

Draftable Comparison Export

This document is an exported comparison with limited functionality, generated by Draftable Desktop. To access full functionality, use Draftable's powerful comparison viewer in any of our products.

Left document: multi_level_tcs.pdf

Right document: multi_level_tcs_revisions.pdf

What is this document?

This is a comparison of two documents. The two documents are interleaved such that the left document is displayed on even pages and the right document is displayed on odd pages.

Is there a specific way I should view this file?

This document is intended to be viewed in Two Page Continuous mode (or sometimes called 'Two Page Scrolling'). It should open in this mode by default when using Adobe Acrobat and most popular PDF readers.

If the document opens in a different view, you can often change this in the settings. In Adobe Acrobat, go to **View > Page Display > Two Page Scrolling**.

Why are there blank pages?

Blank pages are inserted to keep both documents as aligned as much as possible.

How do I read the changes?

Text deleted from the left document and, hence, not in right document is highlighted red. Text added to the right document and, hence, not in left document is highlighted green.

Tip for printing

When printing this document, we recommend printing double-sided and include this first page. This will result in the matching text being displayed on different pages and easily readable, much like a book.

For more information

Draftable offers powerful document comparison solutions for all use-cases. To view our products, please visit our website: draftable.com.

Bayesian multi-level modelling for predicting single and double feature visual search

Anna E. Hughes^a, Anna Nowakowska^b, Alasdair D. F. Clarke¹

^a*Department of Psychology, University of Essex, Colchester, CO4 3SQ, UK*

^b*School of Psychology, University of Aberdeen, Aberdeen, AB24 3FX, UK*

Abstract

Performance in visual search tasks is frequently summarised by “search slopes” - the additional cost in reaction time for each additional distractor. While search tasks with a shallow search slopes are termed efficient (pop-out, parallel, feature), there is no clear dichotomy between efficient and inefficient (serial, conjunction) search. Indeed, a range of search slopes are observed in empirical data. The Target Contrast Signal (TCS) Theory is a rare example of quantitative model that attempts to predict search slopes for efficient visual search. In particular, the search slope in a double-feature search (where the target differs in both colour and shape from the distractors) can be estimated from the slopes of the associated single-feature searches. This estimating is done by using a contrast combination model, and previous work shows that a collinear contrast integration model outperformed other options. In our work, we extend the TCS to a Bayesian multi-level framework. We demonstrate that moving from using a normal distribution to a shifted-lognormal allows for a better fit of previously published data, and in doing so, we show that previous datasets are not sufficient to distinguish between the various contrast combination hypotheses that had been considered. We will conduct experiments that attempt to replicate the key original findings, with some changes to help distinguish between theories.

Keywords: Visual search, Efficient search, Parallel processing

1. Introduction

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. One of the key distinctions in the field

Bayesian multi-level modelling for predicting single and double feature visual search

Anna E. Hughes^a, Anna Nowakowska^b, Alasdair D. F. Clarke¹

^a*Department of Psychology, University of Essex, Colchester, CO4 3SQ, UK*

^b*School of Psychology, University of Aberdeen, Aberdeen, AB24 3FX, UK*

Abstract

Performance in visual search tasks is frequently summarised by “search slopes” - the additional cost in reaction time for each additional distractor. While search tasks with a shallow search slopes are termed efficient (pop-out, parallel, feature), there is no clear dichotomy between efficient and inefficient (serial, conjunction) search. Indeed, a range of search slopes are observed in empirical data. The Target Contrast Signal (TCS) Theory is a rare example of quantitative model that attempts to predict search slopes for efficient visual search. One study using the TCS framework has shown that the search slope in a double-feature search (where the target differs in both colour and shape from the distractors) can be estimated from the slopes of the associated single-feature searches. This estimation is done using a contrast combination model, and a collinear contrast integration model was shown to outperform other options. In our work, we extend TCS to a Bayesian multi-level framework. We investigate modelling using normal and shifted-lognormal distributions, and show that the latter allows for a better fit to previously published data. We propose running a new fully within-subjects experiment to attempt to replicate the key original findings, with some changes to help distinguish between theories.

Keywords: Visual search, Efficient search, Parallel processing

1. Introduction

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. One of the key distinctions in the field

has been between efficient (also referred to as parallel or pop-out) and inefficient (serial) search. These are often studied in the context of the regression slope between the number of distractors and mean reaction time, which has been termed the *search slope*. When the search slope is shallow (usually positive, but occasionally negative e.g. ?), the search is called efficient or parallel, and the addition of more non-target distractors has little impact on an observers difficulty in finding a target. When the slope is steeper, each additional distractor has a noticeable impact on increasing difficulty, and the search is described as inefficient or serial. However, the distinction between these types of search is often less clear in real experimental data, with a range of different search slopes being seen for different types of targets and distractors (Cave and Wolfe, 1990).

In the current study, we are interested in what has traditionally been termed efficient or parallel search, and the factors that affect search slope in these conditions. Recent work has suggested that for efficient search, there is a logarithmic relationship between distractor set size and reaction time, and that this relationship can be modified by target-distractor similarity (Buetti et al., 2016), providing evidence that search behaviour in parallel search is more complex than has previously been assumed. This observation has formed the basis of the ‘Target Contrast Signal (TCS) Theory’ (Lleras et al., 2020), which aims to provide a means of predicting observer search slopes for new search arrays by quantifying target-distractor differences. For example, by measuring search slopes for conditions in which the distractors differ from the target along a *single feature* (e.g. colour *or* shape), it has been shown that you can predict search times for arrays in which the target differs from the distractors along two features (e.g., colour *and* shape) which we refer to here as *double feature* search. (Buetti et al., 2019). Here, we aim to replicate and extend this work both theoretically and empirically, to test the generalisability of the TCS model, and to suggest ways in which the TCS model could be improved to generate better predictions.

1.1. Previous Work

Many different forms of visual search models have been proposed. One well developed class of models are the saliency models, which aim to predict eye movements during scene viewing, including visual search. They 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

has been between efficient (also referred to as parallel or pop-out) and inefficient (serial) search. These are often studied in the context of the regression slope between the number of distractors and mean reaction time, which has been termed the *search slope*. When the search slope is shallow (usually positive, but occasionally negative e.g. (Rangelov et al., 2017)), the search is called efficient or parallel, and the addition of more non-target distractors has little impact on an observers difficulty in finding a target. When the slope is steeper, each additional distractor has a noticeable impact on increasing difficulty, and the search is described as inefficient or serial. However, the distinction between these types of search is often less clear in real experimental data, with a range of different search slopes being seen for different types of targets and distractors (Duncan and Humphreys, 1989; Cave and Wolfe, 1990; Wolfe, 1998; Liesefeld et al., 2016). Recent work has also attempted to model the variation in search slopes at the boundary between inefficient and efficient search (Liesefeld et al., 2016).

In the current study, we are interested in what has traditionally been termed efficient or parallel search, and the factors that affect search slope in these conditions. Recent work has suggested that for efficient search, there is a logarithmic relationship between distractor set size and reaction time, and that this relationship can be modified by target-distractor similarity (Buetti et al., 2016), providing evidence that search behaviour in parallel search is more complex than has previously been assumed. This observation has formed the basis of the ‘Target Contrast Signal (TCS) Theory’ (Lleras et al., 2020), which aims to provide a means of predicting observer search slopes for new search arrays by quantifying target-distractor differences. For example, by measuring search slopes for conditions in which the distractors differ from the target along a *single feature* (e.g. colour *or* shape), it has been shown that you can predict search times for arrays in which the target differs from the distractors along two features (e.g., colour *and* shape) which we refer to here as *double feature* search. (Buetti et al., 2019). Here, we aim to replicate and extend this work both theoretically and empirically, to test the generalisability of the TCS model, and to suggest ways in which the TCS model could be *modified* to generate better predictions.

1.1. Previous Work

Many different forms of visual search models have been proposed. One well developed class of models are the saliency models, which aim to predict eye movements during scene viewing, including visual search. They 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

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., 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. Finally, by focusing on eye movements, these models do not necessarily provide a theoretical framework for the cognitive processes underlying visual search.

Perhaps the most established class of models of visual search are based around Feature Integration Theory (Treisman and Gelade, 1980), which has been modified and extended by Wolfe and colleagues in the Guided Search Model (Wolfe et al., 1989; Wolfe, 2014). These theories have been developed using data from visual search tasks with discrete sets of abstract items. 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. In such cases, a salient item can capture attention even without resembling the target. 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, these are predominantly qualitative models, and thus it is difficult to use them to make specific quantitative predictions.

TCS falls under a class of models that take a different approach, in that they focus 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). Similarly, work on attention has proposed the notion of 'relative features', where attention is tuned to feature relationships i.e. the appearance of the target relative to distractors in the environment (Becker et al., 2014; Becker, 2010). However, TCS (Lleras et al., 2020) aims to provide a unifying framework that can make quantitative behavioural predictions for visual

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., 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. Finally, by focusing on eye movements, these models do not necessarily provide a theoretical framework for the cognitive processes underlying visual search.

Perhaps the most established class of models of visual search are based around Feature Integration Theory (Treisman and Gelade, 1980), which has been modified and extended by Wolfe and colleagues in the Guided Search Model (Wolfe et al., 1989; Wolfe, 2014). These theories have been developed using data from visual search tasks with discrete sets of abstract items. 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. In such cases, a salient item can capture attention even without resembling the target. 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, these are predominantly qualitative models, and thus it is difficult to use them to make specific quantitative predictions.

TCS falls under a class of models that take a different approach, in that they focus 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). Similarly, work on attention has proposed the notion of 'relative features', where attention is tuned to feature relationships

search based on this general assumption. As such, it is an attractive candidate model for a formal registered replication.

