Noname manuscript No.
(will be inserted by the editor)

The Alejandro Project: Testing the Target Contrast Signal Theory

Replication and Generalisation

Anna Hughes · Anna Nowakowska · Alasdair D. F. Clarke

Received: date / Accepted: date

Abstract Insert your abstract here. Include keywords, PACS and mathematical subject classification numbers as needed.

Keywords First keyword · Second keyword · More

1 Introduction

Background Visual search, where participants are asked to find a target within a cluttered scene, has been extensively studied within psychology. Several models have been developed that can generate testable predictions about how different types of distractors and targets affect search efficiency.

The two most studied families of models of visual search include bottom-up models (also known as visual saliency models) and top-down models. Saliency models rest on the assumption that fixations are directed to objects or locations that are most dissimilar to the background or other objects in the visual display Itti and Koch (2000); Itti et al. (1998); Koch and Ullman (1987). While the original saliency model was able to predict fixation allocation in a visual search task above chance Parkhurst et al. (2002), further research demonstrated that a comparable level of performance could be achieved using a simple central fixation bias heuristic Tatler (2007). The saliency models have since been extended and improved (see for example Zhang et al. (2008)): however, the main issue with this family of models remains their limited usability in complex real-life search arrays Tatler et al. (2011); Koehler et al.

ESRC grant?

F. Author
first address
Tel.: +123-45-678910
Fax: +123-45-678910
E-mail: fauthor@example.com

S. Author
second address

(2014). In addition, in most instances of visual search, the target is clearly defined (i.e. the goal is to find a specific object) and inspecting the most salient areas of the display may in these cases be inefficient.

Another class of models are based around Feature Integration Theory Treisman and Gelade (1980), which has been modified and extended by Wolfe in the Guided Search Model Wolfe et al. (1989); Wolfe (2014). These models combine top-down influences (how closely an item resembles the observer's goal) with bottom-up image properties. For example, if one's goal (top-down processing) is to find a red horizontal bar, all the red and horizontal items in a visual search display will be given greater weight than distractors (e.g. vertical and blue items) in the model. The salience of a given object in the display (how distinctive it is from the surrounding objects) also activates bottom-up processing. For instance, a blue item among red items is ranked higher than red among orange items. Combining bottom-up and top-down sources of activation generates an activation map which generates a prediction of the order in which stimuli are processed in visual search. Thus, these models aim to produce a representation of the visual properties of the distractors at each location in the visual field.

However, more recent work has taken a different approach, focusing solely on representing the difference between targets and distractors. For example, in work on eye movement patterns, it has been proposed that performance in inefficient (serial) visual search is mostly determined by the size of the 'functional viewing field', whose size varies as a function of target-distractor similarity Hulleman and Olivers (2017). One recent model, the Target Contrast Signal (TCS) Theory Lleras et al. (2020) aims to provide a unifying, quantitative framework that can make behavioural predictions based on this general assumption.

Theoretical intro to TCS TCS proposes that behaviour is determined by comparing the target template (held in memory) with every element present in the scene in parallel. This allows the visual system to reject peripheral non-targets quickly; the speed at which items are evaluated is determined by how different the item is from the template through an evidence accumulation process (formally, the slope of the logarithmic function is assumed to be inversely proportional to the overall magnitude of the contrast signal between the target and distractor). The model thus focuses on an initial, efficient processing stage of search; if sufficient evidence is not accumulated during this process, the model posits that a second stage is entered, where attention is deployed serially. TCS has been successful in predicting a number of empirical results, including search performance in heterogeneous scenes based on parameters estimated in homogeneous scenes, both with artificial stimuli Buetti et al. (2016); Lleras et al. (2019) and with real-world objects visualised on a computer display Wang et al. (2017).

