



# A target contrast signal theory of parallel processing in goal-directed search

Alejandro Lleras<sup>1</sup> · Zhiyuan Wang<sup>1</sup> · Gavin Jun Peng Ng<sup>1</sup> · Kirk Ballew<sup>1</sup> · Jing Xu<sup>1</sup> · Simona Buetti<sup>1</sup>

Published online: 5 February 2020

© The Psychonomic Society, Inc. 2020, corrected publication 2020

## Abstract

Feature Integration Theory (FIT) set out the groundwork for much of the work in visual cognition since its publication. One of the most important legacies of this theory has been the emphasis on feature-specific processing. Nowadays, visual features are thought of as a sort of currency of visual attention (e.g., features can be attended, processing of attended features is enhanced), and attended features are thought to guide attention towards likely targets in a scene. Here we propose an alternative theory – the Target Contrast Signal Theory – based on the idea that when we search for a specific target, it is not the target-specific features that guide our attention towards the target; rather, what determines behavior is the result of an active comparison between the target template in mind and every element present in the scene. This comparison occurs in parallel and is aimed at rejecting from consideration items that peripheral vision can confidently reject as being non-targets. The speed at which each item is evaluated is determined by the overall contrast between that item and the target template. We present computational simulations to demonstrate the workings of the theory as well as eye-movement data that support core predictions of the theory. The theory is discussed in the context of FIT and other important theories of visual search.

**Keywords** Visual search · Peripheral vision · Parallel processing · Functional viewing field · Efficient search

## Introduction

The goal of most theories of visual attention and visual search is to predict how attention will move in a scene as a function of the elements in it. Analyses of the search slope (i.e., the slope coefficient of the linear regression of reaction time (RT) as a function of the number of elements) can provide key insights into our understanding of visual attention. Pioneering work by Treisman undoubtedly promoted the importance of this measure. Indeed, in Feature Integration Theory (FIT; Treisman & Gelade, 1980), search slopes were used as the critical measure to distinguish between two fundamentally different modes of information processing that guide attention in a scene. The core idea of the theory is that feature search (i.e., when the target differs from

distractors along a unique visual feature) could be performed in parallel and, as a result, search times should be unaffected by the number of distractors. Indeed, in feature search, the presence of the target would be indexed by a single point of activity in the “feature map” coding for that unique feature, and as such, this activation could be detected in parallel, regardless of the amount of activity on other feature maps. Feature search was compared to conjunction search, where the target shares at least one feature with every other distractor in the display. As a result, all feature maps would register more than one point of activation, making parallel detection of the target impossible. Treisman and Gelade (1980) proposed that attention would then be required to visit separate spatial locations to bind feature values across different feature maps into cohered object representations that could then be compared to the target template. This process naturally takes longer with more items in the display, leading to a steep dependence of RT on set size. Treisman and Sato (1990) later improved on this first model to account for efficient search performance in certain conjunction search tasks, by adding a mechanism of feature-based inhibition that can reduce the activity in two or more feature maps to decrease the activity in certain distractor locations.

In a similar vein, Duncan and Humphreys (1989) proposed the concept of a “search surface.” The idea of a search surface

---

The original version of this article was revised: During production of the article, Figure 4 was incorrectly used twice in the initial article, so it appeared both as Figure 4 and Figure 5 in the article. The original Figure 5 therefore was left out of the article.

---

✉ Alejandro Lleras  
alejandrolleras@gmail.com

<sup>1</sup> Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St, Champaign, IL 61820, USA

is meant to reflect the fact that search slopes seem to be determined by two sets of similarity relations among search items: The search slope is an increasing function of target-distractor similarity but a decreasing function of distractor-distractor similarity. In other words, the larger the target-distractor similarity, the larger the search slope; the larger the distractor-distractor similarity, the smaller the search slope. Wolfe's Guided Search (1994) model was perhaps an early culmination of the efforts in systematically understanding search slope variations. In this context, search slopes are a function of the salience of the target with respect to the distractors. As salience decreases, search slopes increase until search eventually becomes serial. The most important contribution of Guided Search was the proposal that a top-down signal could boost the processing of specific feature maps, increasing the weight of that map on the overall activation map.

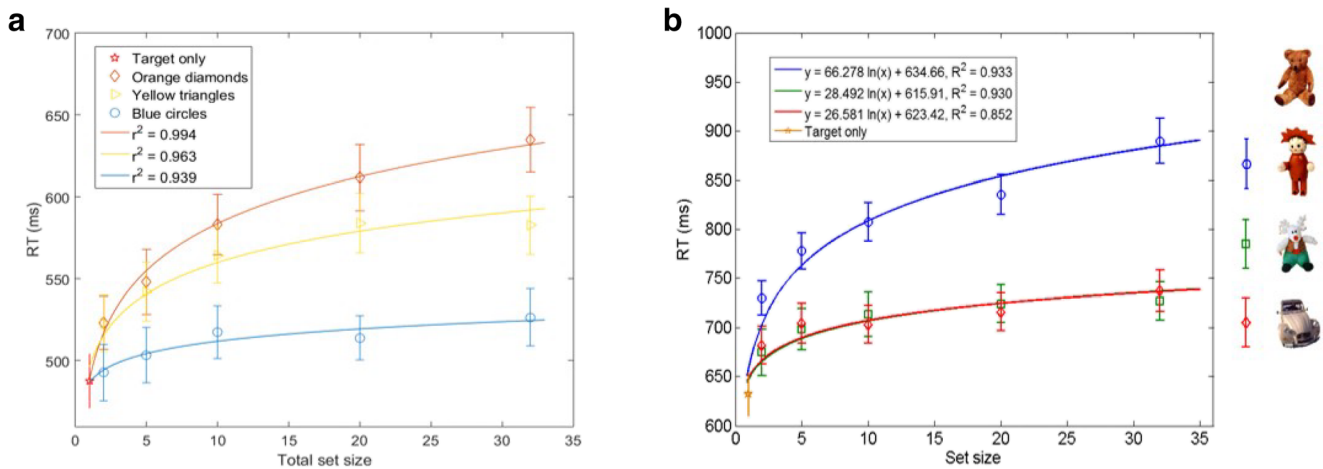
### The importance of parallel processing in visual search

Whereas the models reviewed above focused on understanding serial and/or capacity-limited attention mechanisms, there are two traditions in the visual search literature that placed a major emphasis on the parallel analysis of the scene. First, Signal-Detection Theories (SDTs) of visual search propose that the entire scene is processed and result in a representation of the distribution of distractor observations, as well as a distribution of target observations (e.g., Eckstein, Thomas, Palmer, & Shimozaki, 2000; Palmer, Verghese, & Pavel, 2000; Verghese, 2001). The goal of SDT models is to *predict search accuracy* and not reaction times. Rosenholtz (2001) improved the performance of this family of models: instead of using the entire distribution of distractors, the model performed substantially better when using only summary information about the distractors (average and variance), supporting the idea that peripheral vision is limited in its ability to represent all peripheral items. More directly related to the Target Contrast Signal Theory presented below, Rosenholtz demonstrated that similar performance could also be obtained when peripheral analysis compared display items to one another and, in some proportion of the time, display items to a target template (Relative-Coding-With-Reference-Model). Importantly, SDT theories of visual search demonstrated that a parallel processing mechanism can account for a wide range of visual search results without invoking feature-binding spatial attention. A second tradition consists of theories aimed at predicting eye movements during search or scene viewing. These theories rely entirely on a parallel analysis of the scene to determine where the eyes will move next. We discuss these theories in more detail in the *General discussion*.

Another model that strongly relies on parallel processing is the Texture Tiling Model of Rosenholtz and colleagues (e.g.,

Balas, Nakano & Rosenholtz, 2009; Rosenholtz, Huang, Raj, Balas & Ilie, 2012). This model is a low-level model of peripheral vision that has demonstrated great success at predicting performance across a number of inefficient search tasks (i.e., tasks that produce search slopes of more than 10 ms/item) and can account for a number of well-known visual search phenomena, such as search asymmetries and conjunction search inefficiency. The model assumes a rich set of low-level summary statistics is computed in parallel over pooling regions (that increase in size with eccentricity) across the scene. The statistics are supposed to represent computations that V1-like cells can make over the pooling regions (Freeman & Simoncelli, 2011). The central idea is that search performance must be constrained by the sort of information peripheral vision *can* represent about the objects in the field of view. The authors argue that in crowded displays (when multiple items fall within the same pooling region), visual search does not operate at the object or item level, but rather at the pooled-region level. As a result, search efficiency is determined not by the speed attention can move around the scene and discriminate individual objects, but rather by the ability of the visual system to distinguish pooled representations that contain target features (that have been pooled together with distractor features) from pooled representations that summarize regions containing only distractors. This work has highlighted the importance of considering peripheral processing limitations when trying to understand visual search performance.

Finally, recent work has focused on understanding the systematic variation of RT as a function of set size associated with parallel processing during search (Buetti, Cronin, Madison, Wang, & Lleras, 2016; Wang, Buetti, & Lleras, 2017). Most theories of visual attention include an initial parallel processing of the scene that creates a map, referred to as a saliency map (e.g., Itti & Koch, 2000), activation map (Wolfe, 1994), or a target-distractor similarity map (Adeli, Vitu & Zelinsky, 2017; Zelinsky, 2008). These theories assume that the time to compute this map is negligible or constant, much like signal-detection theories of search (e.g., Palmer et al., 2000; Rosenholtz, 2001). Buetti et al. (2016) took a closer look at the so-called “flat search functions” observed in efficient search and discovered that they were not really flat: In addition to the linear search function typically observed in serial/inefficient search, Buetti et al. (2016) reported that efficient search with a fixed target (e.g., search for a red triangle among blue circles), was best described by a logarithmic RT by set size function rather than by a linear one. Furthermore, the steepness of the logarithmic functions was associated with target-distractor similarity: the more similar the distractors, the steeper the logarithmic functions (e.g., a red triangle among orange diamonds; see Fig. 1, left panel). Note that logarithmic and negatively accelerated search functions had already been reported in multiple studies but were never systematically analyzed (e.g., Briggs & Swanson, 1970;



**Fig. 1** Fixed-target efficient search reaction times (RTs) increase as a logarithmic function of set size. (A) Replotted data from Buetti et al. (2016; Experiment 1A), search for simple colored geometric stimuli. (B) Data from Wang et al. (2017; Experiment 1), search for real-world

Carrasco, Evert, Chang, & Katz, 1995; Corballis, Kirby & Miller, 1972; Kristofferson, 1972; Palmer, Ames, & Lindsey, 1993; Simpson, 1972; Swanson & Briggs, 1969; Treisman & Gelade, 1980). Wang et al. (2017) extended these findings on simple geometric shapes to real-world objects (Fig. 1, right panel).

Importantly, the logarithmic dependency of RT on set size constrains the possible functional architectures underlying visual search, and, in particular, efficient search. Townsend and Ashby (1983) conducted the first systematic investigation of various possible cognitive architectures, distinguishing between properties such as parallel versus serial processing, unlimited versus limited capacity, and exhaustive versus self-terminating completion rules. Buetti et al. (2016) used this framework to propose that the observed logarithmic functions, and modulations of their steepness, were indicative of a processing architecture with parallel processing, unlimited capacity, and an exhaustive termination rule. Buetti et al.'s (2016) simulations also demonstrated that one can easily account for the modulation of the steepness of the logarithmic functions by assuming that the processing time of an individual item is inversely proportional to its dissimilarity to the target.

In the context of these findings, Buetti et al. (2016) proposed that visual processing in visual search evolved over two stages. During the first stage, the visual system accumulates visual information in parallel about each item in the display. Peripheral vision imposes a limit on the resolution and information that the system can acquire in this fashion (e.g., Neider & Zelinsky, 2010; Rosenholtz, 2016; Strasburger, Rentschler, & Jüttner, 2011; Wang, Lleras, & Buetti, 2018). When distractors are sufficiently visually different from the target (referred to as *lures*), the visual system relies on peripheral vision to decide with high confidence that the item is unlikely

to be a target. When distractors are too similar to the target (referred to as *candidates*), the visual system lacks the processing resolution to be able to confidently rule out the possibility that those distractors are not the target. The amount of information necessary for this rejection decision should thus reflect the item's similarity to the target template. As a result, during this first stage of processing, a rejection decision can be made through a relatively coarse, rather than fine-grained, discrimination process. For highly dissimilar items, little information would be required to reject them. However, if a decision cannot be made when the maximal amount of information is reached, the item would then be labeled as a candidate and queued for processing in the second, capacity-limited stage. Therefore, in the context of this model, the accumulation threshold that separated the lure and candidate categories was essentially proposed to arise from the processing limitations in peripheral vision (e.g., Freeman & Simoncelli, 2011; Levi, 2008; Rosenholtz et al., 2012; Strasburger et al., 2011). As a result, candidate items were proposed to require a closer *scrutiny* to be distinguished from the target. This attentive scrutiny would likely require eye movements and/or focused attention to each of these items, until the target is found, producing a steep linear cost to RT. Buetti et al.'s (2016, Experiments 2, 3A–D) findings confirmed that candidates contribute linearly to RTs even in the presence of lures and regardless of the lure-target similarity.

## Mathematical model and first simulations in Buetti et al. (2016)

The architecture proposed by Buetti et al. (2016) was implemented in simulations using equations that govern the

behavior of individual stochastic evidence accumulators (Wang et al., 2017). In the formal model, during stage-one processing, the visual system tries to reject each search item by reaching a decision that it is unlikely to be the target. The model is a drift-diffusion model, where evidence is accumulated in parallel at each location, with unlimited capacity (i.e., the evidence accumulation rate for each item is independent of set size). The amount of information required to reach a decision is proportional to the item's similarity to the target. The greater the target-item similarity, the more information is needed. To account for the resolution limitations in peripheral vision, a maximum decision threshold was introduced. Items that reach this maximum threshold are too similar to the target to be rejected as lures; their location is thus passed on to the second stage of search during which focused attention (or eye movements) is directed to these locations to make a more confident decision regarding these high-similarity items. Information accumulation was modeled as a Gaussian random walk with a positive mean drift rate. Thus, the completion time for each item's processing follows the Inverse Gaussian distribution. Finally, the model had an exhaustive termination rule. Thus, the stage-one completion time for a specific search display was determined by the time taken by the last accumulator to reach threshold.

This initial model allowed us to simulate stage-one processing time as a function of set size and lure-target similarity, and to reproduce all of the observed characteristics in the search functions reported in Buetti et al. (2016). Interestingly, our simulations revealed that the accumulation thresholds, which represent each item's similarity to the target, were proportional to the logarithmic slopes of simulated search functions. This relation supports the assumption that accumulation thresholds are determined by and represent target-lure similarity and is consistent with the findings that logarithmic slopes increase with increasing lure-target similarity.

One successful application of this initial model was a demonstration that the coefficient of the slope of the logarithmic function meaningfully predicts future performance, even in novel search scenarios. That is, this coefficient can be used to predict how long it will take to find a target in novel scenes that contain multiple types of lures (Lleras, Wang, Madison, & Buetti, 2019; Wang et al., 2017). For instance, one can first measure how long it takes for observers to find a target in homogeneous displays that contain all identical lures (e.g., find a red triangle amongst blue circles and a red target amongst yellow triangles). The parameters estimated from those tasks can then be used to predict how long it will take to find a red triangle in a *heterogeneous* display (i.e., a display that contains a mix of blue circles and yellow triangles). Wang et al. (2017) used this approach to determine the best-

approximating “global” equation to predict simulated processing times in heterogeneous displays based on simulated RTs in homogeneous displays. This equation is referred to as “global” because it summarizes the behavior of an entire set of independent 1-D Brownian noise accumulators, which are not directly observable in search behavior, whereas the global equation is. One advantage of a global equation is that once one is identified, there is no longer a need to run simulations of a set of accumulators to understand the behavior of the set. The global equation can then be used to predict behavior under any behavioral experiment that falls within the umbrella covered by the simulation (in this case, predicting heterogeneous display RTs based on homogeneous display RTs).

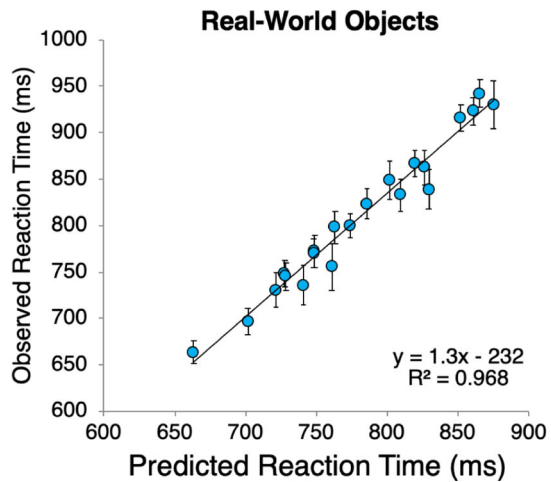
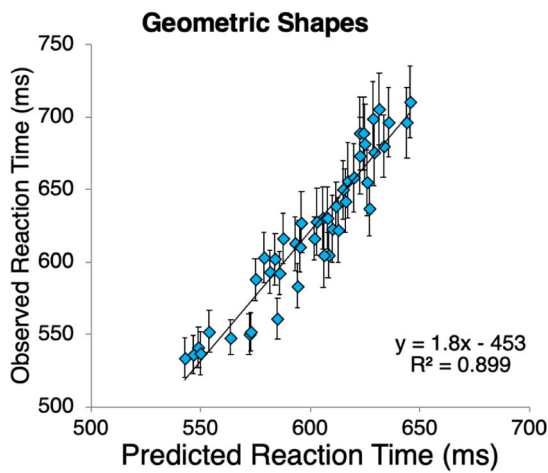
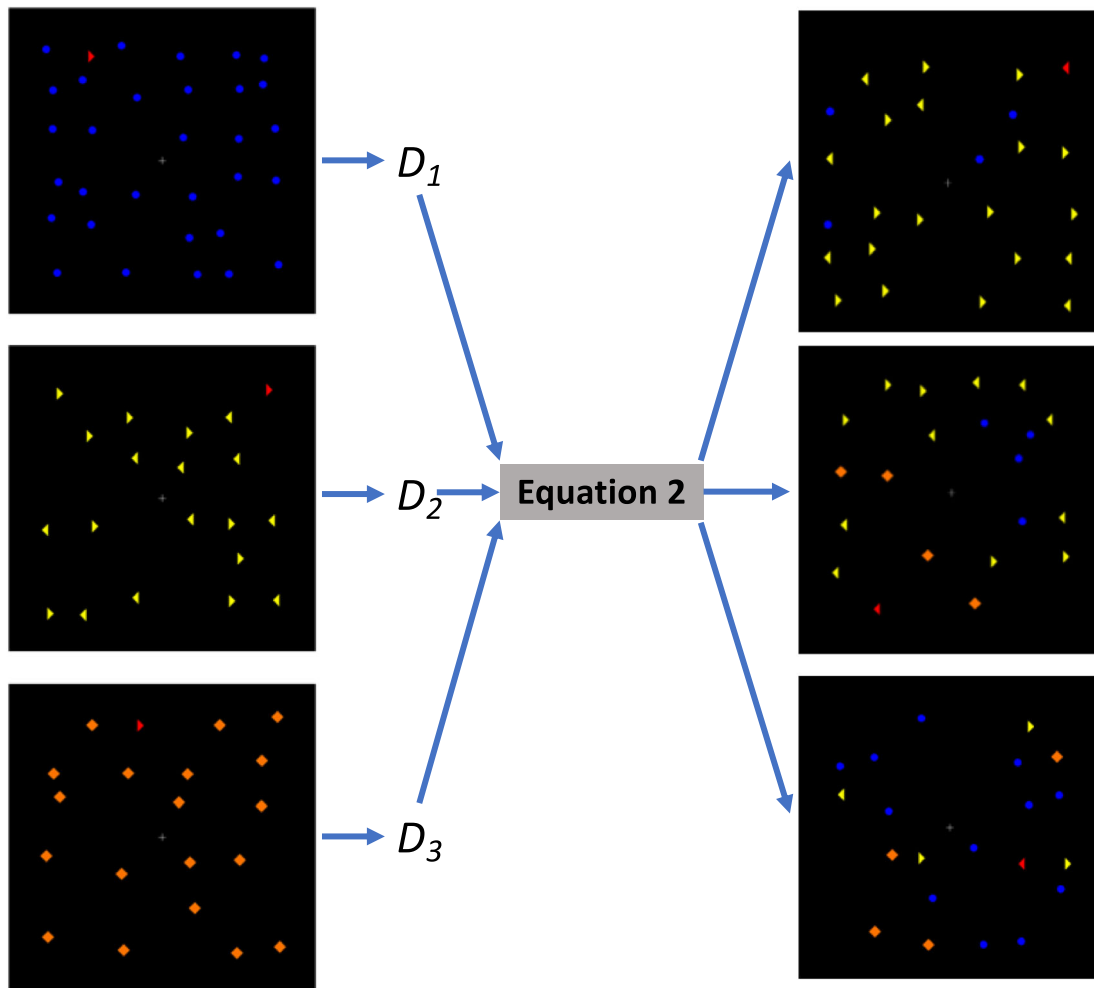
Equation 1, presented below, was the winning equation among the four tested ones. In Equation 1,  $D_j$  indicates the logarithmic slope parameters associated with lures of type  $j$  and are organized from smallest  $D_1$  to largest  $D_L$  (with  $D_0 = 0$ ). This equation assumed an architecture with unlimited capacity, parallel processing, and an exhaustive stopping rule.

$$RT = a + \sum_{j=1}^L (D_j - D_{j-1}) * \ln \left( N_T - \left( \sum_{i=1}^{j-1} N_i \right) * I_{[2, \infty)}(j) + 1 \right) \quad (1)$$

$L$  indicates the number of lure types present in the display,  $N_T$  the total number of lures,  $N_i$  the number of lures of type  $i$ . The constant  $a$  represents the reaction time when the target is alone in the display. The index function  $I_{[2, \infty)}(j)$  indicates that the sum over  $N_i$  only applies when there are at least two different types of lures in the display ( $j > 1$ ). When  $j = 1$ , the second sum is zero.