A key assumption of the TCS model is 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, requiring a sequence of eye movements to search for the target in a serial manner. 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). Table 1 provides an overview of studies investigating the TCS framework to date.

The original version of the TCS model is given below:

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

This is essentially log-linear model in the number of distractors. The intercept, a , corresponds to stimuli in which only the target is present and there are no distractors. L represents the number of different types of distractors present in the display, while N_i represents the number of individual distractors present of type i . N_T is the total number of distractors, i.e., $N_T = \sum_{i=1}^L N_i$. Finally, $I_{[2,\infty]}$ is an index function that equals 0 when there $L < 2$, and 1 otherwise¹. In our paper, we will follow Buetti et al. (2019) and only consider the specific case of $L = 1$, of a target among a homogeneous set of distractors. As such, the Equation 1 simplifies to:

$$\hat{RT} = a + D \log(N_T + 1) \quad (2)$$

¹Note: Equation 1 in Lleras et al. (2020) had j here rather than L . The current version is correct.

i.e. the appearance of the target relative to distractors in the environment (Becker et al., 2014; Becker, 2010). TCS also has features in common with other models that propose parallel identification of all items in a scene, with diffusion based mechanisms for identifying targets from distractors (Moran et al., 2013, 2016). However, TCS (Lleras et al., 2020) aims to provide a unifying framework that can make quantitative behavioural predictions for visual search based on this general assumption. As such, it is an attractive candidate model for a formal registered replication.

A key assumption of the TCS model is 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, requiring a sequence of eye movements to search for the target in a serial manner. 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). Table 1 provides an overview of studies investigating the TCS framework to date.

The original version of the TCS model is essentially a (natural) log-linear model in the number of distractors. The full model contains a variable L , which represents the number of different types of distractors present in the display. However, in our paper, we will follow Buetti et al. (2019) and only consider the specific case of $L = 1$, of a target among a homogeneous set of distractors. In this case, the TCS model can be represented in the following way:

$$\hat{RT} = a + D \log(N_T + 1) \quad (1)$$

The intercept, a , corresponds to search arrays in which only the target is present and there are no distractors. N_T is the total number of distractors.

1.2. Rationale for proposed work

While many aspects of the TCS framework have been tested, with extremely promising results, there remains a great deal of scope for verification of some of

Reference	Overview
Buetti et al. (2016)	For efficient search with a specific target, there is a logarithmic relationship between distractor set size and reaction time. The steepness of this relationship is modulated by distractor-target similarity, with steeper slopes for more similar distractors.
Wang et al. (2017)	Data from homogeneous search arrays can be used to predict search reaction times in heterogeneous displays, using an equation assuming parallel, unlimited capacity, exhaustive processing, and independence of inter-item processing.
Madison et al. (2018)	Logarithmic efficiency in efficient search cannot be explained by crowding in peripheral vision.
Ng et al. (2018)	Logarithmic efficiency in efficient search cannot be explained by eye movements.
Lleras et al. (2019)	Validation of previous results showing data from homogeneous search arrays can be used to predict reaction times in heterogeneous displays. Distractor-distractor interactions can also facilitate processing.
Buetti et al. (2019)	Data from search arrays where the distractors are distinguished from the target by one feature can be used to predict search reaction times in displays with compound stimuli, defined by two features. Reaction times can be predicted using a collinear contrast integration model, which assumes that the overall target-distractor contrast is the sum of the contrasts from the two feature vectors separately.
Lleras et al. (2020)	Full proposal of the Target Contrast Signal Theory, proposing that the initial stage of processing computes a difference signal between each item in the scene and the target template, using this to determine which items in the scene are unlikely to be the target.
Ng et al. (2020)	Attention works in a two stage process, first discarding target-dissimilar distractors in a distributed, parallel way. Focused spatial attention then visits target-similar items at random.

Table 1: An overview of work on the Target Contrast Signal Theory. The key paper for our replication is highlighted.

Reference	Overview
Buetti et al. (2016)	For efficient search with a specific target, there is a logarithmic relationship between distractor set size and reaction time. The steepness of this relationship is modulated by distractor-target similarity, with steeper slopes for more similar distractors.
Wang et al. (2017)	Data from homogeneous search arrays can be used to predict search reaction times in heterogeneous displays containing images of real-world objects, using an equation assuming parallel, unlimited capacity, exhaustive processing, and independence of inter-item processing.
Madison et al. (2018)	Logarithmic efficiency in efficient search cannot be explained by crowding in peripheral vision.
Ng et al. (2018)	Logarithmic efficiency in efficient search cannot be explained by eye movements.
Lleras et al. (2019)	Validation of previous results showing data from homogeneous search arrays can be used to predict reaction times in heterogeneous displays. Distractor-distractor interactions can also facilitate processing when nearby items are similar to each other.
Buetti et al. (2019)	Data from search arrays where the distractors are distinguished from the target by one feature can be used to predict search reaction times in displays with compound stimuli, defined by two features. Reaction times can be predicted using a collinear contrast integration model, which assumes that the overall target-distractor contrast is the sum of the contrasts from the two feature vectors separately.
Lleras et al. (2020)	Full proposal of the Target Contrast Signal Theory, proposing that the initial stage of processing computes a difference signal between each item in the scene and the target template, using this to determine which items in the scene are unlikely to be the target.
Ng et al. (2020)	Attention works in a two stage process, first discarding target-dissimilar distractors in a distributed, parallel way. Focused spatial attention then visits target-similar items at random.
Xu et al. (2021)	Extension of Buetti et al. (2019) to new features (shape and texture), which combine according to a Euclidean metric (orthogonal contrast integration model).

Table 1: An overview of work on the Target Contrast Signal Theory. The key paper for our replication is highlighted.

1.2. *Rationale for proposed work*

While results to date for TCS appear promising, 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, 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; Clarke et al., 2020).

In the current manuscript, we focus on replicating and extending findings from Buetti et al. (2019). In their study, participants searched for a target in a scene of homogeneous distractors (see Figure 1). 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). 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.

We begin by verifying the analysis of Buetti et al. (2019), and then we use their data to show how we are able to extend their modelling, both by incorporating a multi-level design to predict within-subjects effects and by utilising a

116 the key findings to date, and extensions of aspects of the model. In all implementa-
117 tions of TCS so far, predictions of search efficiency (e.g. in heterogeneous scenes)
118 have been made on the average of a group of participants, using data from a dif-
119 ferent group performing a different task (e.g. searching in homogeneous scenes).
120 Thus, we know that TCS can replicate group-level averages between subjects in
121 search well, but we do not know to what extent it is also able to make predictions
122 at the individual level. This is particularly important given that conclusions based
123 on aggregate data can be different from those that take individual differences into
124 account; in one study where participants searched for a target in an array of ran-
125 domly oriented line segments, aggregating the data suggested that participants
126 were using a stochastic search model (Nowakowska et al., 2017). However, when
127 considering each participant individually, it became clear that there was a high
128 level of heterogeneity in responses, with some participants performing close to
129 optimally, and others actually performing worse than chance (Nowakowska et al.,
130 2017). Similarly striking variability has also been reported in other search studies
131 (Irons and Leber, 2016, 2018; Clarke et al., 2020).

132 Taking search time distributions into account is also important for constrain-
133 ing theories of visual search (Wolfe et al., 2010; Liesefeld and Müller, 2020): for
134 example, they have been used to help distinguish between models that make sim-
135 ilar predictions at the level of average reaction times (Moran et al., 2016, 2017).
136 Including subject and trial level data into our implementation of the TCS will
137 therefore further aid model development and assumption testing.

138 We also extend the TCS model into a Bayesian framework, where we begin
139 with existing 'prior' beliefs that are updated with data to give 'posterior' beliefs
140 that can be used for inference (McElreath, 2020). We think this has a number
141 of advantages over frequentist approaches. Perhaps most importantly, Bayesian
142 models are highly flexible. We demonstrate how we are able to specify a model
143 that is able to more accurately represent the distribution of responses (for exam-
144 ple, by specifying a response distribution that avoids predicting negative reaction
145 times) with a relatively complex model structure, that can be fit to a relatively
146 small amount of pilot data: something that would be challenging within a fre-
147 quentist framework. We also believe that Bayesian models offer very intuitive
148 methods for model testing and comparison and straightforward interpretation of
149 results, and we hope that this manuscript can act as a demonstration of these ben-
150 efits, showing how they can be applied to real scientific questions beyond the
151 simplified examples often found in textbooks or tutorials.

152 In the current manuscript, we focus on replicating and extending findings from
153 Buetti et al. (2019). In their study, participants searched for a target in a scene of

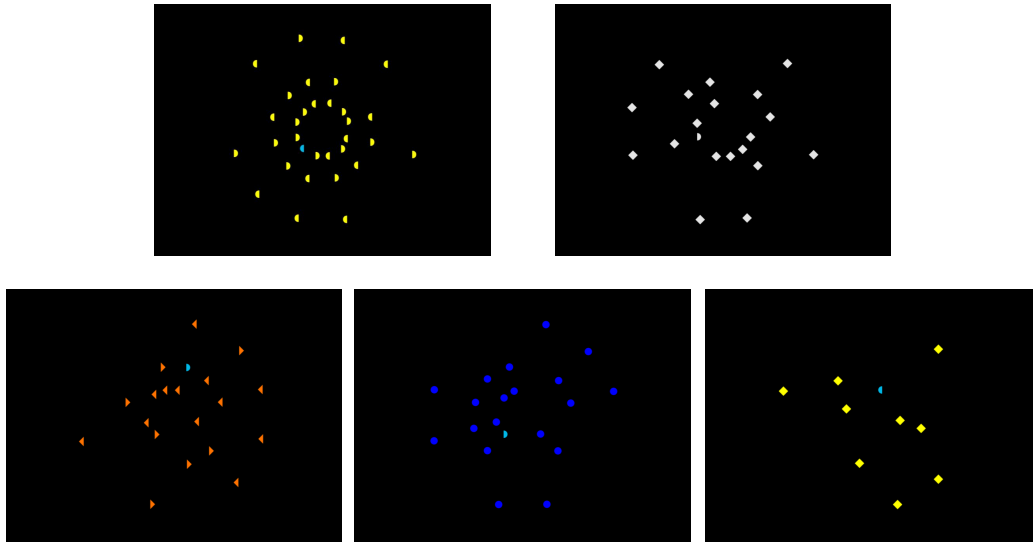


Figure 1: Example stimuli from Buetti et al. (2019) Top left: Expt 1A. Here, the target is a blue semicircle within a set of homogeneous (yellow semicircle) distractors. Top right: Expt 1B. The target is a grey semicircle in circular grey distractors. Bottom left: Expt 2A. The target is a blue semicircle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

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 then describe our proposed replication study and our registered hypotheses.