Limitations, and extending basic stimuli While results to date for TCS appear promising, they remain relatively preliminary, being tested by only one research group so far. There remains a great deal of scope for extending beyond the parameters tested to date (such as the number of distractors) in order to test the robustness and generalisability of the model. In addition, there are some predictions of the model, including its ability to explain search asymmetries, that have yet to be empirically confirmed Lleras et al. (2020).

Within subjects In addition, in all implementations of TCS so far, predictions of search efficiency (e.g. in heterogeneous scenes) have been made on the average of a group of participants, using data from a different group performing a different task (e.g. searching in homogeneous scenes). Thus, we know that TCS can replicate group-level averages between subjects in search well, but we do not know whether it is also able to make predictions at the individual level. This is particularly important given that conclusions based on aggregate data can be different from those that take individual differences into account; in one study where participants searched for a target in an array of randomly oriented line segments, aggregating the data suggested that participants were using a stochastic search model. However, when considering each participant individually, it became clear that there was a high level of heterogeneity in responses, with some participants performing close to optimally, and others actually performing worse than chance Nowakowska et al. (2017). Similarly striking variability has also been reported in other search studies Irons and Leber (2016, 2018).

Conclusion to introduction In the current manuscript, we focus on replicating and extending findings from Buetti et al. (2019). Here, participants searched for a target in a scene of homogeneous distractors. First, parallel search efficiency (measured by the logarithmic search slope) was estimated for cases where the distractors varied from the target in one dimension: either colour (e.g. a cyan target being searched for in either yellow, blue or orange distractors) or shape (e.g. a semicircle target in either circle, diamond or triangle distractors). New participants then searched for the same targets in displays where the distractors were compounds, differing from the target in both colour and shape (e.g. searching for a cyan semicircle in either blue circles, orange diamonds or yellow triangles). Figure 1 shows example stimuli from their paper. The logarithmic search slopes in the initial experiments were then used to predict the logarithmic slopes and reaction times using a number of models. The authors found that the best model was a 'collinear contrast integration model' where the distinctiveness scores were summed along each attribute in the unidimensional experiments, creating an overall contrast score that was used for compound stimuli predictions.

Overview of the current study We first run a replication of Buetti et al. (2019), in an online, within-subjects study. This design allows us to extend the modelling, both incorporating a multi-level design to predict within-subjects effects and by utilising a Bayesian generalised linear model framework to better represent the distribution of responses (e.g. avoiding predicting negative reaction times, accounting for uncertainty in model predictions). We also carry out a direct analytical replication using the same methods as in Buetti et al. (2019) allowing us to ask whether the choice of analysis affects the results.

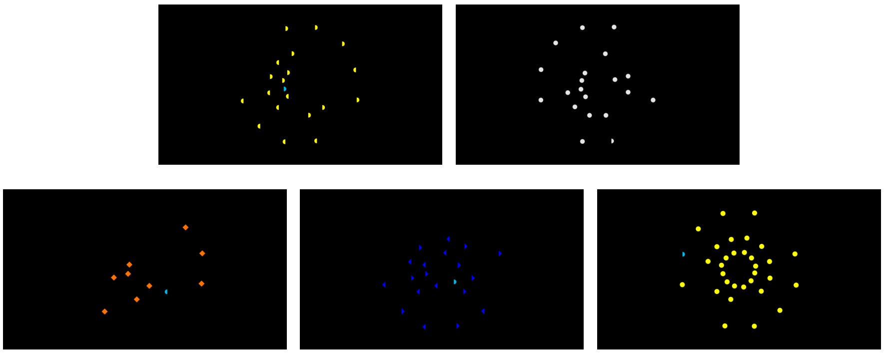

Fig. 1 Example stimuli from Buetti et al. (2019) Top left: Expt 1A. Here, the target is a blue semicircle within a set of homogeneous (yellow semicircle) distractors. Top right: Expt 1B. The target is a grey semicircle in circular grey distractors. Bottom left: Expt 2A. The target is a blue semicircle in orange diamond distractors. Bottom middle: Expt 2B. The target is a blue semicircle in dark blue triangle distractors. Bottom right: Expt 2C. The target is a blue semicircle in yellow circular distractors.