Next, the same approach was used to predict search times in human participants. Specifically, we first estimated the slope coefficients when searching in homogeneous displays in a group of subjects. We then tested whether Equation 1 could predict observed search times in heterogeneous displays based on the parameters observed in homogeneous displays. Critically, a separate group of participants completed the search in heterogeneous displays. The results indicated that Equation 1 accounted for 96.8% of RT variance when pictures of real-world objects were used as stimuli (Wang et al., 2017) and 89.9% when simple geometric figures were used (Lleras et al., 2019).

Finally, Lleras et al. (2019) also demonstrated that this modeling approach can be used to quantitatively measure the extent of inter-item interactions in a search scene (which tend to facilitate search in homogeneous compared to heterogeneous search scenes). These inter-item facilitations were stronger in displays containing simpler (geometric shapes) compared to more complex (real world objects) visual stimuli (Fig. 2). Inter-item interactions were thus indexed by a



**Fig. 2** Predicting heterogeneous search times based on homogeneous search times. Top panel illustrates the procedure. First, the logarithmic slope for homogeneous displays is measured ( $D_i$  values), then those parameters are used along with Equation 2 to predict reaction times in heterogeneous displays (sample displays on the right). Bottom left panel: Search for geometric stimuli: observed vs. predicted reaction times (RTs) in Lleras et al. (2019). Each dot corresponds to one of 45 different types of

lure mixtures across three different experiment. Equation 2 captured 90% of the variance with  $\beta=1.8$ . Bottom right panel: Search for real-world objects: observed vs. predicted RTs in Wang et al. (2017). Each dot corresponds to one of 21 different lure mixtures tested in the heterogeneous search display condition. Equation 2 captured 96% of the variance, with  $\beta = 1.3$



multiplicative factor  $\beta$  as shown in Equation 2.

$$RT = a + \beta * \left( \sum_{j=1}^L (D_j - D_{j-1}) * \ln \left( N_T - \left( \sum_{i=1}^{j-1} N_i \right) * 1_{[2,\infty)}(j) + 1 \right) \right) \quad (2)$$

## Shortcomings of Buetti et al.'s (2016) model

There are four major shortcomings in Buetti et al.'s (2016) model. We first list and describe them before addressing them with a new theory in the next section. The first shortcoming concerns the fact that the thresholds of the accumulation process are proportional to the lure-target similarity, such that lures that are more similar to the target would have higher thresholds and thus take longer to process. This poses the *problem of predetermining the thresholds at the start of a trial*. In other words, how does the system set, a priori, the threshold for each lure in the scene? If the threshold is set at the start of the accumulation process, this is tantamount to saying that the system already knows what the lure being processed is (or at least, its degree of similarity to the target!). It would also imply that the system has sufficient information to determine which items are lures and which are candidates. If this was the case, why would the system bother to go through the accumulation process? Buetti et al.'s model overcame this problem via supervision: we, the modelers, set the threshold for each lure type in each simulation. But a successful model ought to determine the decision time for a given type of lure in an *unsupervised* manner. In other words, a successful model ought to be able to operate on distractors without knowing a priori how similar each of those distractors is to the target. The model ought to react differently (accumulate evidence slower) to distractors that are more similar to the target than to distractors that are very dissimilar from the target.

The second shortcoming concerns the *candidates' contribution to stage-one processing*, when these items are present in the display. While experimentally we reported independent contribution to RT by candidates and lures (Experiments 3A–D, Buetti et al., 2016), all search items on the display, including candidates, must go through the information accumulation process. Because candidates have a higher threshold in that first model, they usually end up being the last ones to complete processing. We should thus expect the number of candidates to also contribute to stage-one processing time in a logarithmic fashion (in addition to their linear contribution during stage-two processing) because stage one only ends after all accumulators have reached threshold, including all the candidate accumulators. This contradicts empirical results (including our own; Buetti et al., 2016) where candidates have

been shown to produce linear RT-by-set size functions, with no evidence of a negative accelerated curve.

A third shortcoming of the initial model is that *the definition of candidates and lures is empirically driven*. That is, lures were defined as those distractors that produced logarithmic search functions and candidates as those that produced linear search functions. We provided no a priori mechanism to distinguish these two types of items, given a specific target item. This concern can be attenuated in part by the predictive successes in Wang et al. (2017) and Lleras et al. (2019), where we demonstrated that once it is determined that a stimulus is a lure with respect to a given target, it will continue to be a lure and hold the same “similarity” relationship irrespective of the context in which it is tested.

A final shortcoming of the model is perhaps the assumption that *the initial parallel processing stage is exhaustive*. That is, although the entire model seems to be predicated on that assumption, it might seem too strong of an assumption. For instance, it is well known that visual processing rates decrease with increasing eccentricity (e.g., Carrasco & Frieder, 1997; Wang et al., 2018). Suppose then that a display has items at various eccentricities, including fairly far into the periphery. It surely sounds extreme to propose that the first stage of processing would continue until all the items are processed, including the more peripheral and slower-to-process items. Surely, in such scenarios there ought to be some rule that terminates processing before the farthest stimuli are fully processed. Such scenarios are consistent with the vast literature on the functional viewing field or useful field of view (e.g., Ball, Beard, Roenker, Miller, & Griggs, 1988; Sanders, 1970; Williams, 1989).

## Target Contrast Signal Theory

Like Buetti et al.'s (2016), the Target Contrast Signal Theory (TCS) is a theory that describes the temporal dynamics of peripheral parallel processing during efficient search, when the target is known and fixed across trials. TCS also addresses the major shortcomings of the Buetti et al.'s initial model. As a brief overview, TCS proposes that in visual search, visual processing begins with the accumulation of a contrast signal between each item in the display and the target template. This accumulation of evidence is toward a non-target threshold and occurs in parallel across all locations in the display. Locations that reach threshold are determined to be unlikely to contain the target and are discarded from further processing. The “end” of stage-one processing is determined by a time-out mechanism: If no accumulators reach threshold during a time interval  $T_0$ , visual processing moves on to stage two where attention (and/or eye movements) is deployed serially to inspect the remaining locations until the target is found or search is terminated. A good analogy for this time-out parameter is cooking popcorn in the microwave. The instructions indicate

that, after the popcorn has started popping, the popcorn has finished cooking once there is an interval of 3 s between pops. Similarly, we propose that after accumulators have started hitting threshold, if a time interval of  $T_0$  elapses without any accumulator completing, then parallel evaluation of the display is terminated and a decision is made to move attention or the eyes towards one of the unresolved accumulators. To be clear, when the time-out elapses, it does not mean that all parallel analysis of the scene ends. Accumulators that had not reached threshold at this point will continue to be analyzed after the eye movement.

## Characterization of TCS

**Computation of a difference signal** Perhaps the most important innovation of CST is with respect to *what* is being accumulated in the accumulators. As did many feature-oriented and similarity-based theories before (e.g., Duncan & Humphreys, 1989; Treisman & Gelade, 1980; Wolfe, 1994; Zelinsky, 2008), in Buetti et al.'s initial model we proposed that each accumulator accumulated perceptual evidence at each location. That is, the accumulators were conceptualized as sampling perception consisting of the features present at each location. TCS instead proposes that the accumulators are not sampling perception but are instead computing a difference signal: the contrast between the visual characteristics of the item in the scene and the target template. The computational goal of stage-one processing remains the same as in the previous model: it consists of determining which items in the display are unlikely to be the target. Thus, the end of the accumulation process for a location occurs whenever evidence reaches a non-target decision threshold, that is, when the system has accumulated enough evidence that the item in that location is visually different from the target template.

In Appendix 1 (“Why accumulate contrast?”) we situate the current TCS proposal about computing a difference signal in the context of current theories and findings in the literature. There is indeed growing evidence that visual search for a specific feature is not achieved by tuning attention to that specific feature.

**Similarity impacts accumulation rates** One improvement over the previous model is that there is no longer a need to provide the model with a known value for the target-distractor similarity at each location. Indeed, the evidence accumulation process for each item is modeled by 1D Brownian motion accumulators, whose mean accumulation rate is a function of the overall target-distractor contrast difference. The model “discovers” the difference as it gathers evidence, resulting in different evidence accumulation rates for different levels of lure-target similarity. The more items are dissimilar to the target (the larger the contrast is), the larger each (noisy) sample of the target contrast will be. Over time, this means the accumulation

rate will be large and the accumulators will reach threshold quickly. In other words, when the visual dissimilarity between the target and the lure is large, every sample will provide a large amount of evidence that the lure is not like the target. In contrast, the more similar target and distractors are (the smaller the contrast signal is), the smaller each (noisy) sample of the target contrast will be. This results in a smaller accumulation rate and accumulators will take relatively longer to reach threshold. That is, when the visual dissimilarity between the target and lure is small (high lure-target similarity), every sample will only provide a small amount of evidence regarding that difference, leading to a slower accumulation of evidence. In sum, the accumulation rate simply tracks the likelihood that any given sample will provide *positive* evidence for a difference between visual properties of the distractor and the target template.

In formal terms, TCS proposes that the slope of the logarithmic function ( $D$ ) is inversely proportional to the overall magnitude of the contrast signal ( $C$ ) between a distractor item and the target template (Equation 3), with a multiplicative constant  $\theta$ .

$$D = \frac{\theta}{C} \quad (3)$$

Finally, the rate of evidence accumulation is likely to be dependent on a number of additional factors such as eccentricity (Buetti et al., 2016), size of the item (Wang et al., 2018), and crowding (Madison, Lleras & Buetti, 2018) precisely because these factors change the magnitude of the contrast signal.

**Fixed threshold for all accumulators** TCS proposes that, all else being equal, there is a fixed threshold for all accumulators, or, more precisely, that the threshold is independent of the lure-target similarity. The threshold can be understood as the amount of evidence that is required for the visual system to confidently reject that an item is not the target. Once an accumulator reaches threshold, the visual system ceases to consider its location as a possible location for the target. Note that in this view, there is no need to “suppress” the location of the rejected lures (for distractor suppression accounts see Arita, Carlisle, & Woodman, 2012; Klein, 1988; MacInnes & Klein, 2003; Moher, Lakshmanan, Egeth, & Ewen, 2014; Müller & von Mühlenen, 2000; Takeda & Yagi, 2000; Thomas & Lleras, 2009; Thomas et al., 2006; Woodman & Luck, 2007). These locations are simply ignored in further processing, and only the locations that have not reached threshold continue to be of potential interest to the system. Finally, TCS does not explicitly represent a target-distractor similarity signal (TDS) as is commonly done in computer vision models of vision and in saliency and activation-map accounts of vision (e.g., Itti & Koch, 2000; Wolfe, 1994). The accumulators

that reach threshold are simply coded as having reached threshold regardless of how long it took for them to do so. Accumulators that did not reach threshold are simply passed on as potential targets for attention, without any ranking, or prioritization. That is to say, TCS does not query or represent the amount of evidence at non-resolved accumulators to prioritize some candidate locations over others. Recent work suggested that there is no similarity-based prioritization amongst the candidates (see Ng, Patel, Buetti, & Lleras, [submitted](#)) and it is likely that other factors might come to play in terms of how eye movements are selected (such as a viewer's scanning preferences and saccade distance, for example).

### Termination rule determined by a time-out mechanism

Because the evidence that is accumulated reflects the contrast signal between the target template and a specific item in the display, the accumulation rate of the target itself will be near zero. Candidates will also have near-zero, or relatively small, accumulation rates. This introduces the problem of termination. If certain items will never reach threshold, what determines whether enough processing of the scene has taken place? We propose that the termination is determined by a time-out mechanism. The time-out parameter ( $T_0$ ) is defined by a time interval during which no accumulators reach the decision threshold. After accumulators start reaching the threshold, if a time interval  $T_0$  elapses without any additional accumulator reaching threshold, the list of locations associated with unresolved accumulators is passed on for consideration as possible targets for inspection by focused attention and/or an eye movement. The goal then will be to scrutinize that location with heightened resolution (e.g., Desimone & Duncan, 1995). In other words, locations on this list are scrutinized during the second stage of processing until the target (if one is present) is found.

There are multiple factors that might impact this time-out parameter. For instance, task instructions might encourage participants to actively move their eyes when searching for the target as in “active search” conditions (e.g., Lleras & von Mühlenen, 2004; Smilek, Enns, Eastwood, & Merikle, 2006). Such a strategy would be modeled by a relatively small time-out parameter. This is in contrast to a “passive” search mode, where participants might prefer to maintain fixation and let accumulators complete as much as possible prior to moving their eyes. Such a strategy would be modeled by a relatively larger time-out parameter. There might also be individual differences in terms of participants' preferences to actively search the display by moving their eyes versus keeping their eyes still, and perhaps certain traits (e.g., hyperactivity) might also be associated with more active searches and correspondingly smaller time-out parameters. The time-out might also be sensitive to the duration of central fixation in the most recent trials, such that it might be reduced if processing on the

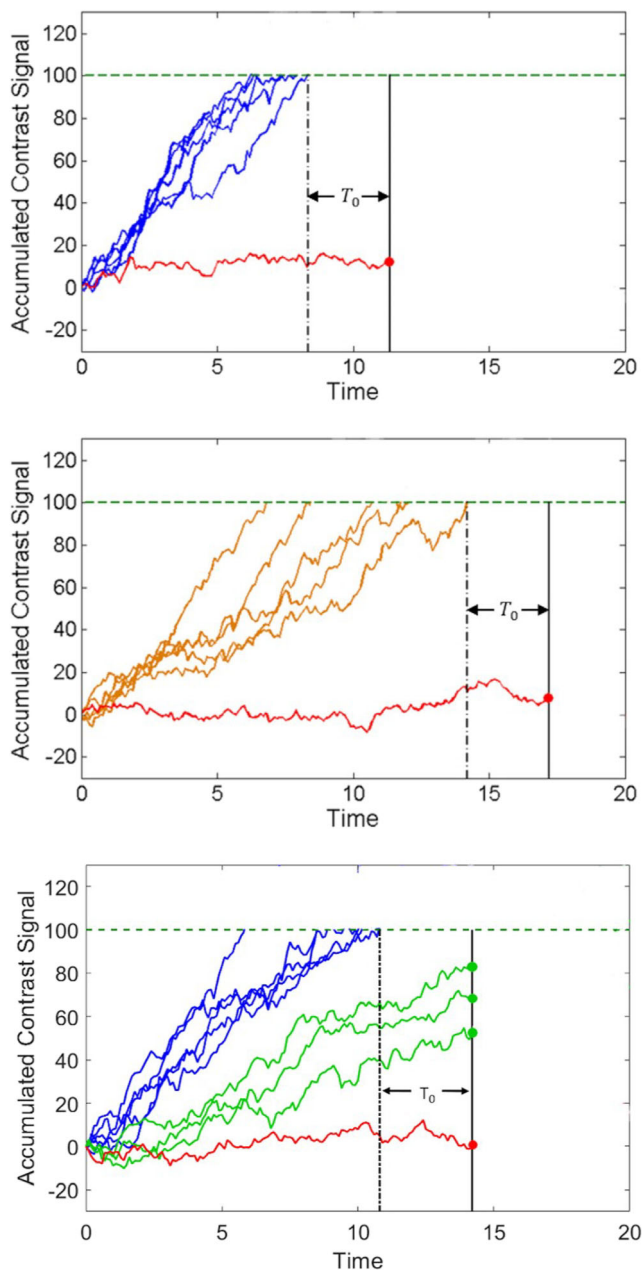
previous trial was easy and might be longer if processing on the preceding trial was more difficult (e.g., Trukenbrod & Engbert, 2014). It is also possible that the default setting for the time-out might be determined by the underlying distribution of fixation durations during normal scene viewing (e.g., Nuthmann, Smith, Engbert, & Henderson, 2010). Similarly, the time-out might also be sensitive to target prevalence effects (e.g., Mitroff & Biggs, 2014; Wolfe, Horowitz, & Kenner, 2005; Wolfe & Van Wert, 2010), such that it is progressively shortened when targets become rare.

As a result, variations in the time-out parameter vary the spatial extent that is processed before a decision to move attention/eye movements is reached. Items near fixation will be processed faster than items farther in the periphery because accumulation rates decrease as a function of eccentricity (e.g., Wang et al., 2018). Therefore, a small time-out parameter is associated with the parallel evaluation of only a small region around fixation because the accumulators in the periphery might not reach threshold before the time-out elapses. More on this will be presented in the section on eye movements and the discussion sections, but it is important to note that the concept of a functional viewing field (e.g., Hulleman & Olivers, 2017) is an emergent property of the proposed architecture.

A simulation of stage-one processing according to TCS is shown in Fig. 3 (top and middle panels), which illustrates the accumulation of a contrast signal in a search display containing one target and several lures. The top and middle panels show how the contrast signal accumulates when the dissimilarity between the target (red T) and the lures is very high (blue circles, top panel) or moderate (orange crosses, middle panel). The red line represents the accumulation associated with the target item. The blue and orange lines represent the noisy accumulation process for the blue and orange lures, respectively. Due to noise in the accumulation process, the individual lures reach the decision threshold at different time points, even when they are all identical. However, in most cases, these time points are close to each other and are within the time-out interval. The target itself has a near-zero contrast signal against the target template and will not reach threshold. Once a time interval longer than the time-out elapses without accumulators reaching threshold (in the figure, that occurs after all lure accumulators reach threshold), attention will directly move towards the location of the only accumulator that did not reach threshold (the target).