2. The Target Contrast Model

We first describe the original Target Contrast Model, as presented in Buetti et al. (2019) and verify that we can successfully replicate the original analysis (see Supplementary Materials). This is followed by Section 2.2 in which we detail our proposed modifications and show how they change the interpretation of the data collected by Buetti et al. (2019).

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

154 homogeneous distractors (see Figure 1). 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). 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. In our registered replication, we will attempt to verify the conclusions of Buetti et al. (2019), that the collinear contrast integration model does indeed offer the best characterisation of contrast signal combinations in visual search within the TCS framework.

171 We begin by verifying the analysis of Buetti et al. (2019). We then describe our proposed replication study, showing with pilot data how we are able to extend their model of how multi-dimensional contrasts are calculated, both by 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).

178 2. The Target Contrast Model

179 We first describe the original Target Contrast Model, as presented in Buetti et al. (2019) and verify that we can successfully replicate the original analysis (both using frequentist modelling and Bayesian modelling; see *Supplementary Materials*).

183 2.1. TCS modelling overview

184 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,

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). Examples of all these stimuli are shown in Figure 1.

The *Target Signal Contrast* theory is built around a linear model for predicting mean reaction times from the logarithm of the number of distractors (see Equation 2). In particular, the TCS theory allows us to predict the value of the logarithmic slope, $D_{c,s}$, in this condition based on the corresponding D_i in the single feature search 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 logarithmically with N_T , the number of distractors. (See Equation 2.) In the original model the error distribution was assumed to be normal. Thus the results of Experiment 1 were used to calculate D_i , 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 ($D_{c,s}$).

Fitting the model specified in Equation 2 to the data, we obtain the values for D_c and D_s given in Table 2. As can be seen, the more similar the distractors are to the target, the steeper the slope parameter is.

feature	D_c	a	feature	D_s	a
blue	76.8	575.1	triangle	141.1	551.6
yellow	16.0	575.1	diamond	77.2	537.7
orange	9.8	575.1	circle	62.1	555.8

Table 2: A table of D_i values for Experiment 1a and 1b

As a is the intercept in our model, it represents how long observers take to find a target when $N_T = 0$, i.e., there are no distractors. As such, it should be independent of both shape and colour. However, this leads to some ambiguity in

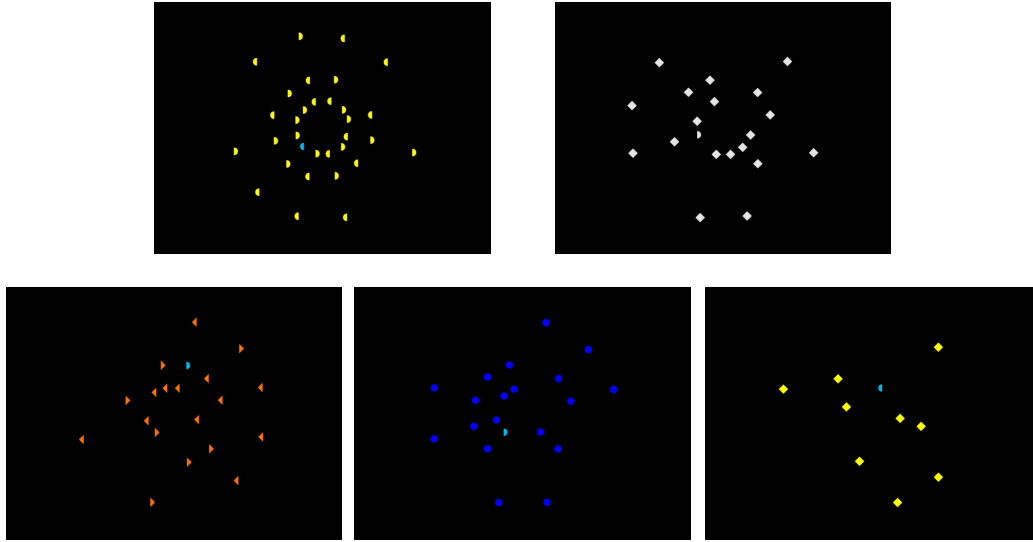


Figure 1: Example stimuli from Buetti et al. (2019) Top left: Expt 1A. Here, the target is a blue semicircle within a set of homogeneous (yellow semicircle) distractors. Top right: Expt 1B. The target is a grey semicircle in circular grey distractors. Bottom left: Expt 2A. The target is a blue semicircle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

188 with participants searching for the semicircular target within triangle, circle or di-
 189 amond distractors). In Experiments 2a, 2b and 2c, participants again searched for
 190 a cyan semicircle, but this time, the distractors differed in both shape and colour.
 191 We will refer to these conditions as *double features*. Note, unlike in standard con-
 192 junction searches, in this paradigm, the distractors are all identical with respect to
 193 these features (i.e, orange triangles). Examples of all these stimuli are shown in
 194 Figure 1. Buetti et al. (2019) also carried out a replication of their basic results
 195 using slightly different target and distractor stimuli (Experiments 3 and 4).

196 The *Target Signal Contrast* theory is built around a linear model for predicting
 197 mean reaction times from the logarithm of the number of distractors (see Equation
 198 1). In particular, the TCS theory allows us to predict the value of the logarithmic
 199 slope, $D_{c,s}$, in this condition based on the corresponding D_i in the single feature
 200 search experiments.

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; see Table 2). Here, we follow that method in order to replicate their results exactly. Unfortunately, this leads to a situation in which the model's prediction of how long participants take to find a target when $N_T = 0$ depends on the features of distractors that are not present in the stimulus! As such, in our proposed new version of the TCS theory, we will estimate a at the level of an experiment (e.g. Experiments 1 and 2).

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 3) 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):

$$D_{c,s} = \min(D_c, D_s) \quad (3)$$

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 $D_{c,s}$ can be represented as²:

$$D_{c,s} = \frac{1}{\sqrt{\frac{1}{(D_c)^2 + (D_s)^2}}} \quad (4)$$

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:

$$\frac{1}{D_{c,s}} = \frac{1}{D_c} + \frac{1}{D_s} \quad (5)$$

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 $R^2 = 0.915$.

²Note: there is a small mistake in Equation 4 in Buetti et al. (2019). The version given here is correct.

201 *2.1.1. Calculating the intercept, a , and the logarithmic slope parameter, D_i*

202 Experiments 1a and 1b and 3a and 3b were used to calculate the logarithmic
 203 slope parameter D_i . In all experiments, the number of distractors varied, allowing
 204 the data to be used to fit a log-linear model for reaction times, where reaction
 205 times increase logarithmically with N_T , the number of distractors (see Equation
 206 1). In the original model the error distribution was assumed to be normal. Thus
 207 the results of Experiments 1 and 3 were used to calculate D_i , for each type of
 208 distractor. When colour varied, we will refer to D_c , for $c = 1, 2, 3$. Similarly for
 209 shape we will denote this (D_s), and the compound features are denoted as ($D_{c,s}$).

210 Fitting the model specified in Equation 1 to the data, we obtain the values for
 211 D_c and D_s given in Table 2. As can be seen, the more similar the distractors are to
 212 the target, the steeper the slope parameter is.

feature	D_c	feature	D_s
blue	76.8	triangle	141.1
yellow	16.0	diamond	77.2
orange	9.8	circle	62.1

Table 2: A table of D_i values for Experiment 1a and 1b. See *Supplementary Materials* for full values for all experiments.

213 *2.1.2. Estimating $D_{c,s}$, the logarithmic slope parameter for compound features*

214 In the context of the current experiments, the core idea of TCS theory is that
 215 we can estimate the (natural) logarithmic slope parameter for a double feature
 216 visual search from the slopes parameters in the two independent single feature
 217 searches i.e., $D_{c,s} = f(D_c, D_s)$. Buetti et al. (2019) tested three different models
 218 for predicting D for compound colour-shape stimuli. The best feature guidance
 219 model (Equation 2) suggests that when the target and lures differ in two dimen-
 220 sions, participants will choose to attend to whichever feature dimension is the
 221 most discriminable (i.e. has the smallest D value):

$$D_{c,s} = \min(D_c, D_s) \quad (2)$$

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

$$D_{c,s} = \frac{1}{\sqrt{(\frac{1}{D_c})^2 + (\frac{1}{D_s})^2}} \quad (3)$$

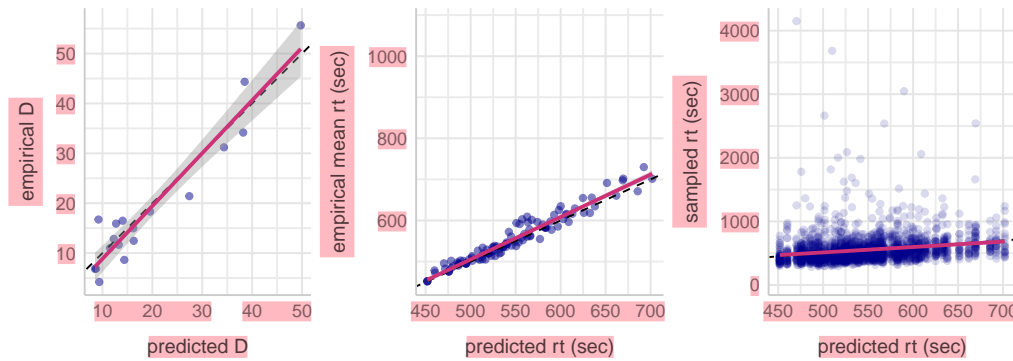


Figure 2: (left) The collinear method for calculating D offers a good prediction. (centre) Using the TCS to predict reaction times. (right) Each dot now represents a randomly sampled reaction time from an observer.