2 The Target Contrast Model

We first describe the original Target Contrast Model (and verify that we can replicate the original analysis - see Supplementary Materials). This is followed by our proposed extension.

2.1 TCS modelling overview

In Experiment 1a of Buetti et al. (2019), participants searched for a cyan semicircle target among blue, yellow or orange semicircular distractors i.e. they searched for a target that differed from the distractors by a *single feature* (colour). The experiment was then repeated (1b) using a different single feature (shape, with participants searching for the semicircular target within triangle, circle or diamond distractors). In Experiments 2a, 2b and 2c, participants again searched for a cyan semicircle, but this time, the distractors differed in both shape and colour. We will refer to these conditions as *double features*. Note, unlike in standard conjunction searches, in this paradigm, the distractors are all identical with respect to these features (i.e, orange triangles).

The *Target Signal Contrast* theory allows us to predict the value of the logarithmic slope, $D_{\mathrm{c,s}}$, in this condition based on the corresponding D_j in the single feature seach experiments.

2.1.1 Calculating the intercept, a, and the logarithmic slope parameter, D_i

Experiments 1a and 1b (referred to jointly as Experiment 1) were used to calculate the logarithmic slope parameter D_i. In both experiments, the number of distractors varied, allowing the data to be used to fit a log-linear model for reaction times, where reaction times increase linearly with $\log(N_T)$ (the log of the number of distractors) 1. In the original model the error distribution was assumed to be normal.

$$\hat{RT} = D_i \log(N_T + 1) + a \tag{1}$$

Thus the results of Experiment 1 were used to calculate the logarithmic slope, for each type of distractor. When colour varied, we will refer to D_c, for $c = 1, 2, 3$. Similarly for shape we will denote this (D_s), and the compound features are denoted as ($D_{c,s}$). As a represented the intercept in our model, it is independent of both shape and colour, as it represents how long observers take to find a target when $N_T = 0$, i.e., there are no distractors. However, this leads to some ambiguity in how a should be defined; in Buetti et al. (2019), a was calculated for each feature in each sub-experiment (e.g. experiments 1a, 1b, 2a, 2b, 2c). Here, we follow that method in order to replicate their results exactly. However, it is also possible to calculate a at the level of an experiment (e.g. experiments 1 and 2), and this is the approach we take for our re-analysis (see below).

2.1.2 Estimating $D_{c,s}$, the logarithmic slope parameter for compound features

In the context of the current experiments, the core idea of TCS theory is that we can estimate the logarithmic slope parameter for a double feature visual search from the slopes parameters in the two independent single feature searches. I.e., $D_{c,s} = f(D_c, D_s)$.

Buetti et al. (2019) tested three different models for predicting D for compound colour-shape stimuli. The best feature guidance model (Equation 2) suggests that when the target and lures differ in two dimensions, participants will choose to attend to whichever feature dimension is the most discriminable (i.e. has the smallest D value):

$$D_{\mathrm{c,s}} = \min(D_{\mathrm{c}}, D_{\mathrm{s}}) \tag{2}$$

The orthogonal contrast combination model instead suggests that independent feature dimensions comprise a multidimensional space, where an object can be described by the overall vector in this space, and thus $\mathrm{D_{overall}}$ can be represented as:

$$D_{\mathrm{c,s}} = \frac{1}{\sqrt{\frac{1}{(D_{\mathrm{c}})^2 + (D_{\mathrm{s}})^2}}} \tag{3}$$

Note: we had to change this a little to make it match! !!!! *******

Finally, the collinear contrast integration model also assumes independence of feature dimensions, but assumes that while the visual features create a multidimensional space, the contrast between them is unidimensional. As D is assumed to be inversely proportional to contrast, the equation can be written as follows:

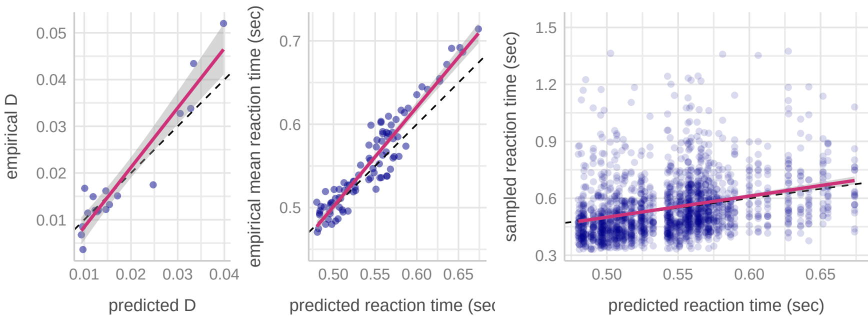

Fig. 2 (left) The collinear method for calculating D offers a good prediction. (centre) Using the TSC to predict reaction times. (right) Each dot now represents a randomly sampled reaction time from an observer.

method	Experiment 2	Experiment 4	Both Experiments
best feature	0.960	0.946	0.944
orthogonal contrast	0.961	0.958	0.945
collinear	0.956	0.966	0.936

Table 1 A table of R^2 values

$$\frac{1}{D_{\mathrm{c,s}}} = \frac{1}{D_{\mathrm{c}}} + \frac{1}{D_{\mathrm{s}}} \tag{4}$$

Buetti et al. (2019) found that with their dataset, the collinear contrast integration model was best able to predict $D_{c,s}$ from D_c and D_s, with $R^2 = 0.915$. We verified we were able to replicate the results of Buetti et al. (2019) using the dataset available on OSF (https://osf.io/f3m24/)[1] and using the exclusion criteria originally applied[2]; see *Supplementary Materials* for details. See Figure 2(left panel).

However, the evidence that the collinear contrast model outperforms the orthogonal contrast model is somewhat weak. The above R^2 statistic was calculated over Experiments 2 and 4 together. If we analyse each experiment seperately, we find that the orthogonal contrast model offers the best fit for the Experiment 2 data (see Table). Furthermore, there is very little difference between the top two methods. This will be discussed further below (Section...)

2.1.3 Estimating mean reaction times

The TCS theory now allows us to predict mean reaction times:

$$RT = Dc,s \log(N_T + 1) + a \tag{5}$$

As can be seen in Figure 2 (centre panel), these predictions are very close to the observed values.

their empirical RT results with an $R^2 = 93.27\%$.

[1] downloaded on 28th August 2020

[2] We also applied some additional exlusion crit!

2.1.4 Discussion

While TCS theory offers a good prediction of search slopes and corresponding mean reaction times for double feature search, there are two related limitations. Firstly, it is unable to account for individual differences between observers, only the changes to the sample average. Secondly, it cannot account for the distribution of reaction times over multiple trials. Figure 2 (right panel) shows clearly that these factors generate high levels of variability within the individual trial-level data. To address these issues, we propose a second version of the TCS that makes use of multi-level modelling techniques.

2.2 A multi-level TCS

Switching from a linear regression model to a multi-level model will allow us to compute D for each participant, while simultaneously estimating the trial-to-trial variance. We also switch from a frequentist to Bayesian framework, as this allows us to naturally account for the uncertainty in the model's predictions.

However, switching from linear regression to a mutli-level model raises the problem of which distribution to use for modelling reaction times. Using a normal distribution is unlikely to be satisfactory, as it is unable to account for the skew frequently seen in reaction time distributions, and also allows the possibility of negative reaction times. We can account for both of these problems by using a log-normal distribution, rt lognormal(μ, σ)). We will also test whether a slightly more complex extension of this model, the *shifted lognormal* model (which allows the distribution to be offset to the right i.e. mimicking the patterns seen in reaction time data, where valid responses begin at around 100ms) offers any improvement in model fit.