Therefore, in TCS, the addition of the time-out parameter introduces a new termination rule that is *not exhaustive*. When all distractors are lures, then, in most cases, we can expect all lure accumulators to reach threshold before the time-out parameter. In this case, the model will behave as a lure-exhaustive model (Fig. 3, top and middle panels). However, we can also expect that in some situations, not all





**Fig. 3** Simulated stage-one processing when one target (e.g., a red T, shown as a red line) is accompanied by five lures that are highly dissimilar (e.g., blue circles, shown as blue lines in top panel) or moderately dissimilar to the target (e.g., orange crosses, shown as orange lines, middle panel). The bottom panel shows simulated stage-one processing when the target is accompanied by five dissimilar lures (e.g., orange crosses, shown as orange lines, middle panel) and three candidate items (e.g., red Ls, shown as green lines). The lines represent the noisy accumulation of contrast signal over time for each item. The green dashed line represents a fixed decision threshold set at 100 (arbitrary units). In all panels, the accumulator associated with the target item (red line) produces a near-zero contrast signal and never reaches threshold. The two vertical lines indicate the points at which the time-out  $T_0$  counter was started and then when it ended, triggering the decision to move attention (or the eyes) to one of the accumulators that did not reach threshold. Time and contrast signal are in arbitrary units

accumulators will reach threshold. As previously mentioned, this is the case for instance when participants decide to

actively move their eyes, spending little time at each location (this corresponds to a very small time-out parameter). Similarly, if many items in the display are far in the periphery, the accumulators for these items will have small drift rates; even if all the elements are indeed lures, a time-out might elapse before all lures reach threshold. At that point in time, attention will move to one of the unresolved accumulators, which will include lure items. Thus, TCS predicts that under certain situations, attention (and eye movements) will “false alarm” and deploy to a lure, rather than a target, even under efficient search conditions. Below, we provide eye-movement evidence consistent with this prediction. Evidence that attention false alarms to “distractors” even in efficient search has also been recently presented by Rangelov, Müller, and Zehetleitner (2017).

Because candidates are quite similar to the target, their accumulators ought to rarely reach threshold and thus, their presence and number do not contribute to overall processing time during the parallel evaluation of the scene. This is an important key distinction from the previous model. Indeed, since candidate accumulators fail to reach threshold (in most cases), and parallel processing times are determined by the completion time of last accumulator that reaches threshold before the time-out, this property of the new model explains why candidates do not contribute logarithmically to RT. Figure 3 (bottom panel) illustrates the accumulation of a contrast signal in a search display containing both lures (e.g., orange crosses) and candidates (e.g., red Ls) in addition to the target (e.g., a red T). The contrast signal for candidates takes longer to accumulate because of the small contrast signal they generate. When candidates are not overly similar to the target, the candidate-target contrast signal is not entirely zero. In such cases, candidate accumulators might eventually reach threshold. However, we propose that in most of such cases, the time-out parameter  $T_0$  will be reached prior to that time (as visualized in Fig. 3, bottom panel). Thus, parallel processing will actually terminate prior to any of the candidate accumulators reaching threshold.

In sum, the introduction of the time-out parameter addresses two of the main shortcomings of the preceding model: (i) the known fact that the number of candidates only contributes linearly to RT (not logarithmically) and (ii) the observation that even efficient search is not always exhaustive of the entire display (Rangelov et al., 2017).

**Redefining “lures” and “candidates”** Due to the stochastic nature of evidence accumulation, it is possible that in a small proportion of cases some candidate accumulators might reach threshold before the time-out elapses. In this case, those locations will be discarded during parallel processing of the scene and will not be scrutinized by focused attention/eye movements later on. This observation highlights the *probabilistic*

distinction between lures and candidates. That is, in most cases, stimuli that are sufficiently different from the target template ought to be rejected in parallel by the visual system in the vast majority of cases and stimuli that are sufficiently similar to the target template are candidates for scrutiny by focused attention and eye movements. That said, the combination of the processing properties of peripheral vision<sup>1</sup> and the duration of the time-out parameter do create *exceptions to those rules*. For instance, a lure that is relatively similar to the target template and that appears far in peripheral vision might have a slow accumulation rate that will not hit threshold before the time-out elapses. Similarly, a candidate that is visually not too similar to the target and that appears near fixation might have a faster accumulation rate and might be rejected in parallel prior to the time-out being reached (because of increased processing resolution of vision near fixation; see Fifić, Townsend, & Eidels, 2008). Thus, instead of defining lures and candidates with respect to their similarity to the target, it would be more precise to define them based on the type of processing that the visual system uses to categorize them as non-target items. Thus, lures are items that have reached threshold and the visual system can confidently discard in parallel whereas candidates are items that have yet to reach threshold and the visual system requires close inspection or scrutiny to rejecting them. Note, though, that these labels tend to be stable for any given target-distractor pair, as evidenced by the success in predicting performance across different groups of participants and display complexity (Lleras et al., 2019; Wang et al., 2017).

Finally, it is important to remember that in TCS the target is processed just like any other candidate. This means that regardless of where the target is in the periphery (near or far), it will accumulate next to nothing in terms of contrast because the goal of peripheral analysis is to find evidence to discard unlikely target locations. So, even if the target appears near fixation (but not at fixation), it should be treated (blindly) as any other peripheral distractor and evaluated to see if its location can be rejected. The accumulation process will not reach threshold and the target location will therefore be amongst the set of possible target locations. This is different from what would happen if and when the eyes are looking directly at the target. It is likely that foveal information accumulates not towards a non-target threshold, but rather, towards a target-threshold, i.e., the perceptual information at the fovea is compared to the target template for a positive match.

<sup>1</sup> Similar to the Texture Tiling Model of Rosenholtz and colleagues, TCS is premised on the fact that there are processing limitations in peripheral vision. These limitations constrain the speed at which the target-contrast signal is computed at any given eccentricity and the confidence with which peripheral locations are classified. Note that these limitations are not just in terms of decreased visual acuity (i.e., they are not equivalent to a Gaussian blurring of the image input, as proposed by Zelinsky, 2008) but reflect a deeper limitation in how information is processed in the first place (see Rosenholtz, 2016; Strasburger et al., 2011).

## Highlights of Target Contrast Signal Theory (TCS)

What TCS can do:

1. TCS provides an account for the variation of RT as a function of set size observed in efficient search tasks, which has been overlooked by previous theories. This variation in RTs is systematic and informative: the logarithmic slope coefficient measured in a typical efficient search task (one target, all identical lures) indexes the dissimilarity between the target and lure stimuli.
2. Using Equation 2, the logarithmic slope coefficients can be used to predict performance in future experiments where the same target and lure stimuli are again paired, even in heterogeneous search displays (e.g., Lleras et al., 2019; Wang et al., 2017).
3. Current theories typically predict that the eyes ought to always move directly to the target in efficient search tasks, yet in many occasions they do not (Ng, Lleras & Buetti, 2018; Rangelov, Muller & Zehetleimer, 2017). TCS provides a framework for understanding the occurrence of eye-movements to non-target elements.
4. TCS provides a framework for evaluating and quantifying lure-lure similarity effects (referred to as *homogeneity* effects) in efficient search, independently from target-lure similarity (Lleras et al., 2019).
5. TCS provides a framework for understanding processing costs associated with varying levels of target-distractor similarity: lures incur logarithmic processing costs but no linear processing costs, while candidates incur linear processing costs but no logarithmic processing costs.
6. TCS provides an intuitive account for the phenomenon of search asymmetries due to the fact that the target-contrast signal is computed as a function of the visual properties present in the target template (not as a function of the properties present in the distractor).
7. TCS provides a framework for understanding the impact of stimulus factors such as visual crowding and stimulus size as these factors are integral to determining the quality of peripheral representations.

Below follows a list of the most important limitations of TCS (see also the *Limitation* section in the *General discussion*):

1. TCS has yet to include a formal description of how processing unfolds during stage two and how eye movements are chosen when multiple possible target locations are present in the scene.
2. TCS is not a theory of oddball search, which has a very different RT × set size function (e.g., Bravo & Nakayama, 1992; Buetti et al., 2016).

3. TCS is not a theory of attentional capture. Certain forms of capture (contingent-capture) might well be understood within TCS framework but a more comprehensive theory of all capture (surprise capture, bottom-up capture) is beyond the scope of the theory.
4. TCS is a computational, cognitive theory, that is not necessarily neurally inspired. We do not know where the neurons required to represent the target-contrast signal reside or if they even exist (see Appendix 1 for a discussion on contrast signals in the brain). This computation might be represented in neuron assemblies or be distributed along different levels of the visual system.

A more thorough discussion of the limitations of the TCS is presented in Appendix 3.

## Simulation of TCS

As mentioned earlier, it is important to make a distinction between two different types of equations in our work. Like in Buetti et al. (2016), each accumulator in TCS is modeled as a Gaussian random walk with a positive mean drift rate, such that the completion time for each item still follows the Inverse Gaussian distribution. There is, therefore, a first set of equations: those that drive the behavior of individual accumulators. These equations reflect the underlying mechanism of the model: that contrast evidence is accumulated at varying rates and with various degrees of noise. Note that these equations cannot be tested behaviorally. A second set of equations summarize the behavior of a group of accumulators, such as Equations 1 and 2. Unlike the equations governing individual accumulators, these global equations can be tested empirically (both in human experiments and computer simulations) and used to predict performance in novel scenarios (e.g., Lleras et al., 2019; Wang et al., 2017).

Below, we present a series of simulations that illustrate the basic properties of TCS. The goal of these simulations was to demonstrate that the underlying architecture of the model (i.e., multiple stochastic accumulators processing information in parallel until a time-out is hit) can reproduce the patterns of behavior we now know characterize efficient search. Examples of such patterns are: The presence of a logarithmic RT by set size function; modulation of the logarithmic slope function by target-lure contrast signal; RTs in lure-heterogeneous displays determined by Equation 2; and the observation that candidates do not contribute to stage-one processing costs. In addition, we also present simulations that demonstrate other aspects of behavior emerging from the TCS architecture. These simulations evaluated for instance, the impact of the noise parameter on processing times and what

conditions might lead lures to be inspected by attention and/or eye movements during stage-two processing.

**Approach** In all simulations the accumulation threshold was fixed at 100 (arbitrary units, see Fig. 3).<sup>2</sup> The following parameters were varied in the simulations (values were reported in the figures): the number of each type of item in a search display  $N_i$ , the average accumulation rate for each type of distractor item  $k_i$ , and the standard deviation of the Gaussian noise in the accumulation rate  $\sigma_i$ .

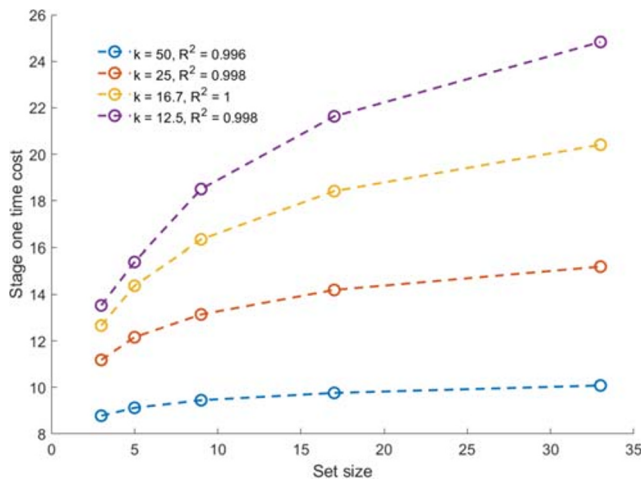
When simulating a visual search trial, for each type of item (denoted by subscript ‘i’ here) characterized by an accumulation rate  $k_i$  and an accumulation noise  $\sigma_i$ , we simulated  $N_i$  random variables that follow an Inverse Gaussian distribution (Fig. 4). Only one target is present per search trial, with a fixed accumulation rate of  $k=0.02$ . The target was present on each trial. The noise parameter was set to the average of all other items (this does not influence the outcome of the simulation). Next, we sampled each random variable based on their respective Inverse Gaussian distribution to determine the processing time for each item in a particular trial. The time-out mechanism was then implemented by sorting the item completion times in ascending order and then locating the first interval between each successive completion time that was greater than  $T_\theta$ . The simulated time cost for stage-one processing was defined to be equal to the completion time of the last item before the timeout.

The simulation was then run multiple times ( $N = 8,000$  for each simulated condition) to obtain a reliable estimate of the expected processing time as well as the distribution of processing times for each specific condition. The results obtained from these simulations were then compared to group-level mean data from participants who searched displays similar to the ones used in the simulations, as we assumed that noise and artifacts present in individual subject data will be averaged out in group estimates.

**Simulation results** The first set of simulations were aimed at validating TCS by verifying that it can reproduce the three following important properties:

- 1) *Logarithmic RT by set size functions that are modulated by lure-target similarity.* When the target was fixed and distractors varied in their accumulation rate  $k$ , the time cost by set size functions at different levels of  $k$  were all logarithmic (Fig. 4). Note that when all lures are identical,

<sup>2</sup> Note that fixing the threshold does not result in the loss of any degrees of freedom in the parameter space, because the Inverse Gaussian distribution is characterized by two free parameters. Thus, varying the accumulation rate parameter  $k$  and the noise parameter  $\sigma$  covers the full spectrum of possible distributions that describe an individual item’s processing time under the current model.

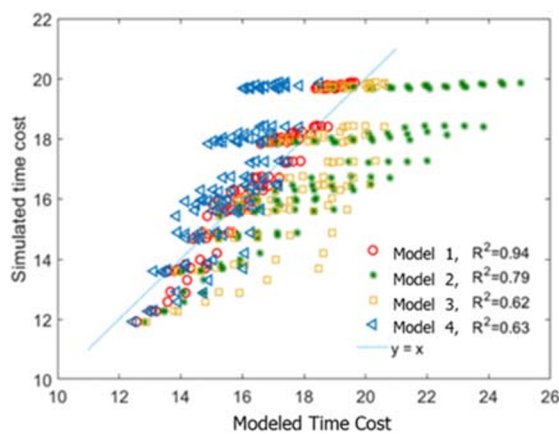


**Fig. 4** Simulations of stage one-time cost (in arbitrary units) according to the Target Contrast Signal theory. The time costs increase logarithmically as a function of set size. Simulations are shown for four different values of the accumulation rate parameter  $k$ , noting that  $k$  is a function the overall magnitude of the lure-target contrast signal. As  $k$  increases, log functions flatten out, indicating faster overall processing times. Fixed parameter values:  $T_0 = 6$ ,  $\sigma = 22.36$ . R-Square values indicate the strength of the logarithmic fit for each condition

Equation 1 simplifies to:

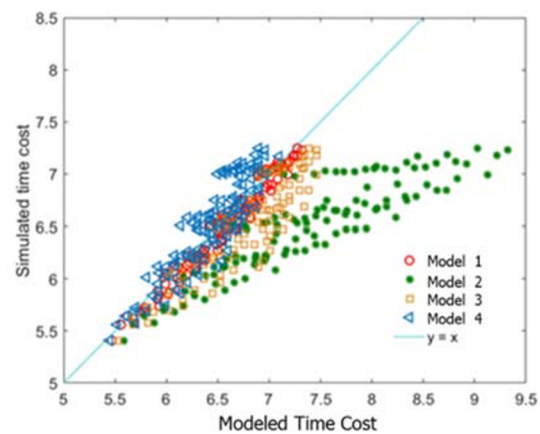
$$RT = a + D_1 * \ln(N_T + 1)$$

Thus, the R square values indicated that the overall completion times in the simulations for a given accumulation rate



matched Equation 1 extremely well. Furthermore, lower  $k$  values (i.e., small contrast signal) led to steeper logarithmic functions, consistent with TCS’ proposal that logarithmic slopes are inversely related to the contrast signal (Equation 3). These simulation results are consistent with human data showing steeper logarithmic functions for high-similarity lures (Buetti et al., 2016; Madison et al., 2018; Ng et al., 2018; Wang et al., 2018).

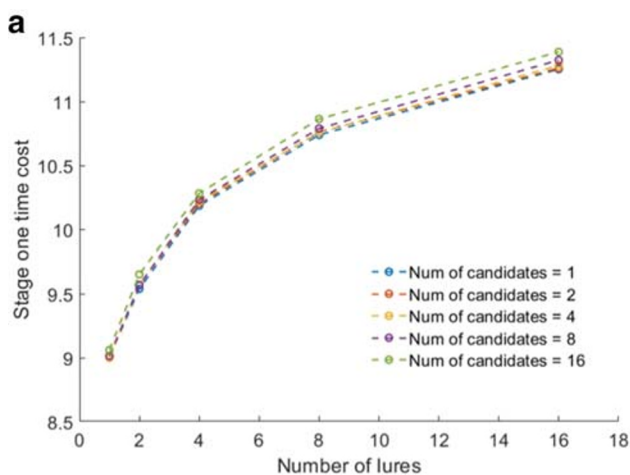
2) *Ability to account for completion times in heterogeneous displays.* Wang et al.’s (2017) contribution was to demonstrate that global Equation 1 best predicted the behavior of the entire set of accumulators that followed the architecture of Buetti et al.’s (2016) model (Fig. 5, right). In that study, the authors compared four models with distinct underlying architecture (described in Fig. 5). The present simulations aimed at verifying that Equation 1 was still the best global equation to predict the behavior of the set of accumulators, as now described in the TCS architecture. As in Wang et al. (2017), the parameters used in the heterogeneous display simulations were the best-fitting logarithmic fits to the completion times in simulated homogeneous displays. The results confirmed that Equation 1 provided the best estimate for the global time costs to process heterogeneous search displays (Fig. 5, left). In other words, in spite of the new mathematical implementations in TCS (new termination rule, most critically), Equation 1 continues to account for completion times in displays simultaneously containing different types of lures. This means that Equation 1 is consistent with TCS and therefore TCS can



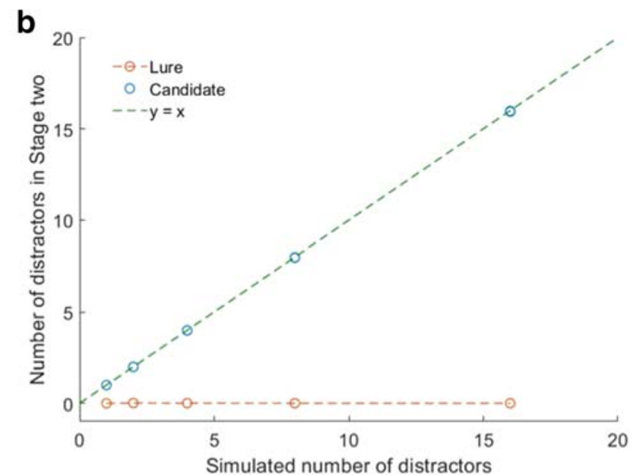
**Fig. 5** Simulated search time costs in heterogeneous displays based on parameters estimated in simulated homogeneous search conditions. Comparison of simulations carried out using TCS (left panel) and initial model by Buetti et al. (2016) (right panel). The critical difference is that in the left panel, the evidence accumulators in the simulations accumulate contrast signal, with the accumulation rate being proportional to the overall contrast signal and the time-out rule determining when processing finishes. In the right panel, the evidence accumulators accumulate perceptual evidence at the same rate across all accumulators and decision thresholds are determined as a function of lure-target similarity. The figures illustrate the ability of four models to account for completion times in lure-heterogeneous simulations were compared, each model reflecting a different underlying architecture. In both simulations, Model 1, corresponding to Equation 1 in the present paper,

was the winning model, falling the closest to the  $y=x$  line (perfect agreement). In addition, the non-winning models had qualitatively similar deviations from the  $y = x$  line. The same set of heterogeneous search conditions were used in these two simulations. For reference, Model 2 represents a distractor discounting model where distractors are rejected by subsets determined by their similarity to the target (e.g., for a red target, all the blue lures would be discarded first, then accumulation for the yellow lures would start until those are rejected, and so forth). Models 3 and 4 represent models where all the lures are rejected at the same processing rate, with Model 3 using a max rule (inspired by Zelinsky, 2008) and Model 4 using an average discounting rate. For more details see Wang et al. (2017). The results show that the architecture of TCS produces overall RTs that are consistent with the predictions of Equation 1 when predicting performance in lure-heterogeneous displays





**Fig. 6** Candidates do not affect stage-one processing time according to TCS. **Left panel:** Simulated stage-one time cost as a function of number of lures depending on number of candidates. No significant modulation by number of candidates was visible. Parameter values:  $T_0 = 4$ ,  $\sigma = 10$ ,  $k_{lure} = 20$ ,  $k_{candidate} = 0.25$ . **Right panel:** Number of distractors that entered stage-two processing versus actual number of distractors in the simulation. Data for candidates fall on the  $y=x$  line, indicating all of them entered stage two. In contrast, data for lures fall on the  $y=0$  line, meaning they were all rejected



predict the RTs in heterogeneous search scenes (as shown in Fig. 2).