We verified we were able to replicate this result using the dataset available on OSF (<https://osf.io/f3m24/>)³ and using the exclusion criteria originally applied; see Figure 2 (left panel) and *Supplementary Materials* for details.

2.1.3. Estimating mean reaction times

Finally, we can use Equation 2 to predict mean reaction times. As can be seen in Figure 2 (centre panel), these predictions are essentially identical to the empirical RT results: $R^2 = 0.93\%$.

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 make use of multi-level modelling techniques.

It is also worth noting that relying simply on R^2 values may give a misleading view of the accuracy of the TCS model. Figure 3 (top) shows the predicted RTs (pink line) for three conditions from Experiment 2 (a, b and c) of Buetti et al.

³downloaded on 28th August 2020

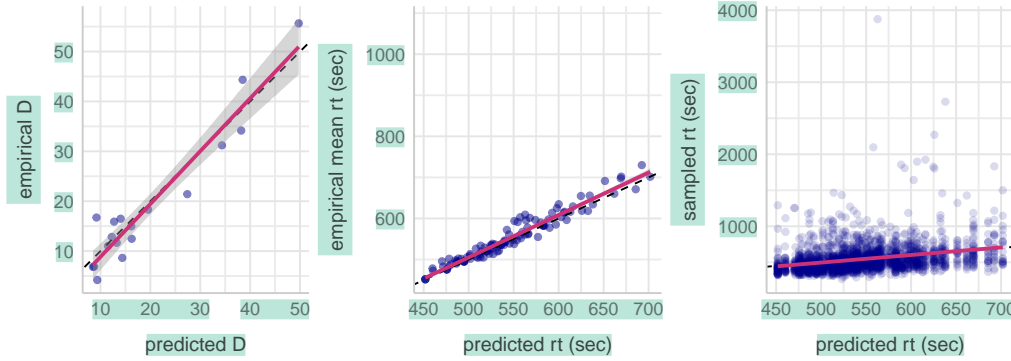


Figure 2: (left) The collinear method for calculating D offers a good prediction. (centre) Using the TCS to predict reaction times. (right) Each dot now represents a randomly sampled reaction time from an observer. Note that there is greater spread in the data points here, due to the fact that there will be trial-to-trial variability due to target position, inter-item distances, observer differences and so on.

225 Finally, the collinear contrast integration model also assumes independence of
 226 feature dimensions, but assumes that while the visual features create a multidimensional space, the contrast between them is unidimensional. As D is assumed
 227 to be inversely proportional to contrast, the equation can be written as follows:
 228

$$\frac{1}{D_{c,s}} = \frac{1}{D_c} + \frac{1}{D_s} \quad (4)$$

229 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 $R^2 = 0.915$.
 230 We verified we were able to replicate this result using the dataset available on
 231 OSF (<https://osf.io/f3m24/>)¹ and using the exclusion criteria originally applied;
 232 see Figure 2 (left panel) and *Supplementary Materials* for details. We show that
 233 we are able to do this using both the frequentist modelling approaches used in the
 234 original paper, and using Bayesian modelling.
 235

236 2.1.3. Estimating a , the intercept parameter for compound features

237 As a is the intercept of the model, it represents how long observers take to find
 238 a target when $N_T = 0$, i.e., there are no distractors. As such, it should be independent of both shape and colour, and can be thought of as the role of non-search
 239

¹downloaded on 28th August 2020

240 processes (such as motivation, motor preparation etc.) that influence reaction time.
241 In Buetti et al. (2019), a was calculated for each sub-experiment. Here, we follow
242 that method in order to replicate their results exactly.

243 2.1.4. *Estimating mean reaction times*

244 Finally, we can use Equation 1 to predict mean reaction times. As can be
245 seen in Figure 2 (centre panel), these predictions are essentially identical to the
246 empirical RT results: $R^2 = 0.93\%$.

247 2.1.5. *Discussion*

248 While TCS theory offers a good prediction of search slopes and corresponding
249 mean reaction times for double feature search, there are two related limitations.
250 Firstly, it is unable to account for individual differences between observers, only
251 the changes to the sample average. Secondly, it cannot account for the distribution
252 of reaction times over multiple trials. Figure 2 (right panel) shows clearly that
253 these factors generate high levels of variability within the individual trial-level
254 data. To address these issues, we propose adapting TCS to make use of multi-
255 level modelling techniques. Multi-level models allow us to take into account the
256 hierarchical structure of the data (i.e. that each participant completes multiple
257 trials) in a way that does not require averaging, meaning that we are able to model
258 participant variability as well as group-level effects (Gelman and Hill, 2006).

259 2.2. *A multi-level TCS*

260 Switching from a linear regression model to a multi-level model will allow
261 us to compute D for each participant, while simultaneously estimating the trial-
262 to-trial variance. We also switch from a frequentist to Bayesian framework, as
263 this allows us to naturally account for the uncertainty in the model's predictions.
264 However, switching from linear regression to a multi-level model raises the prob-
265 lem of which distribution to use for modelling reaction times. Using a normal
266 distribution is unlikely to be satisfactory, as it is unable to account for the skew
267 frequently seen in reaction time distributions, and also allows the possibility of
268 negative reaction times. We can account for both of these problems by using a
269 log-normal distribution. We will also test whether a slightly more complex ex-
270 tension of this model, the shifted lognormal model (which allows the distribution
271 to be offset to the right i.e. mimicking the patterns seen in reaction time data,
272 where valid responses begin at around 100ms) offers any improvement in model
273 fit. Note that a Wald, or inverse Gaussian distribution, would also be a reasonable
274 distribution choice for this data given that TCS is based on a diffusion process

(2019), plotted next to the mean of means for each condition. It is clear that for at least some conditions, the TCS model is doing a poor job of predicting the empirical data, even when allowing quite rigid constraints on the model; for example, the intercept, a , is set at the level of each sub-experiment and thus would be expected to be perfect. To some extent, this could reflect the fact that the R^2 measure of goodness-of-fit simply measures how tightly clustered the points are around a line with arbitrary intercept and slope. It is possible therefore for the R^2 value to be high, but for the line to deviate significantly from an intercept of 0 and a slope of 1, which is the unity line required for the predictions to be a good match to the empirical data. For our reimplementing of the TCS model, we will therefore assess model fit by computing the log marginal likelihoods via bridge sampling. This directly assesses how well our values of $D_{c,s}$ are able to predict RTs, avoiding the issues with the R^2 measure.

Finally, if we look at the different predictions made by the three proposed feature combination models, we can see that in most of the parameter space explored in Buetti et al. (2019), the models all make very similar predictions (Figure 3, bottom). In Experiment 2, we are only able to discriminate between models in the *blue* conditions, whereas in the *yellow* and *orange* conditions, they make essentially identical predictions. As such, if we want to distinguish between these three models, we will have to use combinations of features that lead to larger target-distractor differences. We will therefore use yellow, green and blue distractors in our replication study where participants are required to find a cyan target, hopefully generating a wider range of target-distractor differences.

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 multi-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, $\text{rt lognormal}(\mu, \sigma)$. 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

275 e.g. (Moran et al., 2013): we chose not to use this distribution as it often leads to
276 computational issues, which would make it harder for others to reproduce or build
277 on our approach later.

278 3. Hypotheses

279 We plan an experiment to test the extent to which the original results in Buetti
280 et al. (2019) replicate and generalise, using our new modelling approach.

281 3.1. Proposed Modifications to Experimental Design

282 In order to better test the above, and increases sensitivity, we propose to make
283 the following changes to the experiment described in Buetti et al. (2019):

284 1. **Within-subjects design.** This modification should give us greater power to
285 detect differences between different models, as well as allowing us to in-
286 vestigate how individual differences in the single-feature task might explain
287 differences in the double-feature task.

288

289 2. **Increase target-distractor similarity.** If the distractors are a very differ-
290 ent colour from the target, they may not distinguish well between different
291 contrast models. We will therefore run a version of the experiment where
292 the target is a red semicircle, with distractors being either orange, purple or
293 pink.

294

295 3.2. Registered Hypothesis

296 1. **Shifted lognormal model.** We hypothesise that a shifted lognormal model
297 will give the best fit to our single-feature data, when compared to a lognor-
298 mal and a normal model.

299

300 2. **Log-linear effect of N_T .** We will test the TCS model assumption that N_T
301 has a log-linear effect by testing models with and without the log of this
302 term. We expect that this will confirm the results previously seen in papers
303 testing TCS i.e. that the log-linear approach will be best.