2.2.1 Calculating the logarithmic slope parameter, D_i

We began in a similar fashion to section 2.1.1, although this time, we fitted each model three times using a (i) normal, (ii) lognormal, and (iii) shifted lognormal distribution (please see Supplementary Materials for full details of out modelling approach, including prior predictive checks and model fit diagnostics). The three models were compared using leave-one-out cross validation, and the results showed that the shifted-lognormal model (illustrated in Figure 3) was given 100% of the weight. Therefore, we used only this model for the rest of the analysis.

We then extracted the posterior distributions for the logarithmic slope parameters as before. The only differences were that i) this approach allowed us to obtain the full posterior distribution for each D_c and D_s, rather than just the maximum likelihood fit, and ii) the units are now in $\frac{\log rt}{\log N_T + 1}$ See Figure 4(bottom).

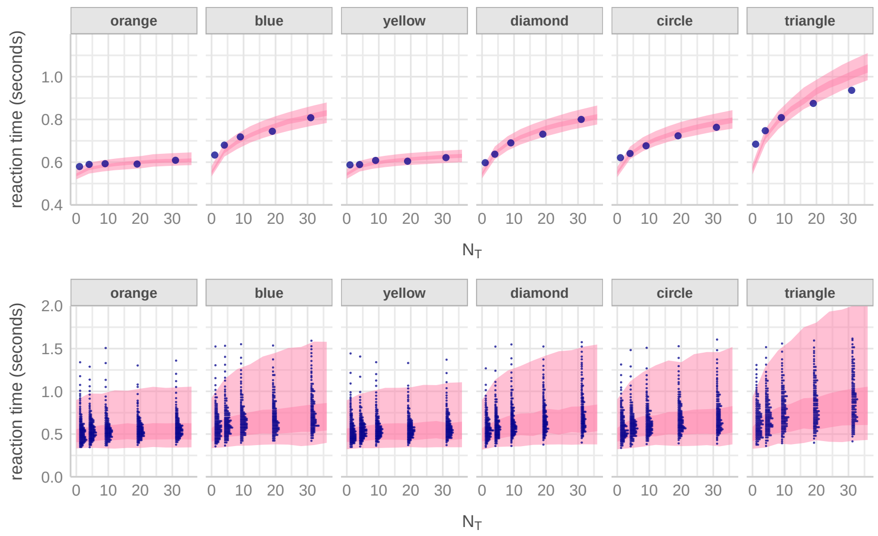

Fig. 3 (*top*) The shaded regions show the model's estimate (53% and 97% intervals) of the average participant's mean reaction time, while the points show the empircal mean reaction time. (*bottom*): The shaded regions now indicate the distribution of reaction times (over a new simulated group of participants) generated by the model. The points now represent the 100 quantiles from the empirical data.

2.2.2 Estimating $D_{c,s}$, the logarithmic slope parameter for compound features

We predict $D_{c,s}$ from D_c and D_c in the same way as the original TCS, except this time we use 1000[3] samples from the posterior distributions, rather than just the maximum likelihood fit. These predictions are shown in Figure 4(bottom). We can see that as with the original model, we still find the *best feature* is the worst of the three methods. We additionally calculated a fourth prediction method, the *mean method* which simply takes the averages of the *collinear contrast* and *orthogonal contrast* models.

2.2.3 Estimating other parameters and predicting reaction times

Unlike the origianl TCS, before we can use the predicted values of $D_{c,s}$ to generate reaction times, we first have to estimate some additional parameters:

- α - the intercept for the shift[4] parameter.
- ϕ_a - the random intercepts for the linear predictor for μ.
- ϕ_α - the random intercepts for a.
- σ - the residual (trial-to-trial) variance.