3) *Candidates should not contribute to stage-one processing.* A simulation was run with varying number of lures and candidates (defined by large and small accumulation rates  $k$ , respectively). As shown on Fig. 6 (left), stage-one time costs vary only minimally when candidates were added to the simulation. Figure 6 (right) shows the number of items that did not reach threshold before the time-out elapsed as a function of distractor type and number of distractors. One can see that with the chosen parameters, lure locations are never passed on to stage two while all candidate locations are.

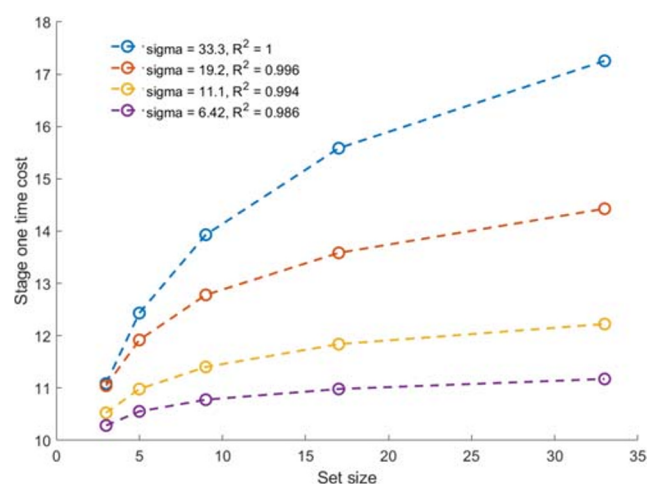
The next set of simulations were aimed at identifying novel properties of TCS.

4) *Impact of the noise parameter on stage-one processing time.* In the model, the time cost for stage-one processing is modulated by the noise parameter  $\sigma$  in addition to the rate parameter  $k$ . Given a constant  $k$  rate, increasing  $\sigma$  should lead to steeper logarithmic curves. This property was confirmed by the simulations (Fig. 7) and suggests that factors such as visual crowding that increase the noise in the accumulation process ought to result in steeper logarithmic curves. Evidence consistent with this property of TCS was presented in Madison et al. (2018) where it was shown that search efficiency was reduced (i.e., the logarithmic slopes became steeper) in displays where crowding was possible compared to displays where crowding was minimized. Furthermore, stimulus size might also be a factor that impacts noise, with smaller stimuli in the periphery leading to larger noise in the accumulation process. Evidence consistent with this property was presented in Wang et al. (2018) where it was shown that search efficiency increased

during stage one. This simulation is meant to illustrate an ideal case where the accumulation rates are sufficiently different, and the time-out not too long or too short, to allow accurate triage of the stimuli. As mentioned in the text, a stimulus that is treated as a candidate in the periphery may be screened as a lure when presented close enough to fixation, if the distractor-target similarity is not too high. Similarly, lures that are presented sufficiently far in the periphery may have accumulation rates so low that they end up being candidates for eye movements (see the text for more details)

(i.e., the logarithmic slopes become shallower) as the size of stimuli in the periphery increased.

5) *Conditions leading lures to be scrutinized in stage two.* The second set of simulations evaluated the likelihood that target-dissimilar items do not reach the threshold before the time-out elapses and enter stage two as a result (Fig. 8). We refer to this condition as a “false alarm” as it creates the possibility that attention and/or eye movements will be directed to lures. The probability of these events depends on the values of the time-out parameter  $T_0$ , the accumulation rate  $k$ , and noise parameter  $\sigma$ . As a first example, Fig. 8 (left panel) illustrates



**Fig. 7** Simulated stage-one time cost (in arbitrary units) according to TCS as a function of set size and the accumulation noise parameter  $\sigma$ . As the noise parameter increases, so does the slope of the logarithmic curve. Fixed parameter values:  $T_0 = 6$ ,  $k = 25$ . R-Square values indicate the strength of the logarithmic fit for each condition

simulations where  $k$  and  $\sigma$  were kept constant, while  $T_0$  was varied. These simulations illustrate a search condition with one fixed target and varying numbers of lures. The simulation shows that the likelihood that locations are passed on to stage two rises dramatically for small time-out values. But, as time-out values increase, this likelihood drops to near zero. This function shows a “diminishing returns” situation: increasing the time-out parameter does decrease false alarm rates, but beyond a certain point, the reduction in false alarm rates is perhaps not justified in terms of the costs of additional waiting. For example, whereas increasing  $T_0$  by two-time units from 2 to 4 cuts false alarm rates by more than half, increasing  $T_0$  by the same two units from 6 to 8 only cuts false alarm rates by an almost negligible fraction. Thus, one can expect that the visual system ought to calibrate itself for intermediate values that produce small enough false alarm rates at the shortest possible  $T_0$ .

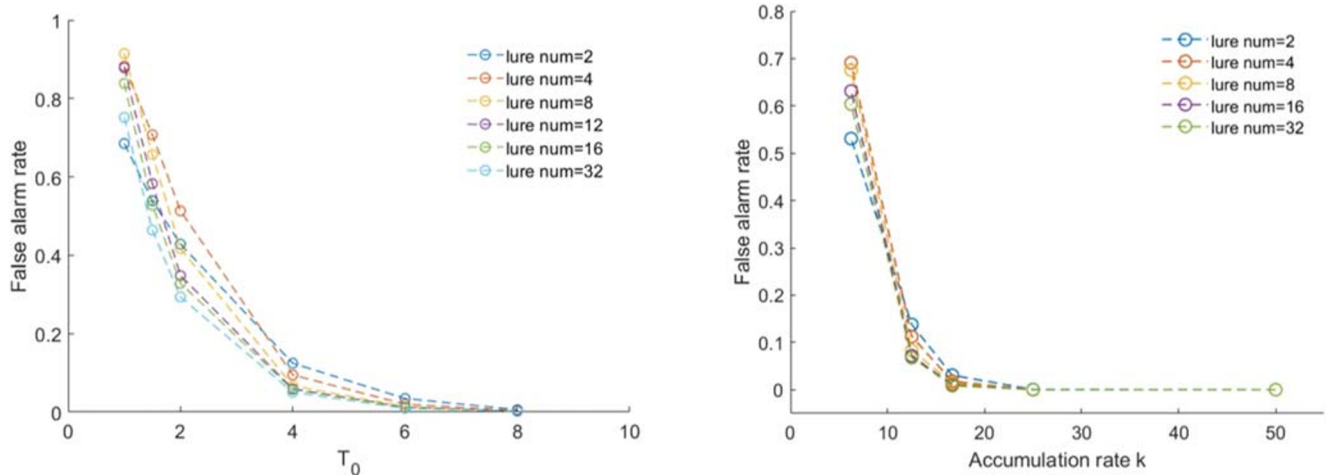
Figure 8 (right panel) illustrates simulations where  $T_0$  and  $\sigma$  were kept constant while  $k$  was varied. As one can intuit, the smaller the accumulation rate, the larger the likelihood that accumulators will not reach threshold and will thus be passed on to stage two. Figure 8, therefore, illustrates conditions where we can anticipate that observers will move their eyes towards lures (rather than to the target). For example, everything else being equal, as a stimulus is presented farther and farther into the periphery, the speed of evidence accumulation for that stimulus decreases (e.g., Carrasco & Frieder, 1997; Wang, Lleras, & Buetti, 2018). Thus, when the time-out parameter is not overly long (in the figure, less than 6 arbitrary units), the likelihood that the same visual stimulus might not reach threshold before the time-out elapses increases with stimulus eccentricity. Because the evidence accumulation rate is a function of lure-target similarity, this set of functions also predicts that, given a fixed eccentricity and a time-out parameter that is not overly long, more similar lures will be more

likely to trigger an eye movement than less similar lures because more similar lures have smaller accumulation rates. Finally, the left panel of Fig. 8 also suggests a solution to avoid eye movements being directed to lures: increasing the time-out parameter. Indeed, as the time-out increases, the likelihood that lure accumulators will not reach the threshold collapses to zero.

In sum, when the time-out parameter is not overly long, TCS predicts that the likelihood of eye movements that are not directed to the target will: (i) increase with target eccentricity and (ii) increase with lure-target similarity. Further, TCS predicts that with sufficiently long time-outs (iii) eye movements (or attentional selection) of non-target elements should be near zero. In the section below, we present evidence consistent with these predictions. The prediction that eye movements to non-target elements will be zero in efficient search (prediction iii) is not new and is actually made by all search models. Nonetheless, TCS makes novel specific predictions for the conditions that ought to lead to eye-movements being directed to non-target elements in *efficient search conditions*. Further, TCS predicts both the occurrence of eye-movements to non-target elements as well as the absence of such eye movements, with variations in a single parameter (the time-out). In this fashion, we believe we have introduced a parameter that adds flexibility to our model in a way that previous models do not.

### Converging evidence from eye-movement data in efficient search

In the present section, we present eye-movement data that support some of the predictions made by TCS in the preceding section. Recently, Ng et al. (2018) published the first study that examined eye movements in an efficient search task that



**Fig. 8** Simulated false alarm rates according to TCS, as a function of the time-out parameter (left panel) and the accumulation rate (right panel). False alarms are instances where a stimulus that ought to have been rejected during the parallel evaluation of the scene is instead passed on

as a possible target for an eye movement or for focused attention. Left panel fixed parameter values:  $k = 16.67$ ,  $\sigma = 12.91$ . Right panel fixed parameter values:  $T_0 = 6$ ,  $\sigma = 12.91$

included target-distractor similarity manipulations as in Buetti et al. (2016). The stimuli were simple geometric shapes that were presented in a circular grid organization to minimize crowding with eccentricities of 4.2, 7.7, and 14.3° of visual angle (Madison et al., 2018; Wang et al., 2018). The goal of the study was to demonstrate that eye movements themselves were not the cause of the logarithmic patterns observed in Buetti et al. (2016). Indeed, previous data from Zelinsky (2008) had shown that, under some experimental conditions, the number of eye movements can increase with set size, in a negatively accelerated fashion (like a logarithmic curve). Thus, it was important to test whether the logarithmic RT patterns in Buetti et al. (2016) were not an artifact arising from the eye movement system, but rather do reflect the sort of parallel, stochastic accumulation of evidence posited. Participants completed the search task both under a free-viewing condition (as in Buetti et al., 2016, and most search tasks, in general) and under a fixed-viewing condition (where they were required to maintain fixation throughout the entire trial).

The results of Ng et al. (2018) showed that even when participants did not move their eyes during search, RTs continued to increase logarithmically as a function of set size (average R-square of 0.89 across all experiments and lure types). This indicated that the lures that we used were sufficiently different from the target such that they could all be confidently rejected via peripheral analysis. In addition, the eccentricity chosen also allowed participants to find the target without moving their eyes. The presence of eye movements in the free-viewing condition would thus be indicative of an internal decision to move the eyes but not of a *need* to move the eyes caused by limitations in peripheral vision. Indeed, when eye movements are allowed, there seems to be a trade-off between continuing to gather evidence through the parallel evidence accumulation process (while maintaining fixation) and moving the eyes to a peripheral location that corresponds to an accumulator that has not reached threshold.

**Target-distractor similarity impacts eye behavior in efficient search** Analyses reported in Ng et al. (2018) indicated that when target-distractor similarity was low, the viewing condition (free- vs. fixed-viewing) did not have a meaningful impact on search efficiency. This follows because in low similarity conditions, participants made fewer eye movements all together than in high similarity conditions and the number of eye movements tended not to be influenced by set size (Fig. 13, bottom, dotted lines). This result is informative regarding the evidence accumulation process and participants' strategy during efficient search: when searching in a display where distractors are very dissimilar from the target, participants seem to take advantage of the fast and parallel accumulation process to find the target. In contrast, when the target-distractor similarity was higher search was *more efficient* (i.e., logarithmic slopes were smaller) in the fixed-viewing

condition (e.g., 35 ms/log unit in Experiment 1B) than in the free-viewing condition (65 ms/log unit). This result suggests that participants sometimes prefer to stop the parallel accumulation process and execute an eye movement. This preference might reflect participants' level of confidence on the quality of the evidence being accumulated, that is, a sense of how well they are sampling the display. When evidence accumulates overly slowly, they may feel that it is more beneficial to redirect the eyes at potential target locations and accelerate these rates. That said, this strategy comes with an increased cost associated with the time needed to execute an eye movement.<sup>3</sup> It is not surprising that participants choose to incur those costs, given that they are likely unaware of them (e.g., Clarke, Mahon, Irvine, & Hunt, 2017; Hunt & Cavanagh, 2009; Mahon, Clarke, & Hunt, 2018) and because in the context of every day's life, we spontaneously move the eyes in these circumstances.

This pattern of results is consistent with TCS. When target-distractor similarity is low, TCS predicts that the accumulation rate of distractors will be high and that all distractor accumulators will reach threshold before the time-out parameter elapses (for most time-out parameter values). On the other hand, when target-distractor similarity is relatively high, TCS predicts an increased likelihood that the time-out will elapse before all accumulators have reached threshold (Fig. 8, right). This in turn triggers an eye movement to one of the unresolved accumulators. Some of these eye movements will be directed to a lure and cause a temporal cost that would not be incurred had participants waited for all accumulators to reach threshold, thus resulting in lower search efficiency. Finally, the results also showed that subjects performed the task with the same level of accuracy under free- and fixed-viewing conditions (98.5% and 97.3%, respectively). In sum, the results validate the decision to include a variable time-out parameter in TCS that determines whether to keep processing information in parallel or whether to move the eyes to an unresolved accumulator. Finally, it is worth remembering that all these results were obtained even though all conditions tested fall within the "efficient" range of search efficiency: all linear slopes observed in the Ng et al. (2018) experiments were smaller than 7 ms/item.

**Where do the eyes go in efficient search** According to TCS, when the time-out elapses prior to all lures having reached threshold, there are multiple possible unresolved accumulators that can be targeted for an eye movement (Fig. 8). Thus, TCS makes a set of specific predictions.

<sup>3</sup> This tradeoff does not necessarily mean that participants are behaving suboptimally when choosing to move the eyes, rather than waiting at fixation. There might be a difference in terms of effort involved in analyzing peripheral information (compared to making an eye movement), and behavior might also not be suboptimal if one considers factors other than search time, such as accuracy, confidence, or energy expenditure.

First, participants should show increased rates of target-uninformed (“guess”) initial saccades when lure-target similarity is relatively high. The term “target-uninformed” reflects saccades that have been generated by information that has accumulated at a non-target location and thus, have been directed to a lure location. According to TCS, accumulators that did not reach threshold are simply passed on as potential targets for attention, without any ranking, or prioritization (Ng et al., *submitted*). TCS predicts that participants will “guess” one location amongst this set of unresolved accumulators.

Second, the farther the target is presented into the periphery, the higher the likelihood that a target-uninformed saccade will be executed prior to the eyes moving towards the target. This follows for the simple reason that if there are multiple unresolved accumulators (one of them being the target), participants will tend to preferentially move their eyes to a location that is nearer to the current fixation (e.g., Zelinsky, 2008). Thus, the larger the target eccentricity is, the less likely its location will be chosen as the destination of the first saccade.

It is worth noting that, although intuitive, these predictions are not necessarily compatible with previous models of search (e.g., Zelinsky, 2008; Wolfe, 1992). Indeed, in efficient search, most models predict that lures are rejected in parallel and the first movement of attention or eye movement is systematically directed to the target. The term “pop-out” is in fact used to characterize the implied automaticity of attentional selection of the target in efficient search. Only when target-distractor discriminability is low do other models predict saccades that do not go directly to the target (e.g., Zelinsky, 2008). Yet, results from Ng et al. (2018) indicated that this was not the case in our displays: target-distractor discriminability was relatively high, even when the target was in the far periphery because participants were able to complete this same task without moving the eyes, with similar levels of accuracy. Furthermore, most models of attention use a prioritization rule, where higher activation locations (i.e., higher TDS signals) have higher priority than lower ones. The lures used in the current experiment have low TDS signals, which are much smaller than the target’s. Thus, it is safe to assume that theories relying on TDS and prioritization rule would not predict many, if any, saccades to lures, in efficient search.

To evaluate these predictions, we performed a series of new analyses (not previously published) on eye-movement data from Experiments 1B and 2 in Ng et al. (2018).<sup>4</sup> In Ng et al.’s Experiment 1B, participants searched for a red triangle, whereas in Experiment 2 they searched for a cyan semi-circle. In both experiments, two sets of distractors were used: orange diamonds and blue circles. As a result, in Experiment 1B, the orange distractors bore high similarity and the blue distractors bore low similarity to the red target, whereas this relationship was reversed in Experiment 2, where orange distractors bore low similarity and blue distractors bore high similarity to the cyan target.

We estimated the probability of a target-uninformed (“guess”) saccade using the MemToolbox (Suchow, Brady, Fougne, & Alvarez, 2013). Originally developed for estimating errors in visual working memory, the MemToolbox allows one to quantify saccade errors as the angular difference between the landing location of the initial saccade and a “perfect” saccade. Here, a perfect saccade refers to a saccade in which target location information is available such that a saccade is executed toward the target. These error data are then fed into a mixture model which assumes, in our experiments, that there are two trial types: one in which the participant has enough evidence to make a target-informed saccade and another in which the participant does not and thus makes a target-uninformed saccade. In this latter case, the direction of the saccade is random with respect to the location of the target. This follows because lures are arranged randomly around the circular grid and there is no consistent spatial relationship between the target and any one lure in the display.