304

305 3. **Contrast model comparisons.** We will test the hypothesis proposed by
306 (Buetti et al., 2019): specifically, that the *collinear contrast ingratiation*

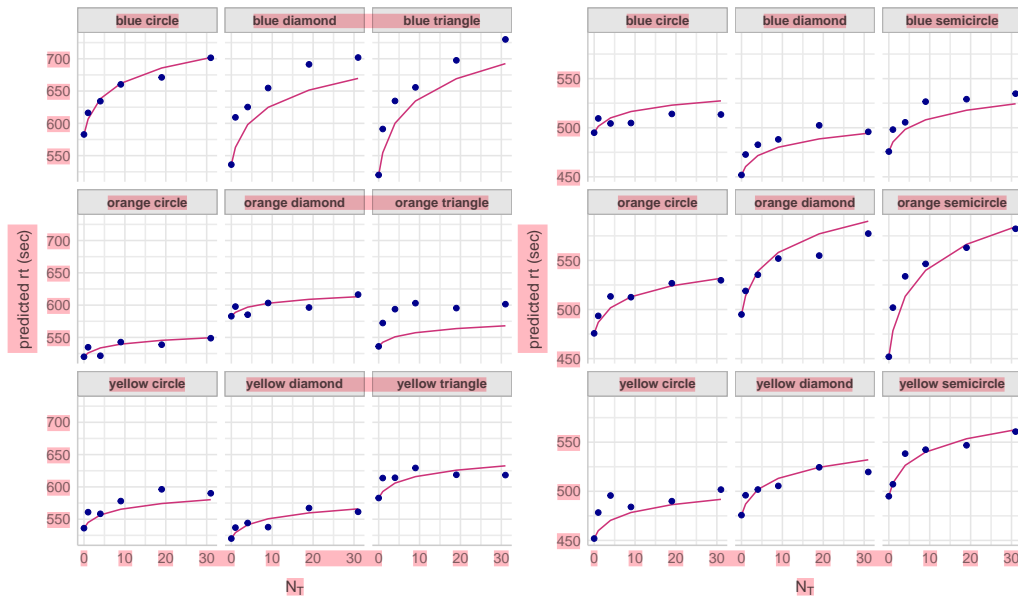


Figure 3: (top) Three conditions from Experiment 2. Blue dots represent the mean of means for each condition, while the line shows the prediction of the TCS theory. (bottom) Estimates of $D_{c,s}$ for the different colours and shape from Buetti et al. (2019). As can be seen, in two of the three colour conditions, we are unable to distinguish between best feature, collinear or orthogonal contrast models.

307 *model* outperforms the *best feature guidance*, and *orthogonal contrast com-*
308 *bination models* for the calculation of D , by calculating and comparing the
309 mean absolute prediction error for each model.

310

311 **4. Reaction time predictions.** We will further test the hypothesis proposed by
312 (Buetti et al., 2019) by testing which model gives the best prediction at the
313 trial-by-trial RT level.

314 We will test each of these hypotheses by calculating the marginal likelihood of
315 the relevant models, and then calculating the poster probabilities. This will give
316 us a probability for each model that represents the likelihood that the model gives
317 the best prediction. We will consider there to be evidence for one model over
318 the others if a given model has a probability above 90%. We will consider there
319 to be strong evidence for one model over the others if that model has a posterior
320 probability above 99%. This approach is most appropriate for our model: other
321 measures of model fit, such as AIC, require an assumption of flat priors (which is
322 not valid for multi-level models) and are based on point estimates (which is not
323 valid for Bayesian models) (McElreath, 2020).

324 3.3. *Planned Explorations*

325 We plan to investigate the effect of individual differences in this paradigm:
326 to what extent performance in the single-feature task can predict performance in
327 the double-feature task for a given individual (Buetti et al. (2019) were not able
328 to investigate this due to the between-subjects design of their study). We plan to
329 do this by specifying a more complex random effects structure for the model, that
330 allows for individual differences across different slopes for different features. This
331 allows us to then study the random effect correlation structure. However, given
332 these models can be challenging to fit, we will do this in an exploratory manner
333 after carrying out our formally registered analysis.

334 One of the benefits of using a multi-level modelling approach is that it is rel-
335 atively easy to extend to incorporate other factors that may contribute to reaction
336 times, such as eccentricity and inter-item distance, which may help to explain
337 behaviour further. To demonstrate this, we will also run exploratory analyses in-
338 cluding a factor for which ring the target is in to assess whether this improves
339 model fit or affects any of the conclusions that can be drawn from the model.

340 3.4. *Pilot Experiment*

341 Full details of a pilot experiment with $n = 4$ participants (960 trials each) using
342 our proposed analyses can be found in *Supplementary Materials*. This suggests

time data, where valid responses begin at around 100ms) offers any improvement in model fit.

We also consider whether it is possible that the brain uses a combination of the different models proposed in Buetti et al. (2019): it is possible that multiple mechanisms could be operating at once. We therefore tested a linear transform model alongside the three original models for predicting D for compound colour-shape stimuli.

Throughout the paper, we will make use of the 53% and 97% *highest posterior density intervals* (HPDI) to summarise probability distributions. These can be thought of as the smallest interval that, given our data and assumptions, contain 53% and 97% of the probability mass. Finally, when comparing models, we assess the goodness-of-fit of the predicted reaction models using bridge sampling to calculate the log marginal likelihoods.

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, (iii) shifted lognormal distribution (please see Supplementary Materials for full details of our modelling approach, including prior predictive checks and model fit diagnostics). The three models were compared using bridge sampling (Meng and Wong, 1996) to calculate the log marginal likelihoods, and the results showed that the shifted-lognormal model (illustrated in Figure 4) was given 100% of the weight⁴. Therefore, we used only this model for the rest of the analysis.

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

The first step in predicting the values for $D_{c,s}$ is to extract the logarithmic slope parameters D_c and D_s from our model outlined above (Figure 5 (top)). This is the same as the original TCS, except here we use samples from the posterior distributions, rather than just the maximum likelihood fit.

We then combine D_c and D_s using the best feature, orthogonal contrast, and collinear contrast integration models to get three different predictions for $D_{c,s}$. In addition, we also calculated a linear transform model to generate a fourth prediction for $D_{c,s}$. These are illustrated in Figure 5 (bottom).

⁴We also tested a version of the shifted lognormal model that did not use the logarithm of the number of distractors i.e. testing this assumption of the TCS model, but found no support for this model being better: see Supplementary Materials for further details.

343 that even with a small sample, we can convincingly demonstrate H1 and H2, how-
344 ever more data will be required to discriminate between the models in H3.

345 4. General Methods

346 4.1. Sample Size: Participants and Trials

347 We plan to test 40 participants during the experiment. Our pilot experiment
348 shows that H1 and H2 are easily demonstrated with 10 times less data, and Buetti
349 et al. (2019) used 20 participants per experiment. Our sample size will therefore be
350 in line with previous work testing H3 and H4. Ethical approval for the study was
351 granted by the University of Aberdeen (application number PEC/4677/2021/2).

352 4.2. Stimuli

353 The targets and distractors are randomly assigned to the display based on an
354 invisible grid. Within each quadrant of the screen, there are three 'spokes' each
355 with four possible target positions (starting from the centre of the screen and mov-
356 ing outwards), creating 36 different target positions in total, in three concentric
357 circles. A small amount of jitter is added to each possible position to make the
358 target locations less predictable.

359 **Distractor and target types:** we will replicate the distractor types used in
360 Buetti et al. (2019), apart from that we will change one distractor colour (from
361 blue to pink) to allow us to discriminate better between different models of the
362 data (see above). There are six single-feature conditions (purple, orange and pink
363 distractors and triangle, circle and diamond distractors) and nine double-feature
364 conditions (all possible pairings of the single-feature conditions). The target is al-
365 ways a red semicircle, except in the trials where the distractors are single-feature
366 shapes (triangles, circles and diamonds) in which case the target is a white semi-
367 circle.

368 **Set sizes:** we will run all the distractor set sizes used in Buetti et al. (2019) (1,
369 4, 9, 19 and 31). We will also run target-only 'zero distractor' trials (60 in total,
370 with 12 being the white semicircle target and the remainder the red semicircle
371 target).

372 The experiments were programmed in PsychoPy and Pavlovia (Peirce et al.,
373 2019). Stimuli were pre-made to generate search array images with 1920×1080
374 resolution.

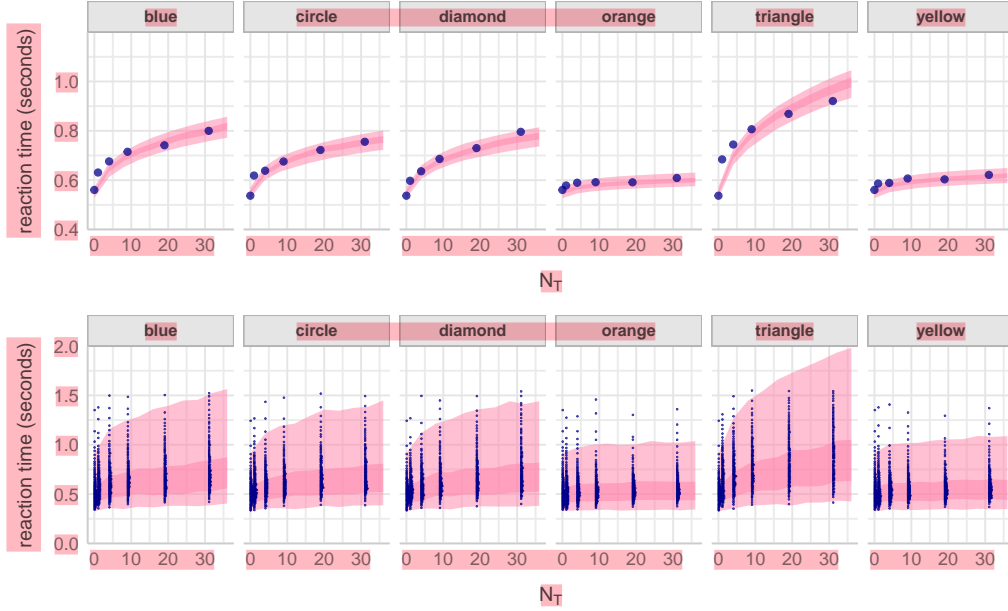


Figure 4: (*top*) The shaded regions show the shifted lognormal model's estimate (53% and 97% intervals) of the average participant's mean reaction time, while the points show the empirical mean reaction time. (*bottom*): The shaded regions now indicate the distribution of reaction times (over a new simulated group of participants) generated by the model. The points now represent the 100 quantiles from the empirical data.

