ. Where to look next? Combining static and dynamic proto-objects in a TVA-based model of visual attention. Cognitive Computation. 2010; 2:326–343.

Wolfe JM. Guided search 2.0: A revised model of visual search. Psychonomic Bulletin and Review. 1994; 1:202–238.

Wolfe JM. What can 1 million trials tell us about visual search? Psychological Science. 1998; 9:33–39.

Wolfe JM, Cave K, Franzel S. Guided search: An alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human Perception and Performance. 1989; 15:419–433. [PubMed: 2527952]

Wolfe, JM.; Gancarz, G. Guided Search 3.0. In: Lakshminarayanan, V., editor. Basic and clinical applications of vision science. Dordrecht: Kluwer; 1996. p. 189-192.

Zelinsky GJ. Using eye saccades to assess the selectivity of search movements. Vision Research. 1996; 36:2177–2187. [PubMed: 8776484]

Zelinsky GJ. A theory of eye movements during target acquisition. Psychological Review. 2008; 115(4):787–835. [PubMed: 18954205]

Zelinsky GJ, Rao R, Hayhoe M, Ballard D. Eye movements reveal the spatio-temporal dynamics of visual search. Psychological Science. 1997; 8:448–453.

Zelinsky GJ, Sheinberg D. Eye movements during parallel–serial visual search. Journal of Experimental Psychology: Human Perception and Performance. 1997; 23:244–262. [PubMed: 9090154]

Zelinsky GJ, Todor A. The role of "rescue saccades" in tracking objects through occlusions. Journal of Vision. 2010; 10(14):29, 1–13. [PubMed: 21191133]

Zelinsky, GJ.; Zhang, W.; Yu, B.; Chen, X.; Samaras, D. The role of top-down and bottom-up processes in guiding eye movements during visual search. In: Weiss, Y.; Scholkopf, B.; Platt, J., editors. Advances in Neural Information Processing Systems. Vol. 18. Cambridge, MA: MIT Press; 2006. p. 1609-1616.