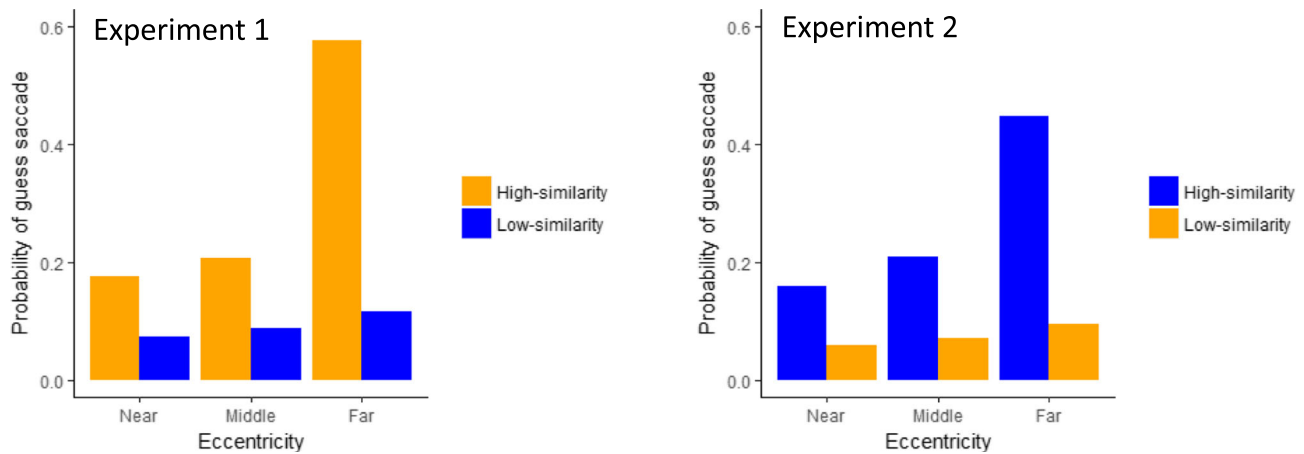
Due to the relatively low number of trials per condition where participants moved their eyes, there were not enough trials to model target-uninformed saccades at the subject level and to perform inferential statistics. Here we present the group-level data (Fig. 9) and the results clearly pass the inter-ocular traumatic test (Edwards, Lindman, and Savage 1963), that is, the difference is so stark that it “hits you” between the eyes. There was a higher probability that the initial saccade was target-uninformed (a “guess” saccade) when lure-target similarity was high (an average of 32.1 % and 32.3% in Experiments 1B and 2, respectively) compared to low (9.3% and 7.5%). In addition, this probability increased with target eccentricity, particularly when lure-target similarity was high: In Experiment 1B, the probability of guess saccades roughly tripled from 20.8% to 57.8% from middle to far eccentricities; in Experiment 2, the probability of guess saccades more than doubled from 21% to 44.8% from middle to far eccentricities. To complement Fig. 10, the landing locations of all first saccades of all participants are displayed in Figs. 11 and 12 (Experiments 1 and 2, respectively). In sum, the patterns of eye movements during free-viewing efficient search are consistent with predictions from TCS.

To confirm the data visualized in Figs. 9, 10, and 11, a series of inferential analyses were conducted. The goal of these analyses was to support the finding that participants executed more target-uninformed saccades when displays had high similarity lures and moved more directly to the target when displays had low similarity lures. As can be seen on Fig. 13, the average distance in degrees of visual angle between the landing location of the first saccade and the target was larger

<sup>4</sup> Data and code can be found on the Open Science Framework: <https://osf.io/pve8d/>

The methods used in these studies can be found in the Appendix.





**Fig. 9** Probability of making a target-uninformed (“guess”) saccade as a function of target eccentricity (4.17, 7.73, and 14.3° of visual angle) and lure-target similarity for Experiments 1B and 2 (left and right panels, respectively) in Ng et al. (2018). Observers were more likely to make a

target-uninformed saccade (not directed toward the target) when lure-target similarity was high, and, in particular, when the target appeared at the far eccentricity

when displays had high-similarity lures, compared to low-similarity lures, in both Experiment 1B (2.4 vs. 5) and Experiment 2 (2.1 vs. 3.9),  $F(1,17) = 63.54$ ,  $p < .001$ ,  $\omega_p^2 = 0.767$  and  $F(1,35) = 65.17$ ,  $p < .001$ ,  $\omega_p^2 = 0.634$ , respectively. This confirms that the initial saccade in low-similarity displays on average landed nearer to the target compared to high-similarity displays, as can be seen in Figs. 10 and 11.

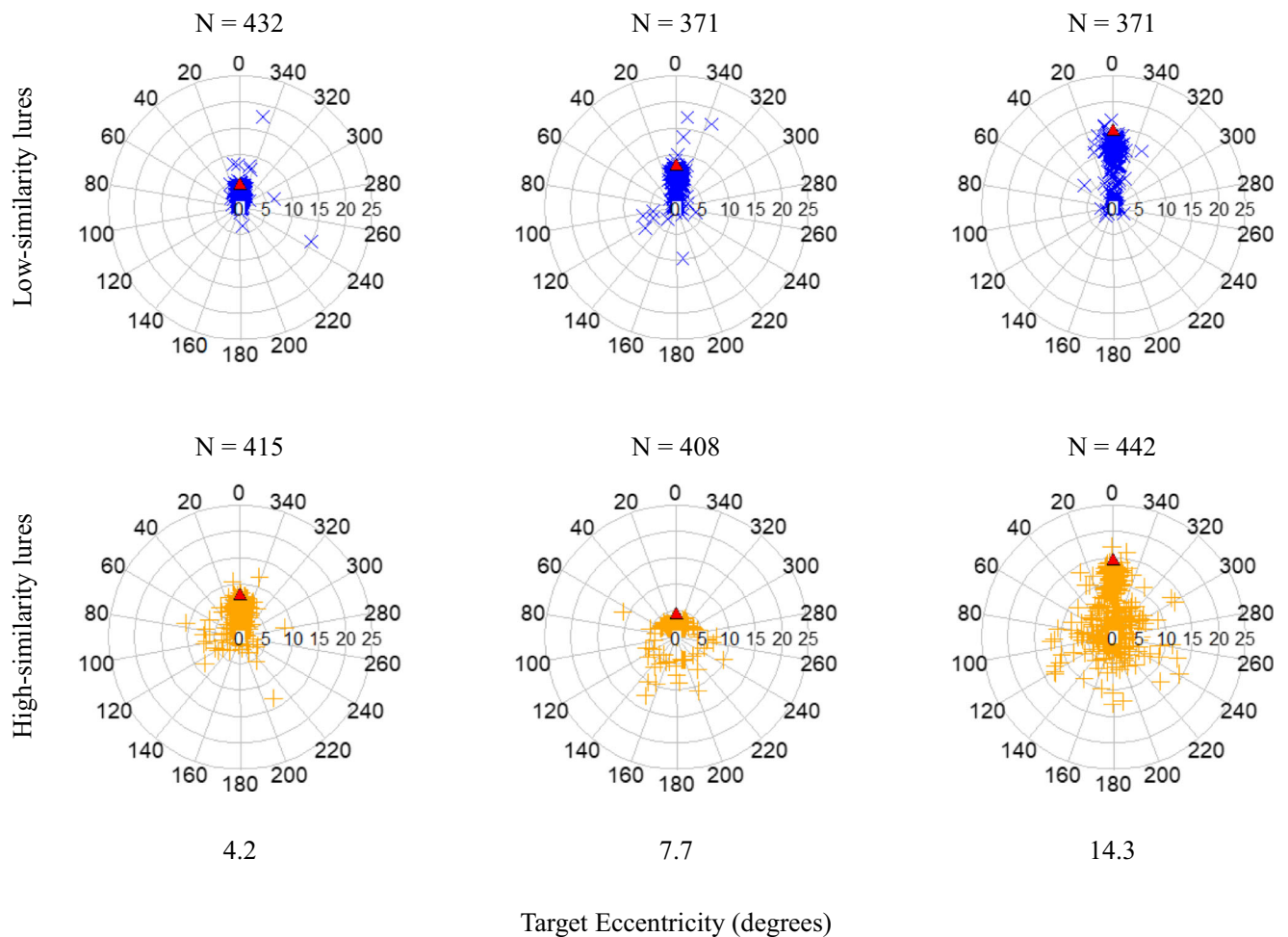
Furthermore, the number of fixations per trial was larger for high than low-similarity lures in both Experiment 1B (2.9 vs. 2.5) and Experiment 2 (2.9 vs. 2.6),  $F(1, 17) = 26.44$ ,  $p < .001$ ,  $\omega_p^2 = 0.572$ ,  $F(1, 35) = 69.93$ ,  $p < .001$ ,  $\omega_p^2 = 0.651$ , respectively. This can also be intuited from Figs. 10 and 11, as the high-similarity lure panels show a much wider dispersion of first saccades and the low-similarity lure panels show a much tighter concentration of saccades near the target location. Finally, the number of fixations also increased with set-size,  $F(2,34) = 9.88$ ,  $p < .001$ ,  $\omega_p^2 = 0.324$ ,  $F(2,70) = 37.12$ ,  $p < .001$ ,  $\omega_p^2 = 0.497$ , for Experiments 1B and 2, respectively. This is expected because the larger the number of accumulators, the larger the likelihood that more accumulators will not reach threshold before the time-out. These results are shown in Fig. 12.

## General discussion

TCS is a theory that focuses on characterizing the temporal dynamics associated with the parallel evaluation of a scene, when an observer is searching for a specific target in an efficient search task. TCS overcomes the major shortcomings of Buetti et al.’s (2016) model. Some key differences are: (i) TCS uses a variable time-out parameter to terminate parallel evaluation of information, rather than an exhaustive termination rule; (ii) under TCS, likely targets contribute linearly to RT (not logarithmically); and

(iii) evidence accumulation in TCS is driven by the magnitude of the contrast signal between visual information at a location and a target template held in mind. Further, the new architecture in TCS is consistent with the equation formulated by Wang et al. (2017) in its ability to predict heterogeneous RT using parameters from homogeneous search conditions. Finally, TCS also makes specific predictions about the sort of eye movements that ought to be observed in efficient search.

TCS is consistent with the idea that search unfolds in sequential stages. That said, TCS places more emphasis on the sort of processing that is used to discard non-targets that on stages. Indeed, at the center of TCS is a differentiation between the (limited) ability of peripheral vision to process and discard unlikely items in parallel across the scene, and the focused-attention processing whereby the eyes are directed to likely target locations, which are inspected with a high degree of accuracy, in a serial manner. It is also important to note that we do not believe parallel evaluation of the scene begins de novo after each eye movement. It is quite likely that the visual system is able to keep track of what has been inspected (and rejected) through a series of attention pointers (e.g., Cavanagh, Hunt, Afraz, & Rolfs, 2010; van Zoest, Lleras, Kingstone, & Enns, 2007). In fact, the results from eye movement analyses are consistent with this idea: when an eye movement was executed to a non-target location, participants lingered there only briefly and then moved the eyes directly to the target. Thus, it is unlikely that they re-started accumulating evidence from scratch. Rather, they might have simply finished accumulating evidence about those lure items in the far periphery that had not reached threshold, allowing them to then quickly identify the target.



**Fig. 10** Circular plot showing the angular error between the first saccade and a “perfect” saccade in Ng et al. (2018) Experiment 1, when a red target was used. The color of the symbols denotes the color of the lure in that condition (blue vs. orange lures). The figure is split into columns according to target eccentricity from central fixation (4.2, 7.7, and 14.3° of visual angle), and rows according to lure-target similarity (blue crosses: low-similarity lures; orange pluses: high-similarity lures). The 0 on the circumference marks the direction where the target is located relative to

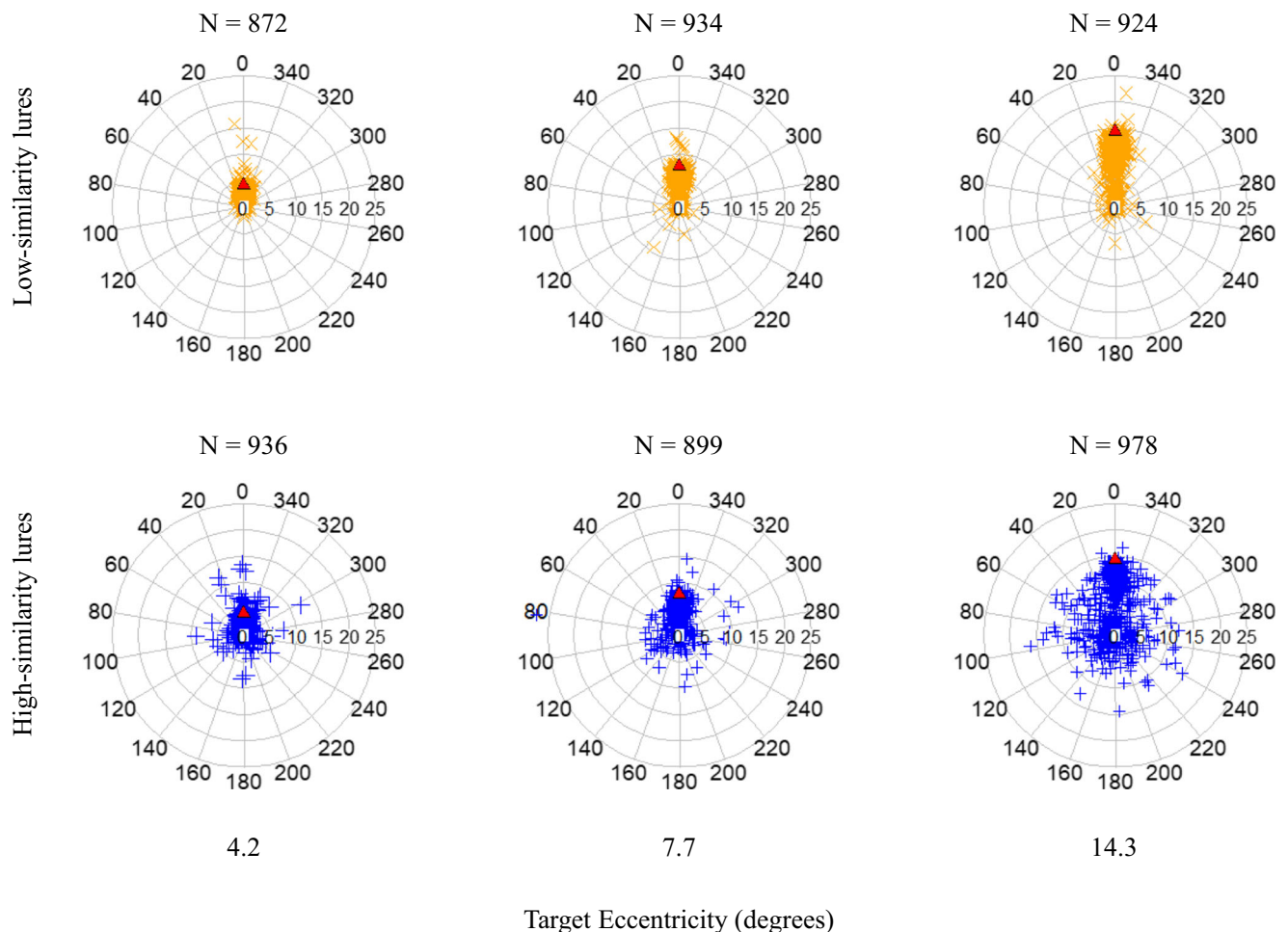
fixation (center of each plot). The radial axes correspond to the *amplitude* of the initial saccade. Each point represents one trial and the total number of initial saccades per panel is indicated above each angular plot. Note that in most panels most of the saccades seem to be directed toward the target (indicated by the red triangle on the 0 angular error line), making it difficult to visualize all these correct saccades. Initial saccades were more likely to be made away from the target when lure-target similarity was high, especially at larger eccentricities

One obvious shortcoming of TCS is that it does not have an exact analytical solution. On the other hand, a comparable advantage is that TCS is easy to simulate, so it can be used to predict performance (as in Lleras et al., 2019; Wang et al., 2017). Additionally, TCS is characterized by a set of parameters that can be studied by analogy. For instance, one can assume that some experimental factor will impact the noise in the accumulation process (e.g., crowding, Madison et al., 2018), or that another one will impact the rate of evidence accumulation (e.g., eccentricity, Wang et al., 2018), or that yet another will impact the time-out rule (e.g., instruction manipulations, Ng et al., 2018). The impact of these factors can be simulated and then one can compare whether or not human data are in line with the patterns observed in corresponding simulations.

**Relation to other visual search theories**

**Feature Integration Theory and Guided Search** Perhaps one of the most lasting and enduring legacies of Feature Integration Theory (and Guided Search) has been the emphasis on feature-specific processing as a key determinant to human performance in visual cognition tasks and as a sort of currency for visual attention. Indeed, although FIT did not propose the specific feature tuning of attention, this mechanism was proposed as one of the first improvements on FIT’s ideas (e.g., Wolfe, Cave, & Franzel, 1989; Wolfe, 1994), and it has ever since become a norm to propose that attention is given to a specific feature value and processing of that given feature is consequently changed.

A growing number of studies have challenged the view that feature-specific processing (and feature-specific modulations)



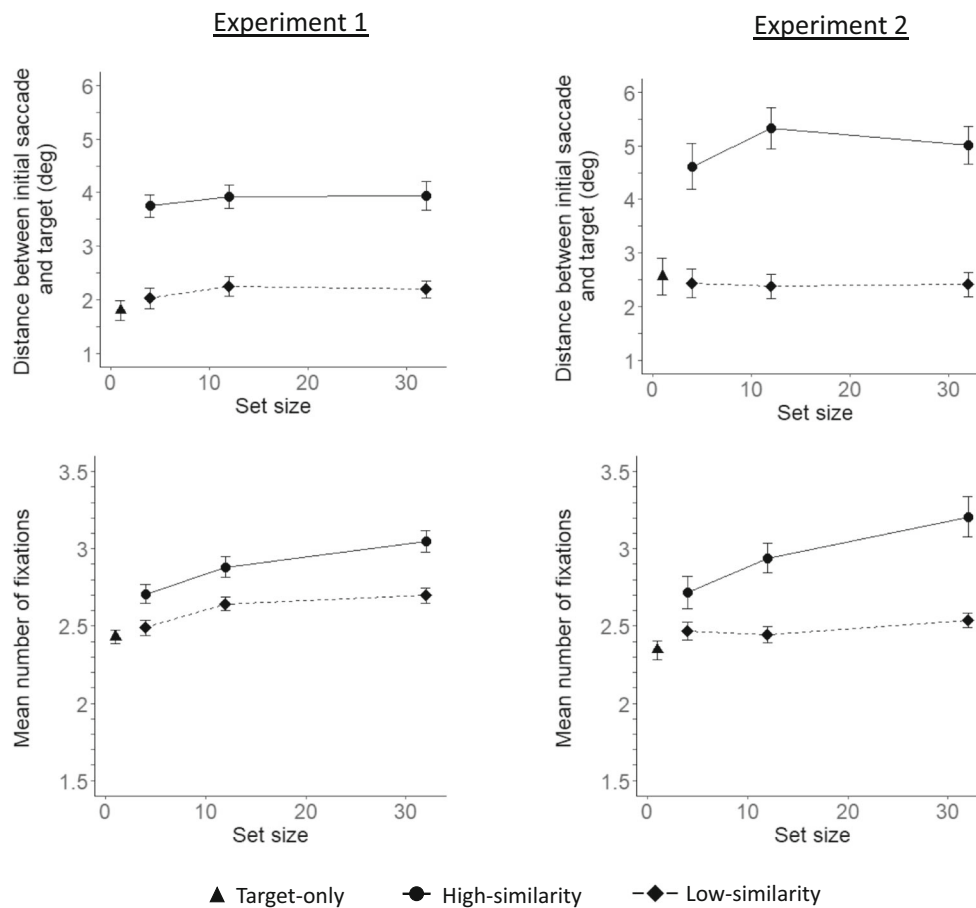
**Fig. 11** Circular plot showing the angular error between the first saccade and a “perfect” saccade in Ng et al. (2018) Experiment 2, where a cyan target was used. The color of the symbols denotes the color of the lure in that condition (blue vs. orange lures). Note that the lures used in Experiment 2 were identical to the ones used in Experiment 1 (results shown in Fig. 10), only the target changed across experiments. The figure is split into columns according to target eccentricity from central fixation (4.2, 7.7, and 14.3° of visual angle), and rows according to lure-target similarity (orange crosses: low-similarity lures; blue pluses: high-similarity lures). The 0 on the circumference marks the direction where

the target is located relative to fixation (center of each plot). The radial axes correspond to the *amplitude* of the initial saccade. Each point represents one trial and the total number of initial saccades per panel is indicated above each angular plot. Note that in most panels most of the saccades seem to be directed toward the target (indicated by the red triangle on the 0 angular error line), making it difficult to visualize all these correct saccades. Initial saccades were more likely to be made away from the target when lure-target similarity was high, especially at larger eccentricities

guides attention or somehow *improves* performance in visual search. Two decades ago, Moore and Egeth (1998) demonstrated that attending to a specific feature failed to impact the sensory quality of stimuli in a visual search task under data-limited conditions (i.e., when displays were masked). Specifically, the authors showed that attending to a specific feature (e.g., green) failed to improve processing of items containing that feature when the display was difficult to process. These are exactly the conditions where one might expect to see performance improvements, if attention to a specific feature improved processing of that feature. Furthermore, in the past decade, there has been growing evidence in favor of theories of attention where the main force driving attention is not a specific feature value but rather a value that measures a

featural difference (Becker, 2008, 2010, 2013, 2014; Becker, Folk, & Remington, 2013; Becker, Harris, Venini, & Retell, 2014; Becker, Harris, York, & Choi, 2017; Buetti et al., 2016; Lleras et al., 2019; Madison et al., 2018; Ng et al., 2018; Wang et al., 2017, 2018).