For the implementation presented here, in order to predict the data from Experiment 2 (4) we will simply use the values for the above parameters obtained from

[3] check!
[4] ndt

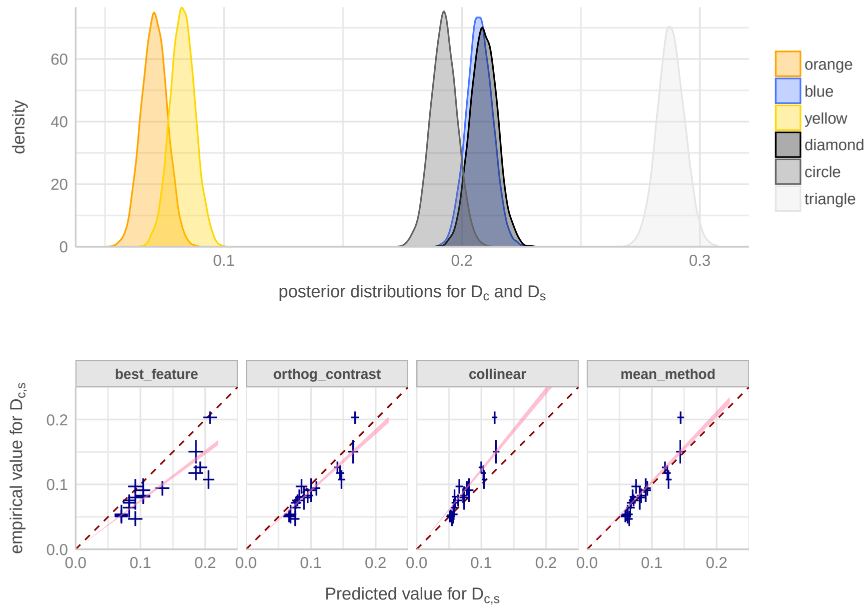

Fig. 4 (*top*) Posterior probability density functions for each D_c and D_s in Experiment 1. As with the original TCS model, when searching for a blue semicircle, triangular distracters are the hardest, while yellow and orange distracters are the eaiest.) (*bottom*) Predicting $D_{c,s}$ from D_c and D_s. Crosshairs indicate 97.5 HPDI for estimates and predictions, while the pink region shows the HDPI for the best fit line.

Experiment 1 (3). Once these have been set, we can use the specified model to generate the HDPI intervals for the expected average performance and the full distribution of reaction times over a number of simulated new observers (Figure 5).

We can summarise how well the model predicts the mean reaction times (Figure 5(left)) by repeatedly i) sampling from the posterior, ii) computing the predicted mean reaction times, and iii) computing the R^2 statistic for the correlation between predicted and empirical mean reaction times. Doing this gives us a HPDI of $[0.75, 0.79]$, which compares with the value of $R^2 = 0.88$ obtained for the original TCS. [DO WE WANT TO COMMENT ON THE DIFFERENCE?]. While our new model is slighlty lower than the original method, as discussed above, it offers substantial improvements over the original model in terms of predicting the full distribution of reaction times over multiple trials and participants.

2.2.4 Discussion

Our extension to a multi-level TCS model allows us to go beyond the original version and predict the full distribution of reaction times over samples of known, or new, observers. Moving from a normal to a shifted-lognormal distribution allows us to