Zingale C, Kowler E. Planning sequences of saccades. Vision Research. 1987; 27:1327–1341. [PubMed: 3424681]
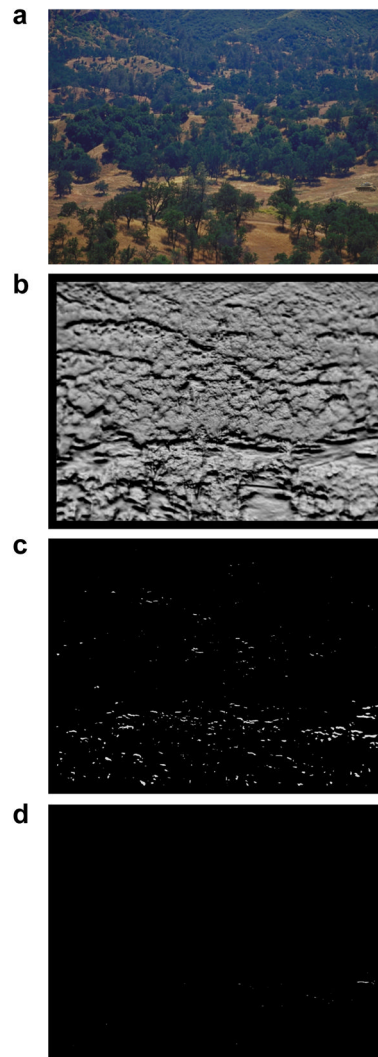
Figure 1.
A search scene (a, target in middle right) and three target maps generated by TAM (b–d). (b) Early during target selection the population of potential saccade vectors covers most of the scene. (c) An intermediate stage in saccade target selection after vectors to unlikely target locations have been pruned from the target map. (d) A late stage in saccade target selection showing only vectors to the target on the target map. Figure adapted from Zelinsky (2008).