There is another reason to challenge theories positing the existence of feature maps (e.g., Treisman & Gelade, 1980; Wolfe, 1994): the implausibility of the number of neurons required for such maps. According to Maunsell and Treue (2006), this number would far exceed the total number of neurons that guide attention in the brain. In fact, most of the color-coding neurons in the early visual system code color contrasts, not specific colors in isolation. This color contrast computation forms the basis for the opponent-color system



**Fig. 12** Eye-movement data from Ng et al. (2018) Experiments 1B (left) and 2 (right) as a function of set size and lure-target similarity. **Top row:** Mean distance between the landing location of the initial saccade and the

target location (in degrees of visual angle). **Bottom row:** Mean number of fixations as a function of set size. Error bars indicate the standard error of the mean

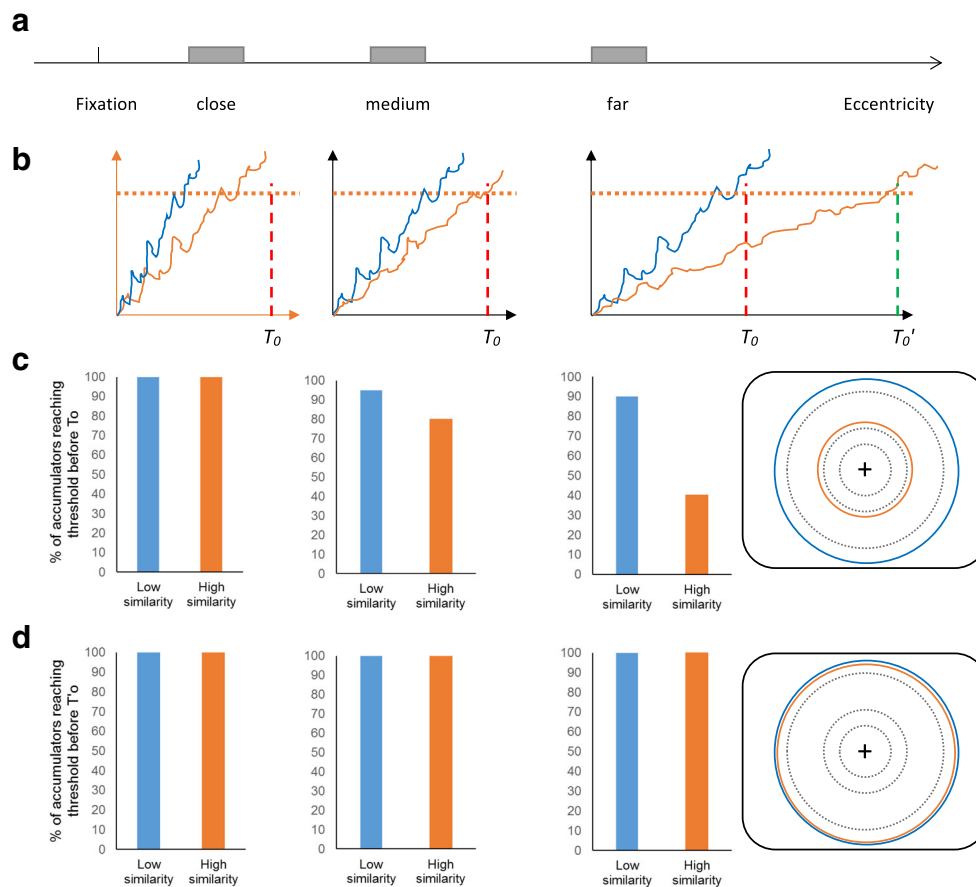
(e.g., De Valois et al., 2000; Hubel & Wiesel, 1967) and is present as early as color-coding neurons in the LGN. Most visual features are also coded in opponent fashion. Thus, it makes sense to build a theory of attention based on the computation of feature comparisons (as opposed to absolute feature values). When observers are searching for a specific target, the comparison is made between the visual information in the scene and the target template in mind. When there is no goal in mind, as when free viewing a scene, the feature comparison might more likely be performed on a spatial, local manner (e.g., Itti & Koch, 2000), such that attention will be attracted to areas containing feature discontinuities. In fact, it is possible that both types of information impact the deployment of attention either jointly or over different time scales, as suggested by the research by Donk and colleagues (e.g., Donk & Soesman, 2010; Donk & van Zoest, 2008; Itti & Koch, 2000; van Zoest, Donk, & Theeuwes, 2004).

TCS also represents a departure from both feature-specific theories of visual processing because its goal is not to produce a representation of the visual properties of distractors at each

location in the visual field, but rather to measure how different each item is from the target.

Finally, one under-appreciated factor in the visual search literature is the extent to which search efficiency is decreased by the presence of distractor heterogeneity, or conversely, the extent to which efficient search is made more efficient by the homogeneity of the displays. Treisman and Gelade (1980) made famous the observation that finding a target defined as a conjunction of features takes much longer (and is less efficient) than search for a target defined by a unique feature. In many conjunction search experiments since, the observation is simply made that conjunction search produces larger slopes than the corresponding feature search task. The confound, however, is that the efficient search displays tend to be inherently more homogeneous (i.e., a red vertical line amongst green vertical lines), whereas conjunction searches tend to be tested in heterogeneous displays, by necessity (i.e., a red vertical line amongst green vertical lines and red horizontal lines). This confound is problematic because, as demonstrated by Lleras et al. (2019), homogeneity facilitation provides very





**Fig. 13** Schematic illustrating the interplay between evidence accumulation, eccentricity, lure-target similarity and the time-out parameter. **Panel A:** diagram illustrating the position of items with respect to fixation, corresponding to the three eccentricities (as used in Ng et al., 2018). **Panel B:** example of stochastic evidence accumulation. The horizontal dotted line (in orange) represents the non-target decision threshold. The vertical dashed line (in red) in the first two panels represents a short time-out parameter ( $T_0$ ). In the third panel, the second vertical dashed line (in green) represents a much longer time-out parameter ( $T_0'$ ). When evidence accumulation (the solid lines) reaches threshold, the corresponding item is discarded. For simplicity, we only illustrate one accumulator per condition and the time-out parameter is shown as being on a fixed location on the x-axis, but it is important to remember that the

time-out is a time interval that is reset every time a lure hits threshold, so its onset is relative to the time lures hit threshold. **Panel C:** the bar graphs illustrate the average percentage of accumulators reaching threshold before the short time-out ( $T_0$ ) at each eccentricity. Due to the slower accumulation rate for high similarity lures, this percentage decreases much faster for these types of lures, as eccentricity increases. **Panel D:** the bar graphs illustrate the average percentage of accumulators reaching threshold before the longer time-out ( $T_0'$ ) at each eccentricity. With sufficiently long time-outs, all accumulators can eventually reach threshold during efficient search. That said, the data from Ng et al. (2018) presented above suggest that observers tend to spontaneously operate with relatively smaller time-outs (more like  $T_0$  than  $T_0'$ ), which makes it look like there is an FVF whose size is determined by target-distractor similarity

large improvements in search efficiency, particularly with simple geometric shapes, the kind of stimuli typically used in conjunction search studies. Compared to homogeneous displays, heterogeneous displays are 1.8 times less efficient (in logarithmic space). Importantly, because Equation 2 provides a very good predictive fit for processing times in heterogeneous displays, one can infer that the manner in which heterogeneous displays are processed are not fundamentally different from how homogeneous displays are processed (aside from the degree of homogeneity facilitation). In other words, one should always expect heterogeneous search displays to be much less efficient to process. And, to demonstrate that conjunction search performance is qualitatively different from

feature search, one should not compare the performance to efficient, homogeneous displays. Rather, one should first evaluate the logarithmic slopes for each of the contrasts that make part of the conjunction search display (i.e., the slope for red vertical target amongst green vertical distractors and the slope for red vertical target amongst red horizontal distractors). Equation 2 can then be used to predict the range of expected processing costs when all those target-distractor pairs are simultaneously present in the scene, that can be then plotted against observed human data in the conjunction search condition (as in Fig. 2). Deviations from a linear fit would indicate that conjunction search requires a different underlying architecture than the parallel evaluation reflected in Equation 2.

**Attentional engagement theory** The attentional engagement theory by Duncan and Humphreys (1989, 1992) most famously contributed to the literature the proposal of the “search surface” as well as that of “spreading suppression.” The search surface summarizes the idea that a search slope is the product of two separate factors: target-distractor similarity and distractor-distractor similarity. As target-distractor similarity increases, the search slope increases and as distractor-distractor similarity increases, the search slope decreases. Spreading suppression is a mechanism whereby identical distractors tend to group together and therefore get rejected as a group, instead of as individual items, thus facilitating target selection. TCS challenges both of these core tenants.

Regarding the concept of the search surface, TCS proposes that there is a marked qualitative discontinuity such that for low levels of target-distractor similarity, efficiency is logarithmic while for higher levels, efficiency is linear. Thus, there is no smooth continuous change in search slope values; this discontinuity reflects the fact that there are different processing mechanisms at play when rejecting non-target items: a parallel, unlimited capacity, resolution-limited mechanism and a serial, limited capacity, potentially unlimited resolution mechanism. In TCS terms, the transition point between the two mechanisms is determined by a combination of factors. First, the average evidence accumulation rate, which is a factor of target-distractor similarity, eccentricity, and size (Wang et al., 2018). Second, the noise in the accumulation process, which is a factor of the quality of the peripheral representations, and thus is impacted by factors like crowding (Madison et al., 2018) and stimulus size (Wang et al., 2018). Third, the time-out parameter, which is sensitive to individual differences and top-down preferences and instruction manipulations (Ng et al., 2018).

The concept that distractors in efficient search are discarded “en masse” or as a group stands in contrast with the proposal that each distractor is processed and rejected separately, a core component of TCS. The logarithmic relationship between RT and set size emerges because of the parallel and stochastic processing of multiple lures at the same time. A group rejection of distractors is inconsistent with such relationship. In addition, as demonstrated by Wang et al. (2017), the best way to account for RTs in heterogeneous displays is to assume that all items are initially processed simultaneously and that individual rejections are determined by each item’s degree of dissimilarity to the target (the more dissimilar lures reaching rejection sooner).

Duncan and Humphreys’ (1989) noted that there was a homogeneity facilitation effect in visual search that inspired their spreading suppression account. Wang et al. (2017) and Lleras et al. (2019) quantified the magnitude of that homogeneity facilitation effect in efficient visual search (indexed by  $\beta$  in Fig. 2 and Equation 2) in a manner that was independent

from lure-target similarity (indexed by  $D_j$  in Equation 2). The existence of this effect implies that the processing efficiency of lures is not independent of other information in the display: displays that contained all identical items were processed more efficiently (by a multiplicative factor of the logarithmic function) than what would have been expected if those lures had been surrounded by non-identical lures. Note that Equation 2 is based on the assumptions that each item is evaluated in parallel, that each item needs some time to be rejected, and that identical neighboring items will speed up the time required to reject those items. The success of Equation 2 in predicting performance across subjects and in novel display arrangements provides a very strong validation of the underlying assumptions of TCS.

From a more general stance, it should be noted that TCS does not include a location-based suppression mechanism, whereby non-target items/locations are suppressed, after they have been categorized as non-targets. Neither does it include a feature-based suppression (or boosting) mechanism. This is not to say that such mechanisms do not exist. Rather, at this point, there is no need to incorporate suppression to account for human data in efficient search.

**Eye-movement based theories** Several theories have focused on the role eye movements play during visual search.

The Target Acquisition Model (Zelinsky, 2008) was developed as a computational model to predict where the eyes would go in a scene. The model uses a similarity score for each region in the scene based on the visual similarity between that region and the target template to form a target-distractor similarity map. Then, a threshold is iteratively raised along this target map with the goal of eliminating locations that are unlikely to contain the target. This leaves only a few locations that have very high probabilities of containing the target, at which point an eye movement is executed toward the most likely location. TAM thus shows several similarities with TCS. For one, both theories emphasize the role of parallel rejection of non-target locations, even though the two theories implement it in different ways. A second important similarity between TAM (as well as its most recent instantiation MASC; Adeli et al., 2017) and TCS is the observation that the quality of the signal (that is the input to the visual system) is strongly dependent on eccentricity (see Rosenholtz et al., 2012; Zhang, Huang, Yigit-Elliott, & Rosenholtz, 2015). Indeed, in TAM, the visual input is progressively blurred as a function of distance from fixation to mimic the visual acuity drop-off in peripheral vision.

The differences between TAM (and MASC) and TCS are important, however. TAM proposes a single thresholding: all activation in the map is judged to be either larger (possible target) or lower than this one threshold (distractor rejection). This single threshold implies that the duration of the process to discard below-threshold distractors is constant or negligible.

Further, this duration is *insensitive* to the similarity relationship between each distractor and the target. Yet, we now know that the lure rejection process, even if it unfolds in parallel, takes different amounts of time depending on each lure's degree of similarity to the target, even when multiple types of lures are simultaneously present in the display (Buetti et al., 2016; Lleras et al., 2019; Wang et al., 2017). Finally, it should be noted that both TAM and MASC have an explicit representation of the target-distractor similarity of each item or region in the display. TCS does not need to represent (or even compute) this variable. The target-contrast signal drives the speed of the evidence accumulation without needing to be represented or remembered, nor is it used to prioritize information among likely candidate locations. Furthermore, TDS maps represent activations across the visual field, with a precision that is only limited by the initial blurring of the input image (meant to mimic the loss of visual acuity in the periphery). Thus, TDS maps do use the TDS signal to rank all likely candidate locations, implying that the visual system is capable of accurately computing those signals, even at the higher end of the similarity scale. TCS, instead, proposes that the visual system has only limited ability to accurately judge the similarity relation between a candidate in the periphery and the target template.

Finally, there are important advantages that MASC has over TCS. For one, MASC is biologically plausible, inspired by the saccade generation mechanisms in the superior colliculus. Second, MASC is a computational model that is image-computable: one can provide an image and MASC can make saccade predictions. Furthermore, it can be used to model both free-viewing and visual search tasks by relying on two distinct methods to compute the TDS map: a saliency map in free-viewing and a true TDS map when features about the target are generally known. That said, it seems to us that MASC and TCS are not fundamentally incompatible. It appears quite plausible that the temporal dynamics of peripheral processing that are the focus of TCS could be implemented into the parallel processing stage of both TAM and MASC. This observation applies more generally at most theories of search that have neglected to model the temporal dynamics of the parallel rejection process in peripheral vision.

A second prominent theory of eye movements and visual search was recently put forward by Hulleman and Olivers (2017), who proposed that performance in inefficient visual search is mostly determined by the size of the functional viewing field (FVF, the area around fixation from which useful information can be obtained without eye movements). The authors proposed that the width of this area is not fixed – it decreases as target-distractor similarity increases. As a result, when target-distractor similarity is low, the width of the FVF is large and many items can be processed and discarded as non-targets in parallel. As target-distractor similarity increases, the width of the FVF decreases, reducing the number

of elements that can be processed in parallel and increasing the need for eye movements to find the target. When target-distractor similarity is high, the area of the FVF is small and only a small number of items (sometimes only one) can be processed and discarded as non-targets during a fixation.

This theory is in many ways compatible with TCS, which was focused mostly on efficient search. For example, Hulleman and Olivers (2017) did not specify any processing cost (or any architecture) for the parallel stage where items are processed within the FVF. Thus, a TCS-style architecture could easily be incorporated into their model to make more precise predictions for the temporal costs incurred in search. In fact, the time-out parameter in TCS has functionally a *spatial* analog. Since evidence accumulation rates decrease with eccentricity, for any given time-out parameter, lures of lower similarity will be able to be processed over a larger area around fixation than lures of higher similarity. This is illustrated on the right-hand side of Fig. 13c: for a small time-out parameter, low-similarity lures at all three eccentricities can be rejected in parallel, producing performance that would be consistent with an FVF that encompasses all eccentricities. In contrast, for the same time-out, high-similarity lures tend to fail to reach threshold at the farthest eccentricity, thus producing performance that would be consistent with a smaller FVF that only encompasses the inner and medium eccentricities.

One shortcoming of Hulleman and Olivers' theory, though, is that it is not clear how it would be adapted to heterogeneous search displays. As proposed, the magnitude of the FVF is determined by distractor-target similarity, which in a way assumes that all distractors are identical or at least share the same overall level of target-distractor similarity. It is impossible, however, to make specific predictions for what this theory would propose when items of different levels of similarity to the target are simultaneously present in the display. Would the FVF be determined by the item with the highest similarity? Or would the average level of similarity in the display determine the FVF? It is unclear. Notice, however, that TCS does make specific predictions for each of these instances. For lure-heterogeneous displays, Equation 2 makes specific predictions for how long processing times will take. For displays combining lures and candidates, we know that candidates will not impact stage-one processing times, only the lures. For displays containing different levels of candidates (some more similar to the target than others), TCS makes the prediction that what matters is the overall number of to-be-inspected locations, not so much the level of similarity of any one candidate to the target (Ng, et al., [submitted](#)).

In sum, although TCS shares a number of similarities with the FVF-oriented proposal of Hulleman and Olivers, we view TCS as a theory that is consistent with their proposal but has the added benefit of making more specific predictions. This is not necessarily a criticism of Hulleman and Olivers' theory since their goal was not so much to propose an exact architecture of

visual search, but rather to sketch out the skeleton of what a good theory of visual search (that more fully embraces ocular constraints) ought to look like. That said, TCS does not impose an FVF of any given size (see also Rosenholtz, 2017 for a similar idea). As shown in Fig. 13, the combination of variations in the accumulation rates and the time-out parameter produce results that are consistent with an FVF that is changing in size as a function of target-distractor similarity. The fact that observers end up processing the scene in what looks like different spatial extents is an emerging property of TCS. Note too, that, as indicated by the eye movement analyses, participants who spontaneously behave as in Fig. 13C (with a smaller FVF for high similarity lures than for low similarity lures) can switch to behaving as in Fig. 13D (where FVF would include all locations in the display), with instruction manipulations. This is easily accommodated by changes in the time-out parameter in TCS and would perhaps be harder to accommodate under Hulleman and Olivers' proposal.