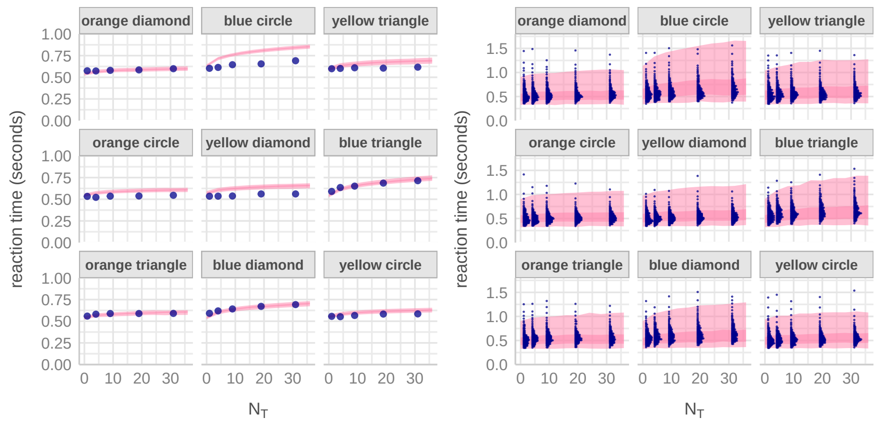

Fig. 5 (*left*) HDPI and empirical mean reaction times for each condition in Experiment 2. (*right*) Similar, for the full distribution of reaction times over trials.

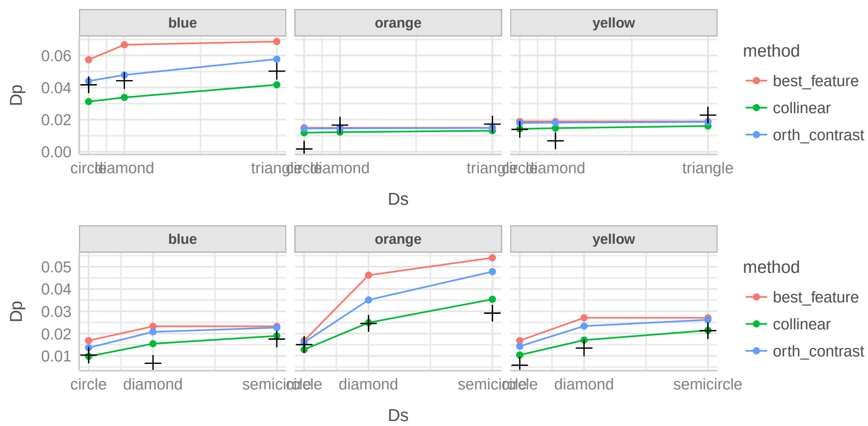

Fig. 6 (Some oof our conditions aren't all that sensitive!

accuratly model the skew seen in the empirical data, and avoid predicting negative response times.

2.3 Sensitivity to different *D* methods

It is hard to distinguish between the three methods! See Figure 6

2.4 Discussion

While we replicate the good predictions using the collinear method if we use a normal distribution, we do not find this that this is the best method if we use a log-normal distribution, and instead, we find the orthogonal contrast method is slightly better. This suggests that (seemingly small) analytical decisions may have important implications for the conclusions drawn.

3 Hypotheses

We plan an experiment to test the extent to which the original results in Buetti et al. (2019) replicate and generalise. As well as following the original, between-subjects, experiment design, in which data from one group of observers in one task is used to predict behaviour of a second group of participants in a different task, we will allow for within-subject comparisons. We will also extend the original experiment to include a larger number of distractors.

3.1 Registered Hypothesis

1. **Replication of Buetti et al (2019) with online data collection.** Specifically, that the *collinear contrast ingratiation model* outperforms the *best feature guidance*, and *orthogonal contrast combination models*. Furthermore, the $R^2 =$ (99% HPDI = [,]) between predicted and observer reaction times.

2. **Larger number of distractors (or targets further in periphery?)** Add an extra ring. Say something about moving from parallel to serial search (although don't use these terms?!!)

3. **Individual differences** To what extent do the individual differences in the single-feature task explain the differences in the double-feature task?

Add something about wanting to test features that allow us to distinguish between the models?

3.2 Planned Explorations!

As these models can be quite challenging to fit, we're making some excuses. But we do plan to use the methods and framework outlined above (and in suppmat) to more formally investigate the individual differences. We plan to do this by specifying a more complex random effect structure, and investigating ...

4 General Methods

4.1 Sample Size: Participants and Trials

Based on the power analysis **??**, we will recruit 20 participants, who will each complete 570 trials (15 distractor types x 6 levels of numbers of distractors, each repeated 6 times, alongside 30 'zero distractor' trials).