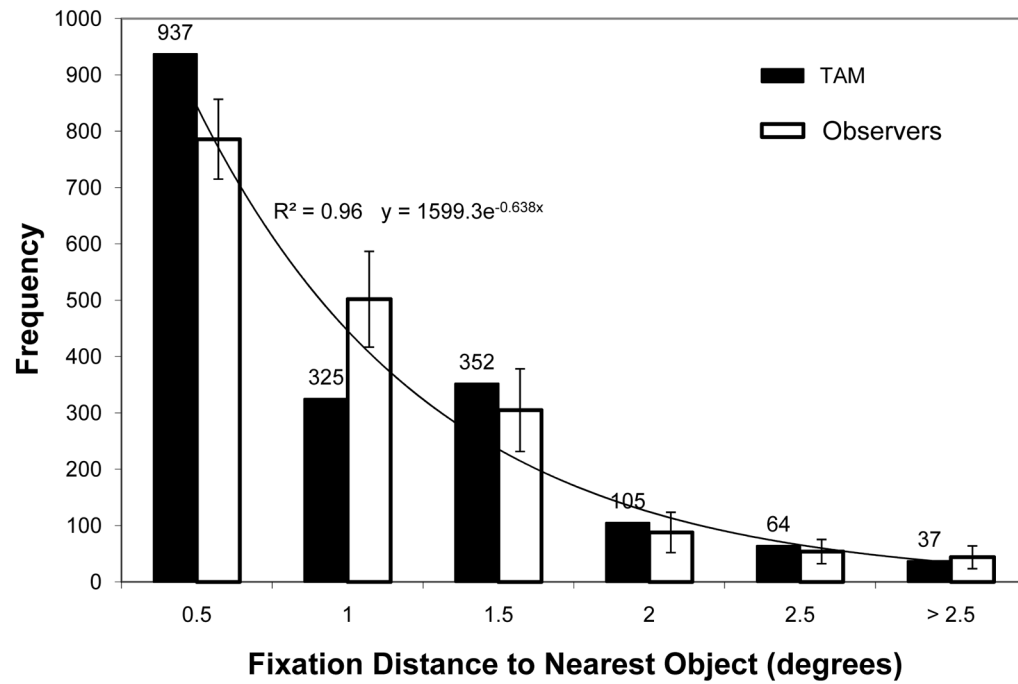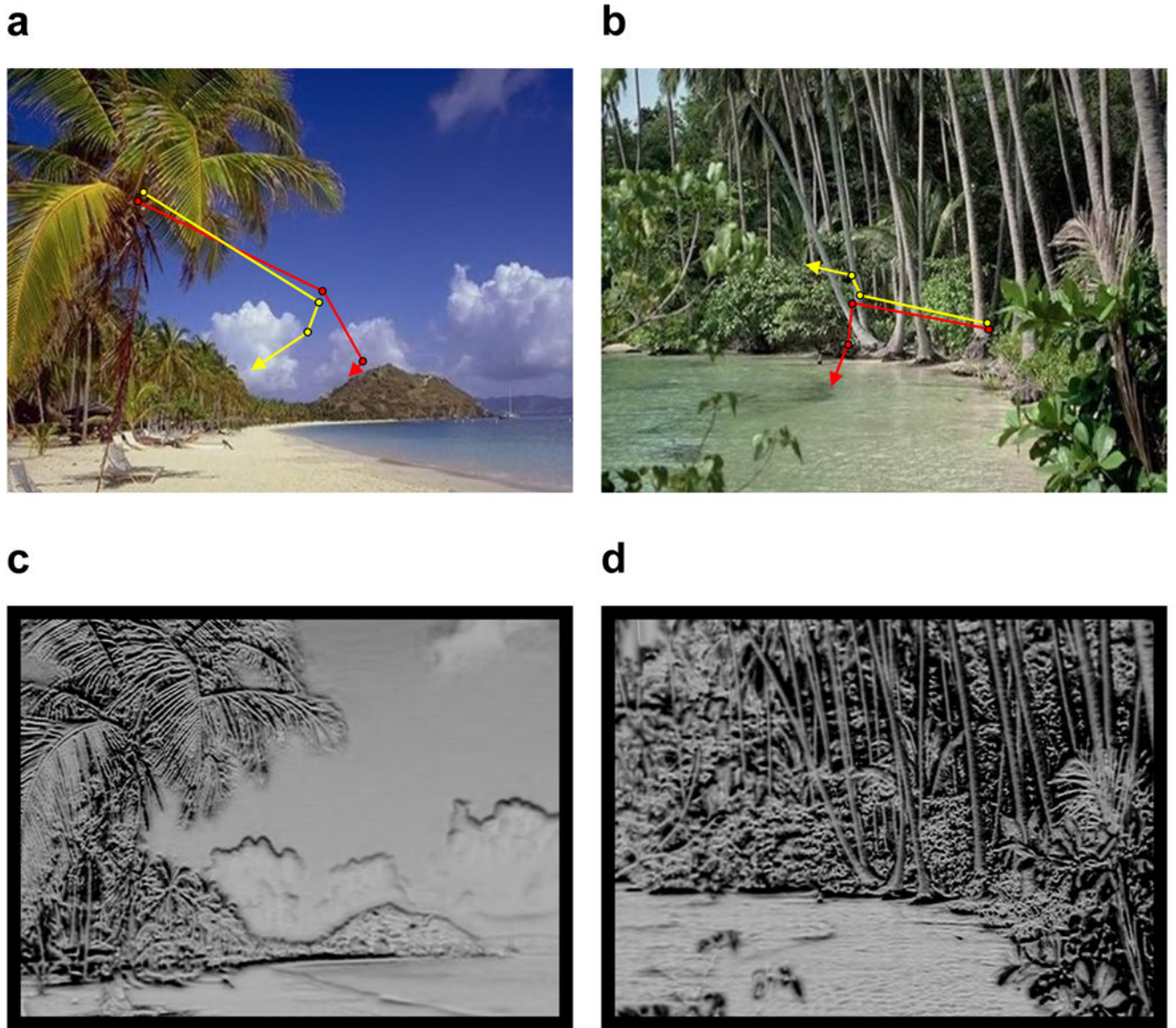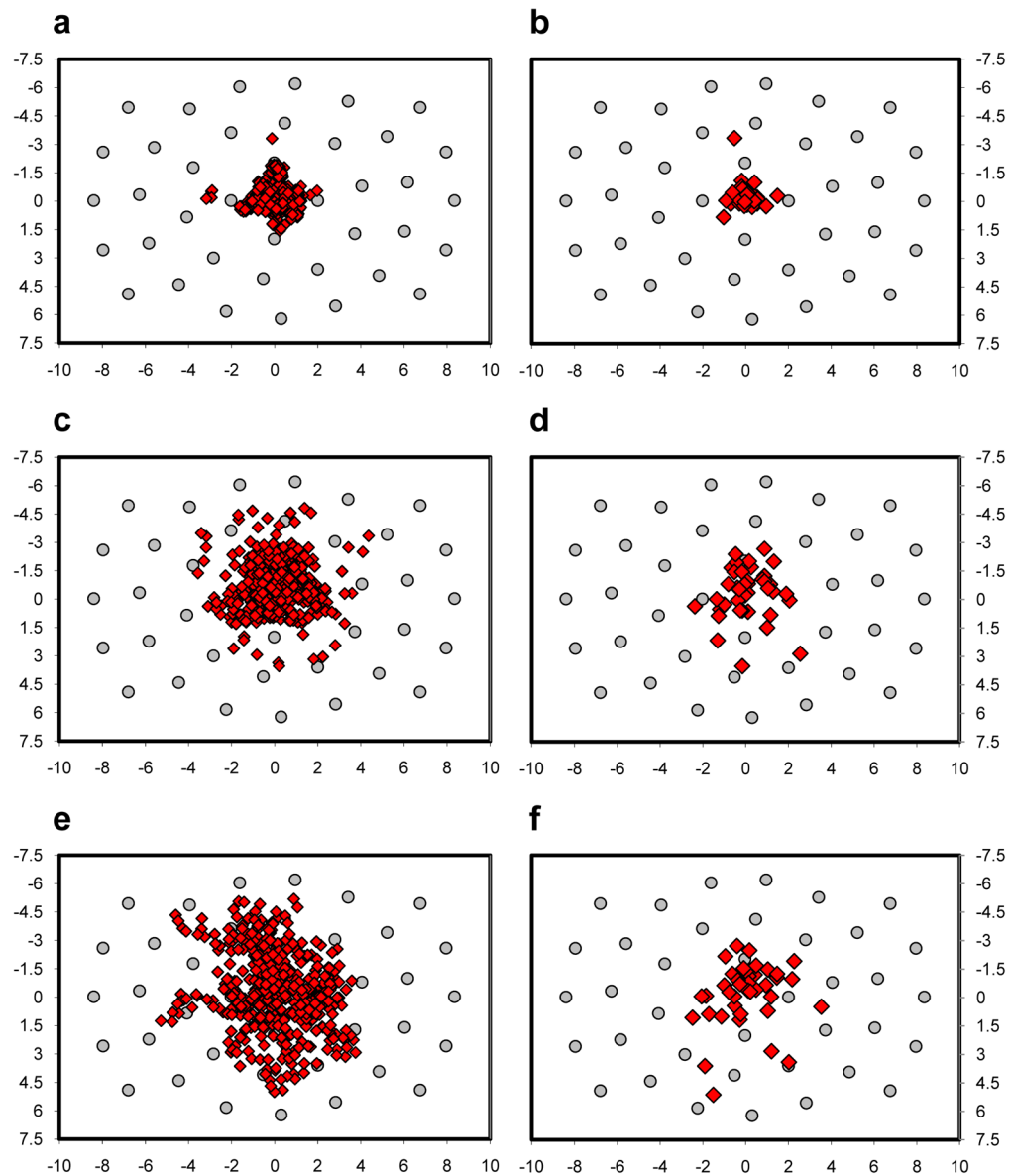
Figure 2.
A histogram of distances between each fixation made during search and the nearest search object, for both human observers (white) and TAM (black). Error bars indicate the 95% confidence interval with respect to the behavioral mean.
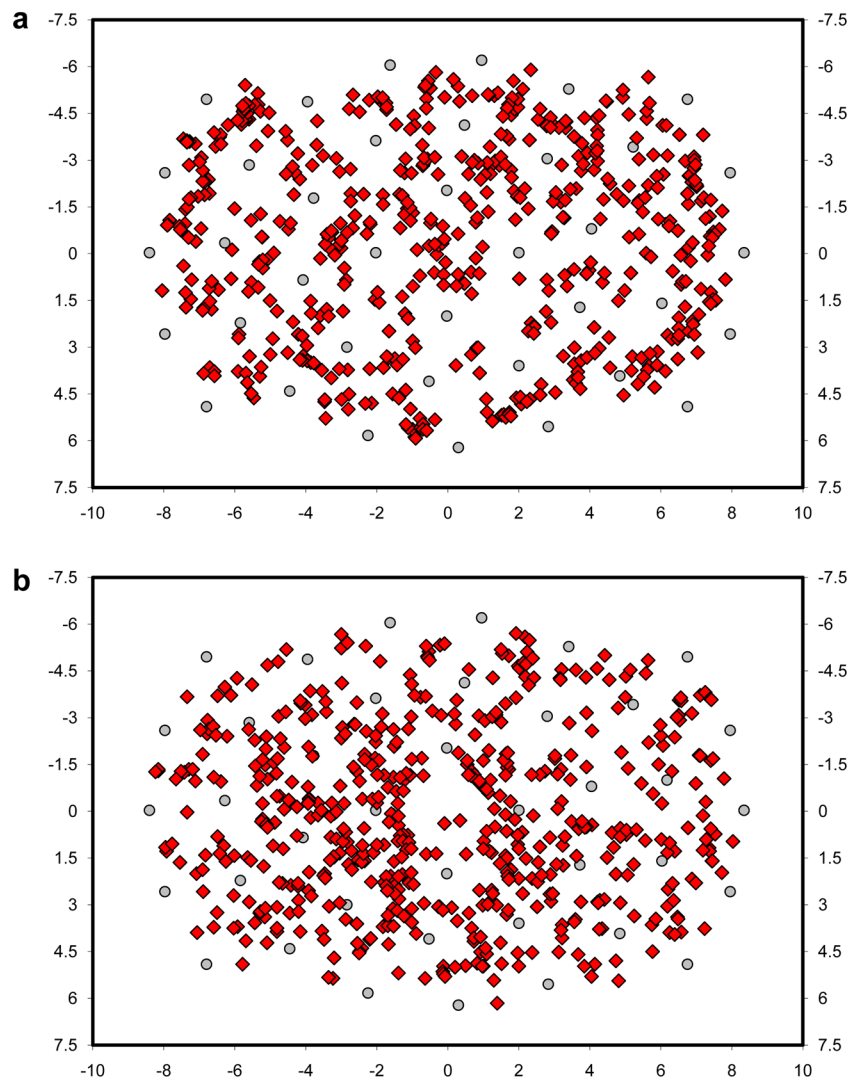
Figure 3.
Scatterplots showing search fixations greater than 1.0° from an object for (a) a representative observer, and (b) TAM. Possible object locations are indicated by the gray circle markers, and axes are in degrees of visual angle relative to center. Note that fixations appearing closer than 1.0° from an object marker are not errors; these fixations are from trials in which no object occupied that location.
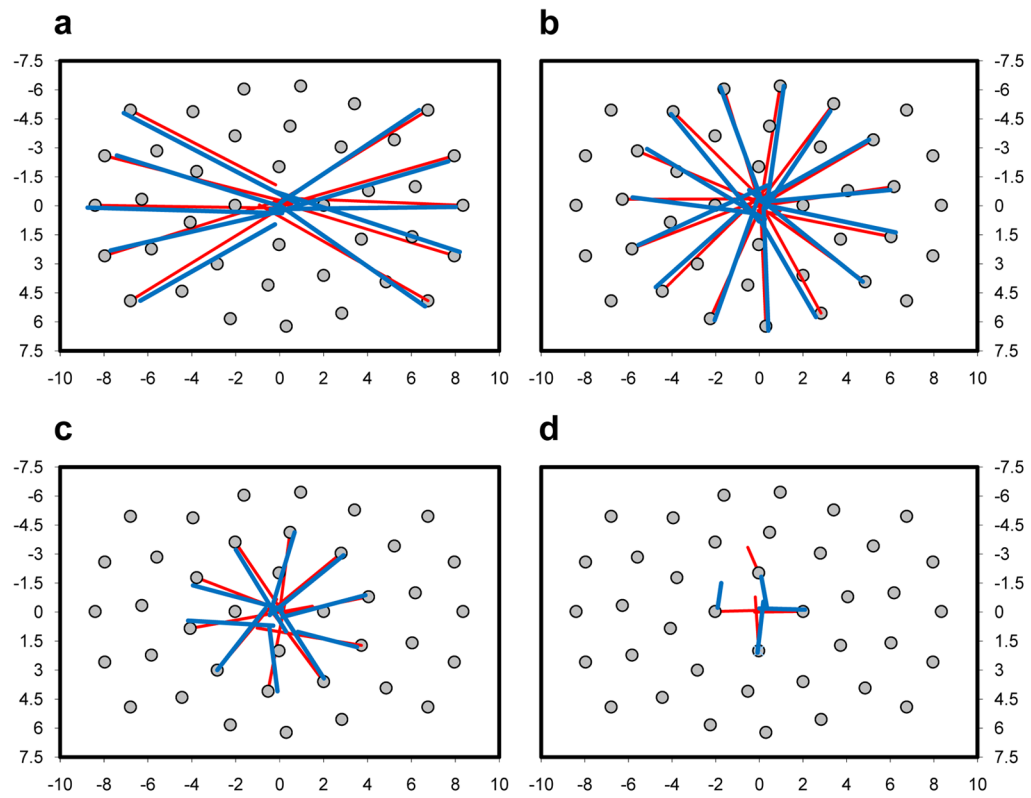
Figure 4.
(a, b) Representative scenes used in Experiment 2. The first three eye movements during search are shown for TAM (red) and an observer (yellow). (c, d) Corresponding target maps from TAM, each based on the starting gaze position.