As a final note, it is also worth mentioning that various aspects of TCS bear similarity to models of saccade generation. Indeed, models like CRISP (Nuthmann, Smith, Engbert, & Henderson, 2010), and the model by Laubrock, Cajar, and Engbert (2013) propose that peripheral analysis of the scene occurs in parallel across the entire scene. These models also include a foveal analysis of the information currently being fixated that takes place concurrently. Another commonality with these models is the idea that evidence is being accumulated in random walk fashion towards a decision to generate a saccade and the timing of that saccade is determined by a saccade timer. The saccade timer can be shortened (saccade towards periphery occurs sooner) or elongated (eyes remain at the current location) as a function of factors like the amount of foveal activation (which inhibits saccades by increasing the saccade timer) or the amount of peripheral activation (which disinhibits saccades by shortening the saccade timer). The time-out parameter in TCS is conceptually similar to the saccade timer in that it can prompt decisions to move the eyes away from the current fixation (although, in efficient search tasks, eye movements are not necessary and the task can be completed without them). These models incorporate factors that can elongate the current fixation, as a function of the difficulty of the processing at fixation, for example, or the duration of recent saccades (Trukenbrod & Engbert, 2014). Thus, these timers and our time-out are both sensitive to processing demands. It seems that the architecture of these saccade generation models could very well be integrated with TCS by having TCS drive the dynamics of the peripheral processing, and including some of the factors that impact the time-out as potential factors modulating the duration of the saccade timer.

**A final note regarding target-distractor similarity theories of attention** TCS distinguishes itself from many similarity-based

models of attention (e.g., Bundesen, 1990; Duncan & Humphreys, 1989, 1992; Zelinsky, 2008) on two fronts. First, TCS posits that the evidence accumulation at each location is driven by target-defining information. The evidence accumulation process consists of an evaluation of the extent to which properties at each location *differ* from the set of properties that define the target. Properties of the distractor that are not present in the target are ignored. This is a departure from most target-distractor similarity signals (TDS), which typically involve a convolution that compares all target features to all distractor features. As a result, traditional TDS signals are symmetrical: they are equally driven by properties in the distractor as they are by properties in the target. In TCS, the target-distractor relationship is asymmetric, such that the contrast signal of a distractor stimulus X to a target stimulus Y is not the same as the contrast signal of a distractor stimulus Y to a target X. This allows TCS to account for the well-known search asymmetry effect first studied by Treisman and Souther (1985): the finding that locating a target Q amongst Os is efficient, whereas finding a target O amongst Qs is not. When the target is a Q, the contrast signal computed at each location will evaluate to what extent the object evaluated is different from the round shape of the Q as well as different from the straight line of the Q. As a result, all Os in the display produce a large contrast signal along the second of the evaluated properties because it is clear that Os do not have any straight line in them. So, finding a Q amongst Os can be done in parallel (to the degree that the straight line can be resolved in peripheral vision). In contrast, when the target is an O, the contrast signal associated with the processing of the Q distractors will be very low. Indeed, all location contain a round shape. Because “roundness” is the only property of the target, it is also the only property that will be queried in the evidence accumulation process. As a result, the search will be inefficient because no accumulators will reach non-target threshold.

Second, TDS-based theories also propose that early vision computes a TDS map of the scene (often with no time cost) and use activation values on that map to drive attention and the eyes. TCS proposes that computing an accurate TDS signal is not necessary to understand efficient search, and in fact, given the processing limitations in peripheral vision, accurate TDS signals might not even be computable by the visual system. Thus, because TCS does not represent the precise similarity relation between candidates and target, likely target locations (i.e., those locations that did not reach threshold before the time-out) should all be equally likely selected by overt or covert attention. We have provided evidence in favor of these lack-of-prioritization in a separate paper (Ng, et al., [submitted](#)). In that paper, displays contained two different types of candidates. Some were extremely similar to the target, while the others were less so. In spite of the large differences in target-candidate similarity between the two types of



candidates, we failed to find any evidence that the more similar candidates were prioritized over the less similar candidates. Indeed, performance in mixed displays were perfectly predicted by a model that assumed random examination of the candidates, irrespective of target-candidate similarity.

## Conclusion

To conclude, a new architecture for early parallel peripheral processing is proposed to predict performance in efficient visual search conditions where participants have a fixed target in mind. The key insight is that this model is based on the computation of a contrast signal between the internal target template representation and visual information at each location in the scene. In doing so, the model moves away from the tradition of feature-specific attentional models and even from visual-contrast based models (e.g., Itti & Koch, 2000). The model also emphasizes that even for extremely fast and efficient visual search tasks, what determines how focused attention is deployed in a scene is neither the result of an automatic computation of what visual signals are present in the scene, nor even the computation of local visual contrasts. Instead, focused attention is deployed as the result of an *active* computation, where the visual system computes a difference score in goal-oriented fashion between an internal representation and the visual stimulation. The new model is a significant improvement of Buetti et al.'s (2016) initial model and thus is capable of explaining a wider range of results (like the occurrence of randomly-directed eye movements in efficient search). Finally, we hope to have demonstrated that the relative ease with which the TCS architecture can be modeled is a boon for both testing as well as developing theoretical predictions that can be precisely tested.

**Authors note** The theory was developed by S.B. and A.L. The simulations were run by Z. W. The eye-movement experiments were designed by G.J.P.N., S.B., and A.L. Eye-tracking data were analyzed by G.J.P.N. All authors contributed to the writing. This project was partially supported by an NSF grant to SB (award number BCS 1921735).

**Open Practices Statement** The data and materials for the experiments are available at the Open Science Framework <https://osf.io/pve8d/>, and both of the experiments reported here were pre-registered.

## Appendix 1: Why Accumulate Contrast?

Typically, the way contrast has been understood to impact visual attention is by using it to denote a visual difference between two nearby regions in the scene (Itti & Koch, 2000). The rationale is that local spatial contrast indexes

conspicuity and that conspicuity attracts attention (e.g., Northdurft, 1991, 1992). Although this approach has had some success in predicting eye movements in free-viewing tasks (Itti & Koch, 2000; Peters, Iyer, Itti, & Koch, 2005), visual conspicuity (or salience) has been shown to play only a minimal role in goal-directed visual search (e.g., Clarke, Dziemianko, & Keller, 2014; Donk & Soesman, 2010; Donk & van Zoest, 2008; Itti & Koch, 2000; Tatler, Hayhoe, Land, & Ballard, 2011; van Zoest, Donk, & Theeuwes, 2004). In comparison, the contrast computed by TCS does not compare neighboring regions but rather visual information at a location to the internal representation of the target template.

Below, we present a series of theories and findings that challenge the idea that visual search for a specific feature is achieved by tuning attention to that specific feature. One early proposal along those lines was put forward by Navalpakkam and Itti (2007), who demonstrated that when searching for a target defined as the most extreme feature along a continuum (e.g., searching for a purple target amongst blue distractors, the purple being the reddest element in the display), because of the width of neural tuning curves, it would be more efficient to “boost” a more extreme feature along the blue-violet axis such as red, than to boost (or tune to) the target feature (here, purple). This follows because the boosting of any specific feature actually results in the boosting of all features along the entire tuning curve of the tuned-to feature (see also, Lee, Itti, Koch, & Braun, 1999; Scolari & Serences, 2010 for other optimal tuning accounts). In other words, if the distractor feature value is sufficiently close to the target feature, boosting the target feature also results in the unintended boosting of the distractor feature. Navalpakkam and Itti (2007) proposed a better strategy in terms of improving signal-to-noise ratio: they proposed that an *optimal* feature to boost would be a feature more extreme than the target feature itself along the distractor-target feature continuum. In that way, one can theoretically find an optimal feature that is sufficiently far from the distractor feature so as not to boost it, while being sufficiently close to the target feature to boost the target feature, thereby improving the signal-to-noise ratio. The key insight here was that, even in theories that propose feature-based tuning of attention, the optimal strategy is not to boost the target-feature itself, but rather, a more extreme (and actually *unseen*) visual feature. This insight implies that the attention system computes the featural difference between the shown target and distractor stimuli and an *unseen* feature. This difference signal is then used to optimally tune perceptual processing and efficiently guide attention to the target.

More recently, target-feature and optimal-feature boosting accounts have been challenged by findings suggesting that attention is not tuned to a specific feature but rather to a

*feature relation* (e.g., Becker, 2014). In the example above, this is equivalent to saying that attention is neither tuned to purple (target-feature boosting) nor to red (optimal-feature boosting), but rather to “redder” because the target is uniquely defined as the element containing the most amount of red in the scene. As a result, attention will be guided to the redder item in a display, irrespective of the specific feature values of the items in the scene. In our example, attention would be captured by a purple stimulus amongst blue distractors, just like it would be by a red stimulus amongst purple distractors because both stimuli are “redder” than their contextual distractors. The “relational” account of attentional tuning has received substantial empirical support (e.g., Becker, 2008, 2013, 2014; Becker, Folk, & Remington, 2013; Becker, Harris, Venini, & Retell, 2014; Becker, Harris, York, & Choi, 2017). A detailed discussion of the merits of that theory is beyond the scope of the current paper, but we do want to highlight that a key insight in this theory is the idea that attention is not guided by the salience of a specific feature (boosted or not). Rather, attention is guided by the output of a *comparison process* that first scores items in the scene along an internally decided feature direction (in our example, the blue-to-red continuum) and then selects the item that scores the highest in this axis irrespective of its underlying (absolute) feature value. In sum, the relational theory is consistent with the idea that attention is guided by a contrast signal along a goal-defined feature axis (in a manner that is *independent* of the feature values themselves that are used to compute that difference score).

A different set of proposals also focused on computing differences between items such as the work on SDT “relative coding” models (e.g., Palmer, Verghese, & Pavel, 2000; Rosenholtz, 2001). These theories proposed that items in the display would be compared against each other and this comparison, in turn, would drive attention. This work was partly inspired by the work of Northdruff (1991, 1992, 1993) who demonstrated that local differences were more important in search than actual feature values themselves. Rosenholtz (2001) introduced the “relative coding with reference” model, in which the majority of the time, items are compared to other items in the display and in some fraction of the time items are compared to the target template in mind. TCS follows this tradition but in a more extreme fashion proposing that in goal-directed search, all comparisons are to the target template.

It is also worth noting that in computer vision, most models no longer use a “target template” representation to guide processing. Rather, these models use target discriminators that basically differentiate the target from other objects. These target discriminators are, in fact, computing something akin to the target-contrast computation proposed by TCS by trying to extract unique differences that help to categorize the target as different from the other elements in the scene.

Finally, we should note that the data in Buetti et al. (2016) also casted doubt on a naïve interpretation of target-feature boosting. Indeed, as shown in Fig. 1, there is a systematic increase in the time it takes to process displays as the lure-target similarity increases. For instance, displays containing one red target and several yellow distractors are processed faster than displays containing one red target and orange distractors. The issue here is that feature boosting is typically implemented as an increase in the accumulation rate for the boosted feature (or attended signal, more generally). That is, an increase in the accumulation rate ought to result in the processing of the item containing that feature to be accelerated, thus, take less time. Unless additional assumptions are made, a target-feature boosting account would necessarily predict that a display with only one boosted item would necessarily take longer to process on average than a display where all items are boosted. For instance, when searching for a red target among orange distractors, both target and distractors will receive some red-related boosting. In contrast, when searching for a red target among yellow distractors, only the target will receive red-related boosting. As a result, the red and orange display ought to be processed faster than the red and yellow display. In sum, target-feature boosting would predict that as target-lure similarity increases, processing efficiency for the entire display increases, thereby *decreasing* RTs. Yet, our findings indicated that overall RT *increase*, not decrease, with increasing similarity. Thus, whereas target-feature boosting might be an appealing and intuitive concept, there are a number of reasons to doubt that it actually plays a role in efficient search.

In sum, the foundational idea behind TCS is that visual search performance is determined by the computation of a difference signal between the internal representation of the target template and the stimuli in the visual scene. This proposal is grounded on similar ones in the literature that have moved away from feature-oriented processing accounts into more relational accounts where the visual system is actively computing a difference signal to guide attention. It is also worth noting that the visual system is a large contrast computing system (see Adelson & Bergen, 1991). Visual contrasts are computed even before visual information leaves the eyeball (Burkhardt & Fahey, 1999). Contrast is the currency of all lateral inhibition signals (Solomon, Sperling, & Chubb, 1993), of the color system (Ekroll & Faul, 2012), of boundary detection (Mareschal & Baker, 1998), and perhaps even of visual categorization (Macé, Thorpe, & Fabre-Thorpe, 2005). Thus, we view TCS as largely consistent with various properties of the architecture of the visual system, such as parallel and unlimited capacity processing of visual stimulation, and the known processing limitations in peripheral vision. In addition, the computations performed by TCS are the sort of computations performed by the visual system, such as such as contrast computations and accumulation of evidence over time.

## Appendix 2: General Methods from Ng, Lleras, and Buetti (2018)

Two experiments were carried out using the same general method. In the first experiment, participants searched for a red triangle target, while in the second one, they searched for a cyan semi-circle target. A different target was used in the second experiment to demonstrate that lure-target similarity, rather than the stimuli themselves, modulates the evidence accumulation process (as Buetti et al., 2016, Experiments 1A and 1B). Both experiments were pre-registered on Open Science Framework (Ng, Buetti, & Lleras, 2016, <https://osf.io/pve8d/>).

### Inclusion-exclusion criteria

The following inclusion-exclusion criteria were pre-registered on OSF (Ng, Buetti, & Lleras, 2016, <https://osf.io/pve8d/>). First, participants who completed less than 90% of the experiment were excluded. These were usually participants who had trouble fixating on the center of the screen or took long breaks. Then, those with less than 90% overall accuracy were excluded. Participants who had overall mean RTs that were 2 standard deviations away from the group mean were then excluded. Participants were also excluded on the basis of eye movements (more elaboration regarding this will be provided in the next section). Lastly, only the first 18 (Experiment 1) or 36 participants (Experiment 2) that met these criteria were included in the analyses.

### Participants

Undergraduate students from the University of Illinois at Urbana-Champaign participated in the experiments in exchange for either course credit in a psychology class or an \$8 monetary compensation. All participants had normal or corrected-to-normal vision, and were determined to be non-colorblind by using the Ishihara color plates. The sample size for both Experiment 1 and Experiment 2 were determined from a power analysis from a pilot experiment (Experiment 1A, Ng et al., 2018).

In Experiment 1, we ran 26 participants (21 females, mean age = 20 years). All participants had overall accuracy levels above 90%. Three participants were removed for excessive eye movements in the fixed-viewing condition (15% of the trials) and two were removed for having overall RTs that were 2 standard deviations away from the group mean. Out of the remaining 21 subjects, there were two who did not have any eye movements in at least one experimental condition. The first subject did not make any eye movements in the target-only condition; the second did not make any eye movements in two out of three setsizes (4 and 32) in the low-similarity

condition, and one out of three (set size 4) in the high-similarity condition. We reasoned that it made little sense to include these participants in the analyses since it would result in missing values and an inability to conduct analyses of variance. We thus removed them, leaving us with 19 participants. It should be noted that this criterion was not included in our pre-registration as we did not foresee it. Since we declared that we will only use 18 participants in our pre-registration, we only took the first 18 participants from the non-excluded participants.

In Experiment 2, the two viewing conditions (free-viewing and no eye movements conditions) were counterbalanced to evaluate effects of practice, with each order having a target sample size of 18 participants (Ng et al., 2018). We ran 54 participants (30 females, mean age = 20 years). We excluded nine participants who did not complete at least 90% of the experiment, five participants who made eye movements in more than 30% of the trials in the fixed-viewing condition, and two subjects who had overall RTs beyond two standard deviations from the mean. Of the remaining 38 participants, there was one participant who did not make any eye movements in the target-only condition. As before, we removed this participant, and analyzed the data from only the first 36 remaining participants that met our inclusion criteria.

### Stimuli and apparatus

The lure stimuli were blue circles and orange diamonds. In Experiment 1, the target was a red triangle and in Experiment 2, the target was a cyan semicircle. The stimuli were distributed across a circular array with 36 locations. The array was made up of three concentric rings spanning 4.17, 7.73, and 14.3° of visual angle. This circular array was utilized to minimize crowding (Bouma, 1970; Madison et al., 2017; Pelli & Tillman, 2008). The distribution of the stimuli across the array was pseudo-random, with the constraint that all stimuli will be distributed equally across the four quadrants of the screen. The quadrants and the concentric rings were used only in the creation of the search array and were not visible to participants.

The search arrays were first created using Psychtoolbox for MATLAB (Brainard, 1997). These were then exported as .png image files to be used with Experiment Builder for use with the EyeLink 1000 eye-tracker. The search array was presented on a black background on a 22-in, (400 × 300 mm) cathode ray tube monitor at a refresh rate of 85 Hz and a resolution of 1,024 × 768 pixels. Each stimulus item subtended .833° of visual angle. Participants viewed the display from a distance of 59 cm with their head on a chin rest. Eye movements were recorded using a tower-mounted EyeLink 1000 eye-tracker at a 1,000 Hz resolution.

## Design and procedure

The three independent variables were: lure type (high- or low-similarity), set size (1, 4, 12, 32), and viewing condition (fixed-viewing vs. free-viewing). There were 14 instead of 16 cells as there was no lure type for the target-only condition [ $2 \times (2 \times 3 + 1) = 14$ ]. There were 32 trials per cell, with a total of 448 trials in the experiment. Viewing condition was blocked (i.e., all trials within the same block were either fixed-viewing or free-viewing), and all trials were randomized within each block. The fixed-viewing block was included to demonstrate that logarithmic search slopes are not artifacts of eye movements but rather an inherent property of stage-one processing (Ng et al., 2018). In Experiment 1, participants always started with the free-viewing block first. In Experiment 2, half the participants started with the free-viewing block first and the other half started with the fixed-viewing block first. We further subdivided each block into eight mini-blocks with 28 trials each. The trials in each mini-block had the same distribution of conditions (i.e., two trials of each condition).

Before the start of the experiment, participants underwent a fixation training exercise (Guzman-Martinez, Leung, Franconeri, Grabowecky, & Suzuki, 2009). In this fixation training, participants viewed a rapidly alternating display that was made up of an equal number of randomly distributed black and white pixels. On each alternation, the color of the pixel reversed such that all white pixels now became black and vice versa. Participants were instructed to maintain their fixation on the center of this display. As long as their eyes were stationary, this rapidly alternating display appeared as a uniformly grey display due to perceptual averaging. Any eye movements caused the perceptual averaging to be disrupted and resulted in the emergence of a random black-and-white dot pattern. This fixation training lasted approximately 2 min.

In the free-viewing block participants were instructed to fixate the central cross at the start of each trial. They were then free to move their eyes once the search array was presented. In the fixed-viewing block participants were instructed to fixate the central cross throughout the trial. In order to ensure that participants were fixating at the central fixation cross before each trial, the experiment was programmed such that recalibration was performed if they were not doing so after 3 s upon the appearance of the central fixation cross. Participants were told that they were free to move their eyes during the blank inter-stimulus interval to prevent fatigue. Before each block, a pseudo-random 9-point calibration was performed. Participants were given a self-paced rest period every 28 trials (i.e., after each mini-block), after which drift correction was performed before resuming the experiment. Participants were also given a short rest between each experimental block.

Each trial began with a central fixation cross which was presented for 1 s, followed by the search array. Participants

were instructed to report the identity of the target by pressing ‘z’ when the target was pointing to the left, or ‘/’ when the target was pointing to the right. The search array remained on the screen until a response was made. Feedback was given only for an incorrect response, in the form of a loud beep.