The R^2 values for the four approaches are shown in Table 3. We can see that as with the original model, we still find the *best feature* is the worst of the three methods, at least in terms of the raw R^2 value. However, it is less clear which of the other methods gives the best performance. While the collinear contrast approach offers the highest R^2 (HPDI: 0.827-0.965), it consistently predicts values that are too low. The orthogonal contrast method gives more accurate predictions, albeit with a slightly lower R^2 (HPDI: 0.748-0.949). Interestingly, the linear combination model also has a high R^2 value (HPDI: 0.849-0.974). In order to distinguish between the different methods, we instead compare goodness-of-fit of the different models for predicting reaction times (see below).

2.2.3. Estimating other parameters and predicting reaction times

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

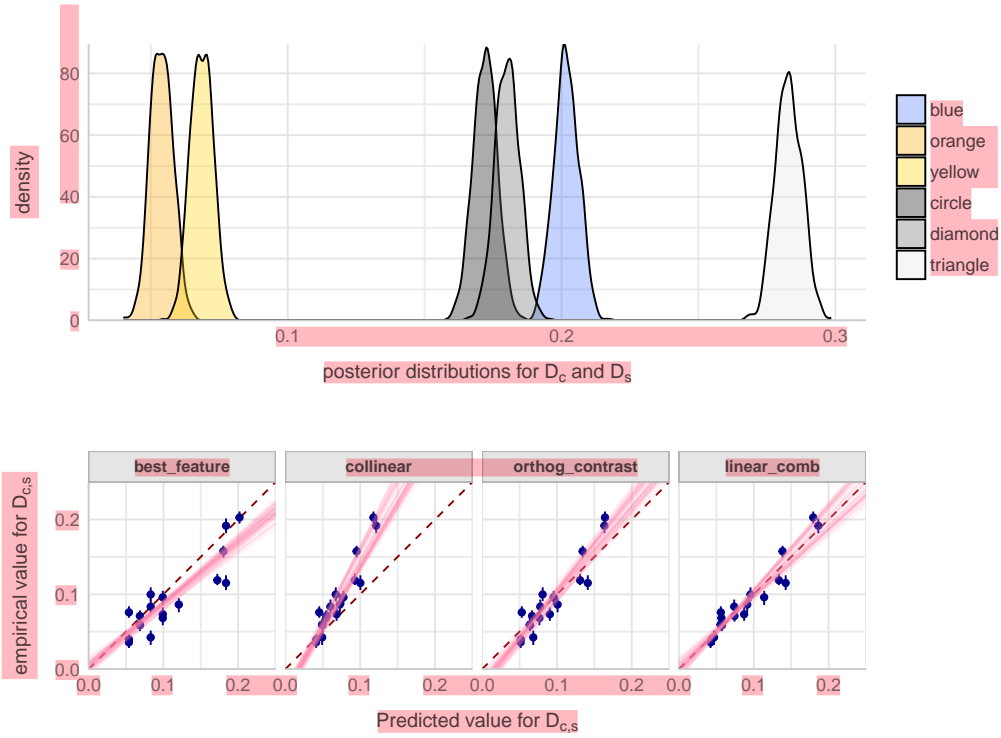


Figure 5: (*top*) Posterior probability density functions for each D_c and D_s in Experiment 1. As with the original TCS model, when searching for a blue semicircle, triangular distractors are the hardest, while yellow and orange distractors are the easiest.) (*bottom*) Predicting $D_{c,s}$ from D_c and D_s . Crosshairs indicate 97.5 HPDI for estimates and predictions, while the pink region shows the HPDI for the best fit line.

- α - the intercept for the shift (ndt) 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 we will simply use the values for the above parameters obtained from Experiment 1. Once these have been set, we can use the specified model to generate the HPDI intervals for the expected average performance and the full distribution of reaction times over a number of simulated new observers (Figure 6).

Method	Intercept	Slope	R^2	Lower	Upper	MW(3)	MW(4)
best feature	0.003	0.831	0.751	0.550	0.919	32.1%	24.0%
orthogonal contrast	-0.019	1.206	0.857	0.748	0.949	33.7%	25.1%
collinear	-0.033	1.812	0.902	0.827	0.965	34.2%	25.5%
linear transform	0.000	1.005	0.916	0.849	0.974	-	25.5%

Table 3: A table to show the logarithmic slope parameters for the four different D models (best feature, orthogonal contrast, collinear contrast and linear transform) along with the R^2 values for the four approaches. The final columns show model weights (MW) based on bridge sampling for a) the three original reaction time models MW(3) and b) the three original reaction time models plus the linear transform model MW(4).

We can summarise how well the model predicts the mean reaction times (Figure 6, left) by repeatedly i) sampling from the posterior, ii) computing the predicted mean reaction times, and iii) calculating a measure of goodness-of-fit of the model predicting RTs from the $D_{c,s}$ values.

At first glance, our model appears to be offering a poorer fit with the data than the original TCS. This is due to our decision to estimate a (the intercept) from the Experiment 1 data, whereas Buetti et al. (2019) estimated three values for a , one for each sub-experiment, and these values were estimated from the Experiment 2 empirical data. With the within-subject design we propose for our replication study, this issue should not arise.

However, our goodness-of-fit measures showed that there was little difference between our models (see Table 3): the posterior model probability of the collinear contrast model was 34.2%, compared to 33.7% for the orthogonal contrast model and 32.1% for the best feature model. There is therefore no strong evidence for one model over another. This conclusion holds if we also include our new linear transform model in the comparison (see Table 3). Our results contrast with Buetti et al. (2019), who concluded that the collinear contrast integration model offered the best fit to their data. They reasoned that this was due to the independence of the two features in their study, and they would not expect this result to generalize to non-independent features, in which case they expect the *orthogonal* contrast combination model to offer the best predictions. However, our analysis suggests that there is no strong evidence that any of the models tested is superior to the others.

Overall, while our new model appears to fit slightly more poorly 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. In addition, as we discuss previously, the R^2

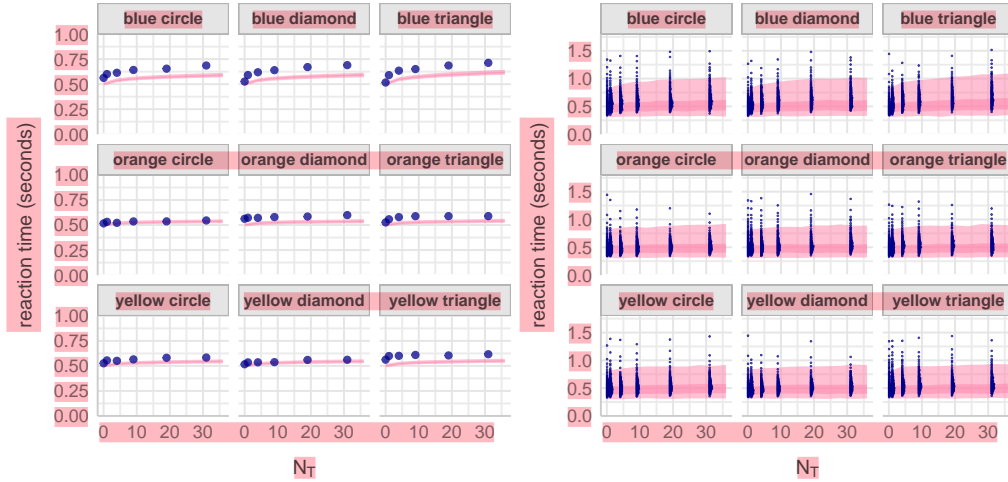


Figure 6: (left) HPDI and empirical mean reaction times for each condition in Experiment 2, using the collinear contrast integration model to predict D values. (right) Similar, for the full distribution of reaction times over trials.

value may not capture goodness-of-fit optimally, and more sensitive goodness-of-fit metrics suggest that the models have equivalent predictive value.

2.3. 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 accurately model the skew seen in the empirical data, and avoid predicting negative response times. Our modelling has also highlighted other issues that we will address during our replication study: firstly, using a within-subjects design allows a to be calculated in a more principled manner, and secondly, we will choose distractor features to give the highest possible D sensitivity, and thus give us greater power to detect true differences between models.

3. Hypotheses

We plan an experiment to test the extent to which the original results in Buetti et al. (2019) replicate and generalise.

3.1. Proposed Modifications to Experimental Design

In order to better test the above, and increase sensitivity, we propose to make the following changes to the experiment described in Buetti et al. (2019):

1. **Within-subjects design.** This modification should give us greater power to detect differences between different models, as well as allowing us to investigate how individual differences in the single-feature task might explain differences in the double-feature task.
2. **Increase target-distractor similarity.** Our reanalysis of Buetti et al. (2019) indicates that the distractors with very different colours from the target do not distinguish well between different contrast models. We will therefore run a version of the experiment where the target is a red semicircle, with yellow, orange and pink distractors.