In the original paper, there were inclusion criteria: search accuracy over 90% and individual average response times smaller than two standard deviations from the group average response time. Even in a lab set up, this normally led to at least one or two participants being excluded.

4.2 Stimuli

4.3 Procedure

4.4 Data Pre-processing

incorrect trials? Poorly behaved participants? RTs that are far too short? Or far too long?

4.5 Analysis Plan

We will follow the analysis given in Section **??**.

5 Results

– blank –

6 General Discussion

Is discriminating between the different models one of our aims? Or is this a discussion point i.e. it's quite hard to do? And therefore maybe a follow up paper?

Acknowledgements Thank you to AL for help and encouragement!

Conflict of interest

The authors declare that they have no conflict of interest.

References

Simona Buetti, Deborah A Cronin, Anna M Madison, Zhiyuan Wang, and Alejandro Lleras. 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, 2016.

Simona Buetti, Jing Xu, and Alejandro Lleras. Predicting how color and shape combine in the human visual system to direct attention. *Scientific reports*, 9(1):1–11, 2019.

Johan Hulleman and Christian NL Olivers. On the brink: The demise of the item in visual search moves closer. *Behavioral and Brain Sciences*, 40, 2017.

Jessica L Irons and Andrew B Leber. Choosing attentional control settings in a dynamically changing environment. *Attention, Perception, & Psychophysics*, 78(7):2031–2048, 2016.

Jessica L Irons and Andrew B Leber. Characterizing individual variation in the strategic use of attentional control. *Journal of Experimental Psychology: Human Perception and Performance*, 44(10):1637, 2018.

Laurent Itti and Christof Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10-12):1489–1506, 2000.

Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.

Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*, pages 115–141. Springer, 1987.

Kathryn Koehler, Fei Guo, Sheng Zhang, and Miguel P Eckstein. What do saliency models predict? *Journal of vision*, 14(3):14–14, 2014.

Alejandro Lleras, Zhiyuan Wang, Anna Madison, and Simona Buetti. Predicting search performance in heterogeneous scenes: Quantifying the impact of homogeneity effects in efficient search. *Collabra: Psychology*, 5(1), 2019.

Alejandro Lleras, Zhiyuan Wang, Gavin Jun Peng Ng, Kirk Ballew, Jing Xu, and Simona Buetti. A target contrast signal theory of parallel processing in goal-directed search. *Attention, Perception, & Psychophysics*, pages 1–32, 2020.

Anna Nowakowska, Alasdair DF Clarke, and Amelia R Hunt. Human visual search behaviour is far from ideal. *Proceedings of the Royal Society B: Biological Sciences*, 284(1849):20162767, 2017.

Derrick Parkhurst, Klinton Law, and Ernst Niebur. Modeling the role of salience in the allocation of overt visual attention. *Vision research*, 42(1):107–123, 2002.

Benjamin W Tatler. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of vision*, 7(14):4–4, 2007.

Benjamin W Tatler, Mary M Hayhoe, Michael F Land, and Dana H Ballard. Eye guidance in natural vision: Reinterpreting salience. *Journal of vision*, 11(5):5–5, 2011.

Anne M Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12 (1):97–136, 1980.

Zhiyuan Wang, Simona Buetti, and Alejandro Lleras. Predicting search performance in heterogeneous visual search scenes with real-world objects. *Collabra: Psychology*, 3(1), 2017.

Jeremy M Wolfe. Approaches to visual search: Feature integration theory and guided search. *Oxford handbook of attention*, pages 11–55, 2014.

Jeremy M Wolfe, Kyle R Cave, and Susan L Franzel. Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and performance*, 15(3):419, 1989.

Lingyun Zhang, Matthew H Tong, Tim K Marks, Honghao Shan, and Garrison W Cottrell. Sun: A bayesian framework for saliency using natural statistics. *Journal of vision*, 8(7):32–32, 2008.