## Appendix 3. Limitations of TCS

One practical problem is that while ideally predictions can be made with mathematics, some mathematical relations between the hypothesized cognitive process and the outcome variables can be analytically intractable. For a wide range of cognitive processing models, there are no explicit equations relating model parameters that characterize the actual mechanism to observable parameters that describe the behavioral results. For example, it has been shown that, while the ex-Gaussian distribution appears to be a good-fitting model for human RT data (Dawson, 1988; Palmer, Horowitz, Torralba, & Wolfe, 2011), it is generally incorrect to directly interpret the parameters in the fitted distribution as each specifying a certain aspect of the underlying processing mechanism (Matzke & Wagenmakers, 2009; although such attempts have still been made, e.g., Moutsopoulou & Waszak, 2012; Steinhauer & Hübner, 2009). TCS also suffers from this problem, mainly because of the hypothesized parallel processing architecture. Under this architecture, the model’s prediction for the overall time cost of stage-one processing is computed as the maximum of  $N$  independent random variables (before a time-out is reached). So, even in cases when the time-out is not a factor (i.e., all lures complete before the time-out is ever reached), the analytical solution for the distribution of the maximum exists only in some specific cases, for example, when the processing time for all items follow the same exponential distribution (which implies all items are equally similar to the target under our current model). With our current assumption that processing time is sampled from an Inverse Gaussian distribution, this problem appears unsolvable. If we consider the case of heterogeneous search (i.e. distractors with different visual features/similarity to the target), then an analytical solution is unavailable even if we were to use the simpler exponential distribution.

On the other hand, the computational model is easy to simulate, which is an advantage of TCS. One can easily simulate a number of accumulators (even with varying accumulation rates) to mimic factors like display heterogeneity or eccentricity changes. For example, the equations that predict heterogeneous search RTs in Wang et al. (2017) and Lleras et al. (2019) were first tested in simulations and then tested on human data. While simulated numerical results are not an exact analytical solution, they can still reliably represent properties of the model and predictions that the model would make. More importantly, it is necessary and convenient to use simulations to understand our current model since it was constructed with a probabilistic nature. Simulation methods



can also be more flexibly adapted to different situations, for example, to simulate results from a group of individuals. They can also be easily adapted to incorporate factors such as inter-trial contingencies (e.g., when colors repeat vs. change across trials) and individual differences (e.g., in terms of default time-out parameters). We know these are important factors to consider (Awh, Belopolsky, & Theeuwes, 2012), though we have yet to study them. At this time, we simply assume that these sorts of effects will average out across trials. Thus, in our analysis we have taken the average group data as the most representative pattern of human behavior and compare that pattern (rather than individual subject data) to the mean patterns observed in the simulations.

A second limitation is that, as of now, TCS is not yet an image-based computational model. We do not see this as an insurmountable issue. Every aspect of the model should be able to be implemented such that the model ought to be able to: identify the number of accumulators to evaluate on any given trial, and compute contrast signals based on a known set of target-defining features. The evidence accumulation rates are proportional to that overall signal, so they need not be determined a priori by the user. When similarity is high, some samples will contain very little or no contrast information, whereas when similarity is low, every sample will likely provide positive evidence that this item is not like the target, allowing to reach threshold in fewer steps.

A third limitation is that the model has yet to instantiate a mechanism for distractor-distractor similarity effects. What we know so far is that lure-lure similarity effects are sensitive to the spatial distribution of the lures and that the magnitude of this interaction might depend on the complexity of the stimuli (Lleras et al., 2019), but much work remains to be done on this front. One might imagine that lure-lure facilitation might occur when identical lures are nearby one another and there might be a limit to the how close together the lures need to be for the advantage to arise. Or, rather than a spatial limit, what might be more important is that there are no other intervening lures in between two identical lures. What appears to be the case is that candidate-candidate similarity effects are of a different nature, that deserve their own study. An initial examination of this topic suggests that candidate homogeneity might reduce the likelihood of revisitations during the serial part of the search (i.e., when candidates are being scrutinized, see Ng et al., submitted). An additional factor that will likely play an important role in distractor-distractor similarity effects is stimulus complexity. Indeed, as suggested by Lleras et al. (2019), stimulus complexity might play a role in the strength of distractor-distractor similarity effects, with simpler stimuli (colored geometric shapes) yielding stronger homogeneity facilitation effects than more complex stimuli (images of real-world stimuli). This might result from the fact that simpler stimuli will have a chance to interact along more levels of the ventral pathway in the brain than more complex objects,

which require processing from more anterior regions of the brain like IT. In sum, distractor-distractor similarity effects deserve further study and more precise modeling.

A fourth limitation is that TCS is currently mostly focused on parallel processing and efficient search. More work is needed to flesh out what happens after parallel evidence accumulation is stopped and several target likely locations need to be inspected. Finally, as mentioned above, TCS has not yet integrated a role for visual contrast (or salience) even though data suggest that salience drives very fast eye movements even in goal-directed efficient search.

A final limitation of TCS is that it does not include much detail regarding the selection of individual target-likely locations in stage two. We assume that as demonstrated by TAM (Zelinsky, 2008), the visual system is likely to prefer making an eye movement to a somewhat nearby location rather than to a farther location. At this point, we rely on randomization of stimuli locations in the experiment to avoid any systematic impact of those types of preferences on the results. In that context, current results from our lab do suggest that all target-likely locations are equally likely to be selected. That is, there is no prioritization of some candidates over others, at least not based on their degree of similarity to the target (Ng et al., submitted).

## References

- Adeli, H., Vitu, F., & Zelinsky, G. J. (2017). A model of the superior colliculus predicts fixation locations during scene viewing and visual search. *Journal of Neuroscience*, *37*(6), 1453–1467.
- Adelson, E. H. & Bergen, J. R. (1991). The plenoptic function and the elements of early vision. In Computational models of visual processing, M. Landy & J. A. Movshon (eds), Cambridge, MA: MIT Press, pp. 3–20.
- Arita, J.T., Carlisle, N.B. & Woodman, G.F. (2012) Templates for rejection: configuring attention to ignore task-irrelevant features. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 580–584.
- Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443.
- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12):13, 1–18, <http://www.journalofvision.org/content/9/12/13>, doi:<https://doi.org/10.1167/9.12.13>.
- Ball, K. K., Beard, B. L., Roenker, D. L., Miller, R. L., & Griggs, D. S. (1988). Age and visual search: Expanding the useful field of view. *Journal of the Optical Society of America, A, Optics, Image & Science*, *5*(12), 2210–2219.
- Becker, S. I. (2008). Can intertrial effects of features and dimensions be explained by a single theory? *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1417–1440.
- Becker, S. I. (2010). The role of target–distractor relationships in guiding attention and the eyes in visual search. *Journal of Experimental Psychology: General*, *139*(2), 247–265.
- Becker, S. I., Folk, C. L., & Remington, R. W. (2013). Attentional capture does not depend on feature similarity, but on target-nontarget relations. *Psychological Science*, *24*(5), 634–647.

- Becker, S. I. (2013). Simply shapely: relative, not absolute shapes are primed in pop-out search. *Attention, Perception & Psychophysics*, 75(5), 845–61. doi:<https://doi.org/10.3758/s13414-013-0433-1>
- Becker, S. I. (2014). Guidance of attention by feature relationships: the end of the road for feature map theories?. In Mike Horsley, Matt Eliot, Bruce Allen Knight and Ronan Reilly (Ed.), *Current trends in eye tracking research* (pp. 37–49) Heidelberg, Germany: Springer. doi:[https://doi.org/10.1007/978-3-319-02868-2\\_3](https://doi.org/10.1007/978-3-319-02868-2_3)
- Becker, S. I., Harris, A. M., Venini, D., & Retell, J. D. (2014). Visual Search for Color and Shape: When Is the Gaze Guided by Feature Relationships, When by Feature Values ?, 40(1), 264–291. doi:<https://doi.org/10.1037/a0033489>
- Becker, S. I., Harris, A. M., York, A., & Choi, J. (2017). Conjunction search is relational: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 43(10), 1828–1842. doi:<https://doi.org/10.1037/xhp0000371>
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226, 177–178.
- Briggs, G. E., & Swanson, J. M. (1970). Encoding, decoding, and central functions in human information processing. *Journal of Experimental Psychology*, 86(2), 296–308.
- Buetti, S., Cronin, D. A., Madison, A. M., Wang, Z., & Lleras, A. (2016). Towards a better understanding of parallel visual processing in human vision: Evidence for exhaustive analysis of visual information. *Journal of Experimental Psychology: General*, 145(6), 672.
- Bundesden, C. (1990). A theory of visual attention. *Psychological Review*, 97(4), 523.
- Burkhardt, D. A., & Fahey, P. K. (1999). Contrast rectification and distributed encoding by ON-OFF amacrine cells in the retina. *Journal of Neurophysiology*, 82(4), 1676–1688.
- Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Attention, Perception, & Psychophysics*, 57(8), 1241–1261.
- Carrasco, M., & Frieder, K. S. (1997). Cortical magnification neutralized the eccentricity effect in visual search. *Vision Research*, 37(1), 63–82.
- Cavanagh, P., Hunt, A., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, 14, 147–153.
- Clarke, A., Dziemianko, M., & Keller, F. (2014). Measuring the salience of an object in a scene. Talk presented at the meeting of the Vision Sciences Society, St. Pete Beach, FL.
- Clarke, ADF., Mahon, A., Irvine, A. and Hunt, AR., (2017). People Are Unable to Recognize or Report on Their Own Eye Movements. *Quarterly Journal of Experimental Psychology*. 70, 2251–2270
- Corballis, M. C., Kirby, J., & Miller, A. (1972). Access to elements of a memorized list. *Journal of Experimental Psychology*, 94(2), 185–190.
- Dawson, M. R. (1988). Fitting the ex-Gaussian equation to RT distributions. *Behavior Research Methods, Instruments, & Computers*, 20(1), 54–57.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433.
- De Valois, R.L., Cottaris, N.P., Elfar, S.D., Mahon, L.E. & Wilson, J.A. Some transformations of color information from lateral geniculate nucleus to striate cortex. *Proceedings of the National Academy of Science, USA* 97: 4997-5002 (2000).
- Desimone, R. and Duncan, J. (1995) Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, 18, 193–222. <https://doi.org/10.1146/annurev.ne.18.030195.001205>
- Donk, M., Soesman, L. (2010). Saliency is only briefly represented: evidence from probe-detection performance. *Journal of Experimental Psychology Human Perception and Performance*, 36, 286–302.
- Donk, M., van Zoest, W. (2008). Effects of saliency are short lived. *Psychological Science*, 19, 733–739.
- Duncan, J., & Humphreys, G. (1992). Beyond the search surface: visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance*, 18(2), 578.
- Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & psychophysics*, 62(3), 425–451.
- Edwards, W., Lindman, H., & Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193–242
- Ekroll, V., & Faul, F. (2012). Basic characteristics of simultaneous color contrast revisited. *Psychological Science*, 23(10), 1246–1255.
- Fifić, M., Townsend, J. T., & Eidels, A. (2008). Studying visual search using systems factorial methodology with target-distractor similarity as the factor. *Perception & Psychophysics*, 70(4), 583–603.
- Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, 14(9), 1195–1201.
- Hubel, D. H., & Wiesel, T. N. (1967). Cortical and callosal connections concerned with the vertical meridian of visual fields in the cat. *Journal of neurophysiology*, 30(6), 1561–1573.
- Hulleman, J., & Olivers, C. (2017). The impending demise of the item in visual search. *Behavioral and Brain Sciences*, 40, E132.
- Hunt, A. R., & Cavanagh, P. (2009). Looking ahead: The perceived direction of gaze shifts before the eyes move. *Journal of Vision*, 9(2009), 1–7. doi:<https://doi.org/10.1167/9.9.1>
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10–12), 1489–1506.
- Klein, R. M. (1988). Inhibitory tagging system facilitates visual search. *Nature*, 334, 430–431.
- Kristofferson, M. W. (1972). When item recognition and visual search functions are similar. *Perception & Psychophysics*, 12(4), 379–384.
- Laubrock, J., Cajar, A., Engbert, R., (2013). Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. *Journal of Vision*, 13(12):11. 1–20.
- Lee, D. K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature Neuroscience*, 2, 375–381.
- Levi, D. M. (2008). Crowding-An essential bottleneck for object recognition: A mini-review. *Vision Research*, 48(5), 635–654.
- Lleras, A. & von Mühlenen, A. (2004). Spatial context and top-down strategies in visual search. *Spatial Vision*, 17(4–5), 465–482.
- Lleras, A., Wang, Z., Madison, A., & Buetti, S. (2019). Predicting Search Performance in Heterogeneous Scenes: Quantifying the Impact of Homogeneity Effects in Efficient Search. *Collabra: Psychology*, 5(1).
- Macé, M. J., Thorpe, S. J., & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: How robust at very low contrasts? *European Journal of Neuroscience*, 21(7), 2007–2018.
- Madison, A., Lleras, A., & Buetti, S. (2018). The role of crowding in parallel search: Peripheral pooling is not responsible for logarithmic efficiency in parallel search. *Attention, Perception, & Psychophysics*, 1–22.
- MacInnes, W. J., & Klein, R. M. (2003). Inhibition of return biases orienting during the search of complex scenes. *Scientific World Journal*, 3, 75–86.
- Mahon, A., Clarke, ADF. and Hunt, AR., (2018). The role of attention in eye-movement awareness. *Attention, Perception, and Psychophysics*. 80 (7), 1691–1704
- Mareschal, I., & Baker, C. L., Jr. (1998). A cortical locus for the processing of contrast-defined contours. *Nature Neuroscience*, 1(2), 150–154.
- Matzke, D., & Wagenmakers, E. J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, 16(5), 798–817.
- Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in neurosciences*, 29(6), 317–322.

- Mitroff, S. R., & Biggs, A. T. (2014). The ultra-rare-item effect: Visual search for exceedingly rare items is highly susceptible to error. *Psychological Science*, *25*(1), 284–289.
- Moher, J., Lakshmanan, B., Egeth, H., & Ewen, J. (2014). Inhibition drives early feature-based attention. *Psychological Science*, *25*, 315–324.
- Moore, C. M., & Egeth, H. (1998). How does feature-based attention affect visual processing? *Journal of Experimental Psychology: Human Perception and Performance*, *24*(4), 1296–1310.
- Moutsopoulou, K., & Waszak, F. (2012). Across-task priming revisited: Response and task conflicts disentangled using ex-Gaussian distribution analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 367.
- Müller, H. J., & von Mühlelen, A. (2000). Probing distractor inhibition in visual search: Inhibition of return. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1591–1605.
- Navalpakkam, V. & Itti, L. (2007). Search goal tunes visual features optimally. *Neuron*, *53*, 605–617.
- Neider, M. B., & Zelinsky, G. J. (2010). Exploring the perceptual causes of search set-size effects in complex scenes. *Perception*, *39*, 780–794.
- Ng, G. J. P., Buetti, S., & Lleras, A. (2016). The role of eye movements in parallel search: eye movements are neither necessary nor sufficient for logarithmic search functions to emerge. Retrieved from [osf.io/pve8d](https://osf.io/pve8d)
- Ng, G. J. P., Lleras, A., & Buetti, S. (2018). Fixed-target efficient search has logarithmic efficiency with and without eye movements. *Attention, Perception, & Psychophysics*, *80*(7), 1752–1762.
- Ng, G.J.P., Patel, T. N., Buetti, S., & Lleras, A. (submitted). Prioritization in visual attention does not work the way you think it does.
- Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research*, *31*, 1073–1078.
- Nothdurft, H. C. (1992). Feature analysis and the role of similarity in preattentive vision. *Perception & Psychophysics*, *52*, 355–375.
- Nothdurft, H. C. (1993). The role of features in preattentive vision: comparison of orientation, motion and color cues. *Vision Research*, *33*, 1937–1958.
- Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: a computational model of fixation durations in scene viewing. *Psychological review*, *117*(2), 382.
- Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 108.
- Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search?. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 58.
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision research*, *40*(10–12), 1227–1268.
- Peters, R.J., Iyer, A., Itti, L., Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*, 2397–2416.
- Rangelov, D., Müller, H. J., & Zehetleitner, M. (2017). Failure to pop out: Feature singletons do not capture attention under low signal-to-noise ratio conditions. *Journal of Experimental Psychology: General*, *146*(5), 651–671.
- Rosenholtz, R. (2001). Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 985.
- Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Review of Vision Science*, *2*, 227–247.
- Rosenholtz, R. (2017). Those pernicious items. *Behavioral & Brain Sciences*, *40*, e154.
- Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, *12*(4):14, 1–17, <http://www.journalofvision.org/content/12/4/14>, doi:<https://doi.org/10.1167/12.4.14>.
- Sanders, A. F. (1970). Some aspects of the selective process in the functional visual field. *Ergonomics*, *13*(1), 101–117.
- Scolari, M., & Serences, J. T. (2010). Basing perceptual decisions on the most informative sensory neurons. *Journal of Neurophysiology*, *104*, 2266–2273.
- Simpson, P. J. (1972). High-speed memory scanning: Stability and generality. *Journal of Experimental Psychology*, *96*(2), 239–246.
- Smilek, D., Enns, J. T., Eastwood, J. D., & Merikle, P. M. (2006). Relax! cognitive strategy influences visual search. *Visual Cognition*, *14*(4–8), 543–564.
- Solomon, J. A., Sperling, G., & Chubb, C. (1993). The lateral inhibition of perceived contrast is indifferent to on-center/off-center segregation, but specific to orientation. *Vision Research*, *33*(18), 2671–2683.
- Steinhauser, M., & Hübner, R. (2009). Distinguishing response conflict and task conflict in the Stroop task: evidence from ex-Gaussian distribution analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1398.
- Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of vision*, *11*(5), 13–13.
- Suchow, J. W., Brady, T. F., Fougner, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10), 8.
- Swanson, J. M., & Briggs, G. E. (1969). Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology*, *81*(2), 223–229.
- Takeda, Y., & Yagi, A. (2000). Inhibitory tagging to continuous visual stimuli. *Perception & Psychophysics*, *62*, 927–934.
- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of vision*, *11*(5), 5–5.
- Thomas, L. E., Ambinder, M. S., Hsieh, B., Levinthal, B., Crowell, J. A., Irwin, D. E., Kramer, A. F., Lleras, A., Simons, D. J., & Wang, R. F. (2006). Fruitful visual search: Inhibition of return in a virtual foraging task. *Psychonomic Bulletin & Review*, *13*, 891–895.
- Thomas, L. E., & Lleras, A. (2009). Inhibitory tagging in an interrupted visual search. *Attention, Perception & Psychophysics*, *71*, 1241–1250.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. CUP Archive.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97–136.
- Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of experimental psychology: human perception and performance*, *16*(3), 459.
- Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, *114*(3), 285.
- Trukenbrod, H. A., & Engbert, R. (2014). ICAT: A computational model for the adaptive control of fixation durations. *Psychonomic bulletin & review*, *21*(4), 907–934.
- van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human perception and performance*, *30*, 746–759.
- van Zoest, W., Lleras, A., Kingstone, A., & Enns, J. T. (2007). In sight, out of mind: the role of eye movements in the rapid resumption of visual search. *Perception & Psychophysics*, *69*, 1204–1217.
- Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, *31*(4), 523–535.
- Wang, Z., Buetti, S., & Lleras, A. (2017). Predicting search performance in heterogeneous visual search scenes with real-world objects. *Collabra: Psychology*, *3*(1).
- Wang, Z., Lleras, A., & Buetti, S. (2018). Parallel, exhaustive processing underlies logarithmic search functions: Visual search with cortical magnification. *Psychonomic Bulletin & Review*, *25*(4), 1343–1350.
- Williams, L. J. (1989). Foveal load affects the functional field of view. *Human Performance*, *2*(1), 1–28.

- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202–238.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 419–433.
- Wolfe, J.M., Friedman-Hill, S.R., Stewart, M I., & O'Connell, K. M. (1992) The Role of Categorization in Visual Search for Orientation. *Journal of Experimental Psychology*, 18(1): 34–49
- Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Cognitive psychology: rare items often missed in visual searches. *Nature*, 435(7041), 439.
- Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, 20(2), 121–124.
- Woodman, G.F. & Luck, S.J. (2007) Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology: Human Perception and Performance*, 33, 363–377.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115(4), 787.
- Zhang, X., Huang, J., Yigit-Elliott, S., & Rosenholtz, R. (2015). Cube search, revisited. *Journal of Vision*.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.