3.2. Registered Hypothesis

1. **Shifted lognormal model.** We hypothesise that a shifted lognormal model will give the best fit to our single-feature data, when compared to a lognormal and a normal model.
2. **Log-linear effect of N_T .** We will test the TCS model assumption that N_T has a log-linear effect by testing models with and without the log of this term.
3. **Contrast model comparisons.** We will test the hypothesis proposed by (Buetti et al., 2019): specifically, that the *collinear contrast ingratiation model* outperforms the *best feature guidance*, and *orthogonal contrast combination models*. We will also test whether a *linear transform model* is better able to predict the data.

We will test each of these hypotheses by calculating the marginal likelihood of the relevant models, and then calculating the poster probabilities. This will give us a probability for each model that represents the likelihood that the model gives the best prediction. We will consider there to be evidence for one model over the others if a given model has a probability above 90%. We will consider there to be strong evidence for one model over the others if that model has a posterior probability above 99%.

3.3. *Planned Explorations*

We plan to investigate the effect of individual differences in this paradigm: to what extent performance in the single-feature task can predict performance in the double-feature task for a given individual (Buetti et al. (2019) were not able to investigate this due to the between-subjects design of their study). We plan to do this by specifying a more complex random effects structure for the model, that allows for individual differences across different slopes for different features. This allows us to then study the random effect correlation structure. However, given these models can be challenging to fit, we will do this in an exploratory manner after carrying out our formally registered analysis.

4. **General Methods**

4.1. *Sample Size: Participants and Trials*

To determine our sample size, we carried out a simulation study to estimate the effect of reducing the number of trials and increasing the number of participants on our ability to accurately measure the D_i . While our reanalysis of Buetti et al. (2019)’s data is unable to distinguish between the three models of contrast combination, we believe that this is due to the feature values used in their study, rather than low statistical power from sampling (see above and Figure 3). Therefore, we will base our sensitivity analysis around the width of the 97% HPDI when estimating D_{blue} . We have chosen to look at this single feature level as the blue distractor trials offer the greatest discrimination between the three contrast models.

We carried out our simulation sensitivity analysis by generating synthetic data from the shifted lognormal model outlined in Section 2.2, Figure 4, for the blue distractor conditions. This was done for several different sample sizes, both in terms of number of trials and number of participants. For each synthetic dataset, we then refit the model and calculated the width of the 97% HPDI (see Figure 7).

As we are moving from the between-subjects design used by Buetti et al. (2019) to a within-subjects design in which each participant sees all combinations of features, we prefer to decrease the number of trials per person \times condition from 40 to six. Increasing the number of participants from 20 to 150 will allow us to maintain sufficient accuracy in the estimating of logarithmic search slopes.

Ethical approval for the study was granted by the University of Aberdeen (application number PEC/4677/2021/2).

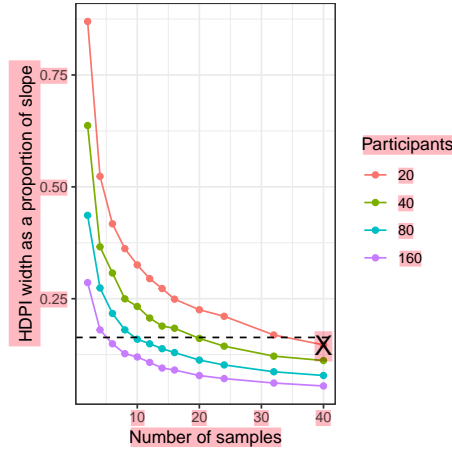


Figure 7: Simulation sensitivity analysis, showing the 97% HPDI width for the blue distractor condition in a shifted lognormal model, across different numbers of trials and participants. The X shows the real HPDI width as a proportion of slope for the blue distractor condition in Buetti et al. (2019).

4.2. Stimuli

The targets and distractors are randomly assigned to the display based on an invisible grid. Within each quadrant of the screen, there are three 'spokes' each with four possible target positions (starting from the centre of the screen and moving outwards), creating 48 different target positions in total, in four concentric circles. A small amount of jitter is added to each possible position to make the target locations less predictable.

Distractor and target types: we will replicate the distractor types used in Buetti et al. (2019), apart from that we will change one distractor colour (from blue to pink) to allow us to discriminate better between different models of the data (see above). There are six single-feature conditions (yellow, orange and pink distractors and triangle, circle and diamond distractors) and nine double-feature conditions (all possible pairings of the single-feature conditions). The target is always a red semicircle, except in the trials where the distractors are single-feature shapes (triangles, circles and diamonds) in which case the target is a white semicircle.

Set sizes: we will run all the distractor set sizes used in Buetti et al. (2019) (1, 4, 9, 19 and 31). We will also run target-only 'zero distractor' trials (30 in total, with 6 being the white semicircle target and the remainder the light blue semicircle target).

The experiments were programmed in PsychoPy and Pavlovia (Peirce et al., 2019). Stimuli were pre-made to generate search array images with 1920×1080 resolution.

4.3. Procedure

The experiment will take place in a controlled lab setting...

Participants will view a fixation cross before viewing a search array: they will press the space bar to continue to the trial. Participants will be told to search for the target among distractors and report if the semicircle target points to the left or right, by pressing either the ‘f’ or ‘j’ key respectively on their keyboard. They will first complete 20 practice trials where they will receive feedback immediately after completing each trial. In the real experimental trials, participants will receive feedback on their average accuracy and reaction time after each block of 30 trials. Participants will complete 19 blocks of trials (570 trials overall) with the order of the stimuli being fully randomised.

In both the practice and experimental trials, the search display will always remain on screen until a response is made, or until 5 seconds had passed.

4.4. Data Pre-processing

Only participants who complete the full experiment will be considered candidates for inclusion in the data analysis. We will apply the same inclusion criteria as the original paper: participants will only be included if their search accuracy over 90% and their average response time is not smaller or larger than two standard deviations from the group average response time.

For participants included in the analysis, we will apply the data cleaning used in Section 2.2 i.e. removing the top and bottom 1% of their data.

4.5. Analysis Plan

We will follow the analysis given in Section 2.2. Full details of the code used are provided in the *Supplementary Material* and on Github at https://github.com/scienceanna/TCS_Bayesian.

5. Results

– blank –

6. General Discussion

– blank –

375 4.3. Procedure

376 Participants will complete the experiment in the laboratory, sitting at a viewing
377 distance of 45cm from the screen. They will view a fixation cross before viewing
378 a search array: they will press the space bar to continue to the trial. Participants
379 will be told to search for the target among distractors (i.e. will be told to per-
380 form a singleton search) and report if the semicircle target points to the left or
381 right, by pressing either the 'f' or 'j' key respectively on their keyboard. They
382 will first complete 16 practice trials where they will receive feedback immedi-
383 ately after completing each trial. In the real experimental trials, participants will
384 receive feedback on their average accuracy and reaction time after each block of
385 120 trials. Participants will complete 8 blocks of trials (960 trials overall i.e. 192
386 trials in each of 5 experiments, consisting of 5 set sizes x 3 distractor conditions
387 x 12 repeats + 12 zero distractor trials) with the order of the displays being fully
388 randomised i.e. all different conditions will be completely intermixed (in the ter-
389 minology of Buetti et al. (2019), participants will complete all of Experiments 1
390 and 2 in the same testing session, rather than having different participants com-
391 plete each separate sub-experiment).

392 In both the practice and experimental trials, the search display will always
393 remain on screen until a response is made, or until 5 seconds had passed.

394 4.4. Data Pre-processing

395 Only participants who complete the full experiment will be considered candi-
396 dates for inclusion in the data analysis. We will apply the same inclusion criteria
397 as the original paper: participants will only be included if their search accuracy
398 over 90% and their average response time is not smaller or larger than two stan-
399 dard deviations from the group average response time.

400 For participants included in the analysis, we will apply the data cleaning used
401 in the pilot data analysis i.e. removing the top and bottom 1% of their data.

402 4.5. Analysis Plan

403 All analysis will be carried out using R (vx.xx)², brms (v.xx.xx) and rStan
404 (vx.xxxx) As discussed above, we will use a mixed-effect models with either nor-
405 mal, lognormal or shifted lognormal distributions.

406 Please see the analysis of our pilot data for a full implementation of our anal-
407 ysis pipeline, including all code (available on Github at [https://github.com/](https://github.com/scienceanna/TCS_Bayesian)
408 scienceanna/TCS_Bayesian).

²Version numbers will be recorded upon completion of final analysis.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by an Economic and Social Research Council grant (ES/S016120/1) to ADFC and employing AN.

References

- Stefanie I Becker. The role of target–distractor relationships in guiding attention and the eyes in visual search. *Journal of Experimental Psychology: General*, 139(2):247, 2010.
- Stefanie I Becker, Christian Valuch, and Ulrich Ansorge. Color priming in pop-out search depends on the relative color of the target. *Frontiers in psychology*, 5:289, 2014.
- 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.
- Kyle R Cave and Jeremy M Wolfe. Modeling the role of parallel processing in visual search. *Cognitive psychology*, 22(2):225–271, 1990.
- Alasdair DF Clarke, JL Irons, Warren James, Andrew B Leber, and Amelia R Hunt. Stable individual differences in strategies within, but not between, visual search tasks. *Quarterly Journal of Experimental Psychology*, 2020.
- 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.

409 5. Results

410 – blank –

411 6. General Discussion

412 – blank –

413 Conflict of interest

414 The authors declare that they have no conflict of interest.

415 Acknowledgements

416 This work was supported by an Economic and Social Research Council grant
417 (ES/S016120/1) to ADFC and employing AN.