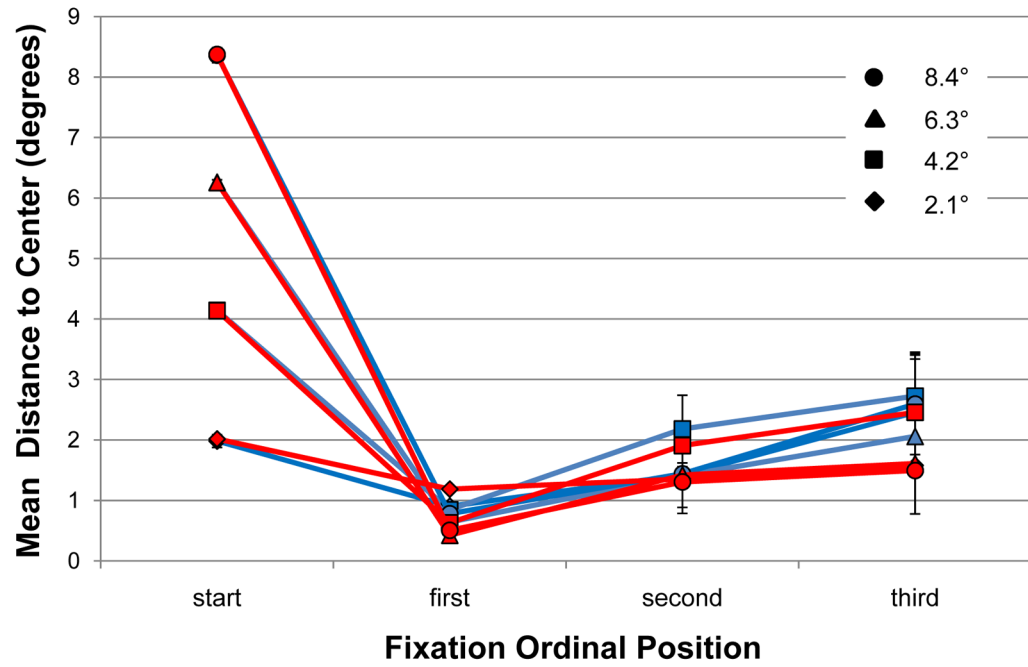
Figure 5.
Scatterplots of fixations following search scene onset for observers (left) and TAM (right). Grey circle markers indicate starting gaze positions, and axes are in degrees of visual angle relative to center. (a, b) First fixations. (c, d) Second fixations. (e, f) Third fixations.
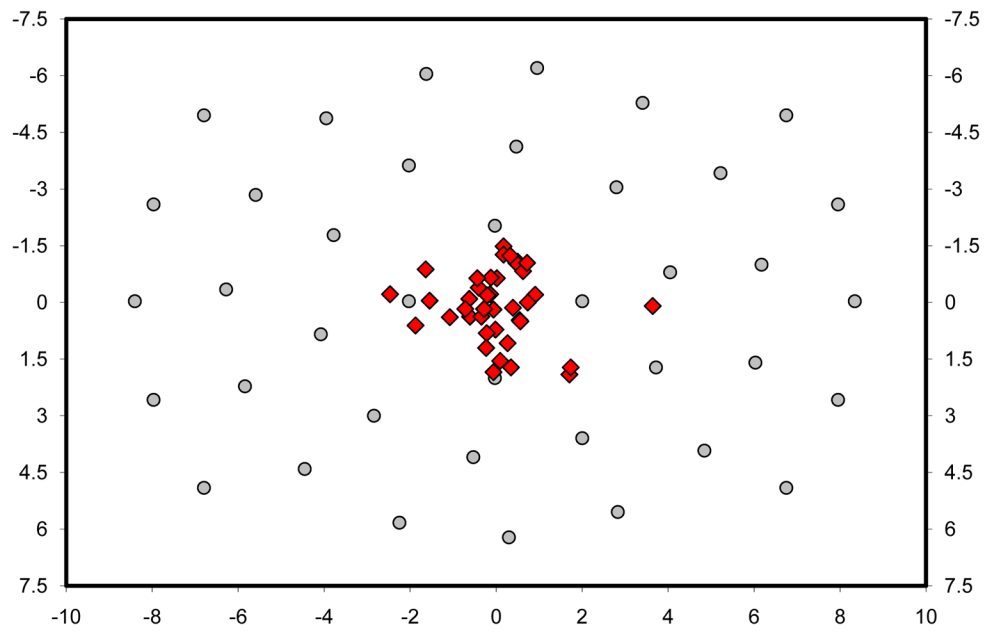
Figure 6.
Initial saccades originating from each of the 39 starting gaze positions (grey circle markers) for one observer (blue lines) and TAM (red lines). (a) Saccades from 8.4° eccentric locations, relative to center. (b) Saccades from 6.3° locations. (c) Saccades from 4.2° locations. (d) Saccades from 2.1° locations.

Figure 7.
Mean distance in degrees between fixation and the center of the scenes as a function of
fixation ordinal position and starting gaze position for observers (blue) and TAM (red).
Error bars indicate 95% confidence intervals.

Figure 8.
Scatterplot of positions obtained by finding the weighted centroid of points on a saliency map computed for each of the 39 target absent scenes used in Experiment 2. Axes are in degrees of visual angle relative to center.