418 References

419 Stefanie I Becker. The role of target–distractor relationships in guiding attention
420 and the eyes in visual search. *Journal of Experimental Psychology: General*,
421 139(2):247, 2010.

422 Stefanie I Becker, Christian Valuch, and Ulrich Ansorge. Color priming in pop-
423 out search depends on the relative color of the target. *Frontiers in psychology*,
424 5:289, 2014.

425 Simona Buetti, Deborah A Cronin, Anna M Madison, Zhiyuan Wang, and Ale-
426 jandro Lleras. Towards a better understanding of parallel visual processing in
427 human vision: Evidence for exhaustive analysis of visual information. *Journal*
428 *of Experimental Psychology: General*, 145(6):672, 2016.

429 Simona Buetti, Jing Xu, and Alejandro Lleras. Predicting how color and shape
430 combine in the human visual system to direct attention. *Scientific reports*, 9(1):
431 1–11, 2019.

432 Kyle R Cave and Jeremy M Wolfe. Modeling the role of parallel processing in
433 visual search. *Cognitive psychology*, 22(2):225–271, 1990.

- 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 Madison, Alejandro Lleras, and Simona Buetti. The role of crowding in parallel search: Peripheral pooling is not responsible for logarithmic efficiency in parallel search. *Attention, Perception, & Psychophysics*, 80(2):352–373, 2018.
- Xiao-Li Meng and Wing Hung Wong. Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, pages 831–860, 1996.
- Gavin JP Ng, Simona Buetti, Trisha N Patel, and Alejandro Lleras. Prioritization in visual attention does not work the way you think it does. *Journal of Experimental Psychology: Human Perception and Performance*, 2020.
- Gavin Jun Peng Ng, Alejandro Lleras, and Simona Buetti. Fixed-target efficient search has logarithmic efficiency with and without eye movements. *Attention, Perception, & Psychophysics*, 80(7):1752–1762, 2018.

- 434 Alasdair DF Clarke, JL Irons, Warren James, Andrew B Leber, and Amelia R
435 Hunt. Stable individual differences in strategies within, but not between, visual
436 search tasks. *Quarterly Journal of Experimental Psychology*, 2020.
- 437 John Duncan and Glyn W Humphreys. Visual search and stimulus similarity.
438 *Psychological review*, 96(3):433, 1989.
- 439 Andrew Gelman and Jennifer Hill. *Data analysis using regression and multi-*
440 *level/hierarchical models*. Cambridge university press, 2006.
- 441 Johan Hulleman and Christian NL Olivers. On the brink: The demise of the item
442 in visual search moves closer. *Behavioral and Brain Sciences*, 40, 2017.
- 443 Jessica L Irons and Andrew B Leber. Choosing attentional control settings in a
444 dynamically changing environment. *Attention, Perception, & Psychophysics*,
445 78(7):2031–2048, 2016.
- 446 Jessica L Irons and Andrew B Leber. Characterizing individual variation in the
447 strategic use of attentional control. *Journal of Experimental Psychology: Hu-*
448 *man Perception and Performance*, 44(10):1637, 2018.
- 449 Laurent Itti and Christof Koch. A saliency-based search mechanism for overt and
450 covert shifts of visual attention. *Vision research*, 40(10-12):1489–1506, 2000.
- 451 Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual
452 attention for rapid scene analysis. *IEEE Transactions on pattern analysis and*
453 *machine intelligence*, 20(11):1254–1259, 1998.
- 454 Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards
455 the underlying neural circuitry. In *Matters of intelligence*, pages 115–141.
456 Springer, 1987.
- 457 Kathryn Koehler, Fei Guo, Sheng Zhang, and Miguel P Eckstein. What do
458 saliency models predict? *Journal of vision*, 14(3):14–14, 2014.
- 459 Heinrich René Liesefeld and Hermann J Müller. A theoretical attempt to revive
460 the serial/parallel-search dichotomy. *Attention, Perception, & Psychophysics*,
461 82(1):228–245, 2020.
- 462 Heinrich René Liesefeld, Rani Moran, Marius Usher, Hermann J Müller, and
463 Michael Zehetleitner. Search efficiency as a function of target saliency: The

- 464 transition from inefficient to efficient search and beyond. *Journal of Experi-*
465 *mental Psychology: Human Perception and Performance*, 42(6):821, 2016.
- 466 Alejandro Lleras, Zhiyuan Wang, Anna Madison, and Simona Buetti. Predicting
467 search performance in heterogeneous scenes: Quantifying the impact of homo-
468 geneity effects in efficient search. *Collabra: Psychology*, 5(1), 2019.
- 469 Alejandro Lleras, Zhiyuan Wang, Gavin Jun Peng Ng, Kirk Ballew, Jing Xu, and
470 Simona Buetti. A target contrast signal theory of parallel processing in goal-
471 directed search. *Attention, Perception, & Psychophysics*, pages 1–32, 2020.
- 472 Anna Madison, Alejandro Lleras, and Simona Buetti. The role of crowding in par-
473 allel search: Peripheral pooling is not responsible for logarithmic efficiency in
474 parallel search. *Attention, Perception, & Psychophysics*, 80(2):352–373, 2018.
- 475 Richard McElreath. *Statistical rethinking: A Bayesian course with examples in R*
476 *and Stan*. Chapman and Hall/CRC, 2020.
- 477 Rani Moran, Michael Zehetleitner, Hermann J Mueller, and Marius Usher. Com-
478 petitive guided search: Meeting the challenge of benchmark rt distributions.
479 *Journal of Vision*, 13(8):24–24, 2013.
- 480 Rani Moran, Michael Zehetleitner, Heinrich René Liesefeld, Hermann J Müller,
481 and Marius Usher. Serial vs. parallel models of attention in visual search: ac-
482 counting for benchmark rt-distributions. *Psychonomic bulletin & review*, 23(5):
483 1300–1315, 2016.
- 484 Rani Moran, Heinrich René Liesefeld, Marius Usher, and Hermann J Muller. An
485 appeal against the item’s death sentence: Accounting for diagnostic data pat-
486 terns with an item-based model of visual search. *Behavioral and Brain Sci-*
487 *ences*, 40:e148, 2017.
- 488 Gavin JP Ng, Simona Buetti, Trisha N Patel, and Alejandro Lleras. Prioritiza-
489 tion in visual attention does not work the way you think it does. *Journal of*
490 *Experimental Psychology: Human Perception and Performance*, 2020.
- 491 Gavin Jun Peng Ng, Alejandro Lleras, and Simona Buetti. Fixed-target efficient
492 search has logarithmic efficiency with and without eye movements. *Attention,*
493 *Perception, & Psychophysics*, 80(7):1752–1762, 2018.

- 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, Klinto Law, and Ernst Niebur. Modeling the role of salience in the allocation of overt visual attention. *Vision research*, 42(1):107–123, 2002.
- Jonathan Peirce, Jeremy R Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. Psychopy2: Experiments in behavior made easy. *Behavior research methods*, 51(1):195–203, 2019.
- 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.

- 494 Anna Nowakowska, Alasdair DF Clarke, and Amelia R Hunt. Human visual
495 search behaviour is far from ideal. *Proceedings of the Royal Society B: Biological Sciences*, 284(1849):20162767, 2017.
- 497 Derrick Parkhurst, Klinto Law, and Ernst Niebur. Modeling the role of salience
498 in the allocation of overt visual attention. *Vision research*, 42(1):107–123, 2002.
- 499 Jonathan Peirce, Jeremy R Gray, Sol Simpson, Michael MacAskill, Richard
500 Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv.
501 Psychopy2: Experiments in behavior made easy. *Behavior research methods*,
502 51(1):195–203, 2019.
- 503 Dragan Rangelov, Hermann J Müller, and Michael Zehetleitner. Failure to pop
504 out: Feature singletons do not capture attention under low signal-to-noise ratio
505 conditions. *Journal of Experimental Psychology: General*, 146(5):651, 2017.
- 506 Benjamin W Tatler. The central fixation bias in scene viewing: Selecting an op-
507 timal viewing position independently of motor biases and image feature distri-
508 butions. *Journal of vision*, 7(14):4–4, 2007.
- 509 Benjamin W Tatler, Mary M Hayhoe, Michael F Land, and Dana H Ballard. Eye
510 guidance in natural vision: Reinterpreting salience. *Journal of vision*, 11(5):
511 5–5, 2011.
- 512 Anne M Treisman and Garry Gelade. A feature-integration theory of attention.
513 *Cognitive psychology*, 12(1):97–136, 1980.
- 514 Zhiyuan Wang, Simona Buetti, and Alejandro Lleras. Predicting search perfor-
515 mance in heterogeneous visual search scenes with real-world objects. *Collabra:*
516 *Psychology*, 3(1), 2017.
- 517 Jeremy M Wolfe. What can 1 million trials tell us about visual search? *Psycho-*
518 *logical Science*, 9(1):33–39, 1998.
- 519 Jeremy M Wolfe. Approaches to visual search: Feature integration theory and
520 guided search. *Oxford handbook of attention*, pages 11–55, 2014.
- 521 Jeremy M Wolfe, Kyle R Cave, and Susan L Franzel. Guided search: an alterna-
522 tive to the feature integration model for visual search. *Journal of Experimental*
523 *Psychology: Human perception and performance*, 15(3):419, 1989.

524 Jeremy M Wolfe, Evan M Palmer, and Todd S Horowitz. Reaction time distri-
525 butions constrain models of visual search. *Vision research*, 50(14):1304–1311,
526 2010.

527 Zoe Jing Xu, Alejandro Lleras, and Simona Buetti. Predicting how surface texture
528 and shape combine in the human visual system to direct attention. *Scientific*
529 *reports*, 11(1):1–13, 2021.

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