

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

6 field has been between efficient (also referred to as parallel or pop-out) and in-
7 efficient (serial) search. These are often studied in the context of the regression
8 slope between the number of distractors and mean reaction time, which has been
9 termed the *search slope*. When the search slope is shallow (usually positive, but
10 occasionally negative e.g. (?)), the search is called efficient or parallel, and the
11 addition of more non-target distractors has little impact on an observers difficulty
12 in finding a target. When the slope is steeper, each additional distractor has a no-
13 ticeable impact on increasing difficulty, and the search is described as inefficient
14 or serial. However, the distinction between these types of search is often less clear
15 in real experimental data, with a range of different search slopes being seen for
16 different types of targets and distractors (????). Recent work has also attempted
17 to model the variation in search slopes at the boundary between inefficient and
18 efficient search (?).

19 In the current study, we are interested in what has traditionally been termed
20 efficient or parallel search, and the factors that affect search slope in these condi-
21 tions. Recent work has suggested that for efficient search, there is a logarithmic
22 relationship between distractor set size and reaction time, and that this relationship
23 can be modified by target-distractor similarity (?), providing evidence that search
24 behaviour in parallel search is more complex than has previously been assumed.
25 This observation has formed the basis of the ‘Target Contrast Signal (TCS) The-
26 ory’ (?), which aims to provide a means of predicting observer search slopes for
27 new search arrays by quantifying target-distractor differences. For example, by
28 measuring search slopes for conditions in which the distractors differ from the
29 target along a *single feature* (e.g. colour *or* shape), it has been shown that you
30 can predict search times for arrays in which the target differs from the distrac-
31 tors along two features (e.g., colour *and* shape) which we refer to here as *double*
32 *feature* search. (?) (but similar paradigms have been known by other names e.g.
33 ‘redundant feature search’ (??)). Here, we aim to replicate and extend this work
34 both theoretically and empirically, to test the generalisability of the TCS model,
35 and to suggest ways in which the TCS model could be modified to generate better
36 predictions.

37 1.1. Previous Work

38 Many different forms of visual search models have been proposed. One well
39 developed class of models are the saliency models, which aim to predict eye
40 movements during scene viewing, including visual search. They rest on the as-
41 sumption that fixations are directed to objects or locations that are most dissimilar
42 to the background or other objects in the visual display (???). While the original

43 saliency model was able to predict fixation allocation in a visual search task above
44 chance (?), further research demonstrated that a comparable level of performance
45 could be achieved using a simple central fixation bias heuristic (?). The saliency
46 models have since been extended and improved (see for example ?): however, the
47 main issue with this family of models remains their limited usability in complex
48 real-life search arrays (??), and even in abstract laboratory search arrays (?). In
49 addition, in most instances of visual search, the target is clearly defined (i.e. the
50 goal is to find a specific object) and inspecting the most salient areas of the display
51 may in these cases be inefficient. Finally, by focusing on eye movements, these models
52 do not necessarily provide a theoretical framework for the cognitive processes underlying
53 visual search.

54 Perhaps the most established class of models of visual search are based around
55 Feature Integration Theory (?), which has been modified and extended by Wolfe
56 and colleagues in the Guided Search Model (??). These theories have been developed
57 using data from visual search tasks with discrete sets of abstract items. These models
58 combine top-down influences (how closely an item resembles the observer's goal) with
59 bottom-up image properties. For example, if one's goal (top-down processing) is to find
60 a red horizontal bar, all the red and horizontal items in a visual search display will be
61 given greater weight than distractors (e.g. vertical and blue items) in the model. The
62 salience of a given object in the display (how distinctive it is from the surrounding
63 objects) also activates bottom-up processing. For instance, a blue item among red items
64 is ranked higher than red among orange items. In such cases, a salient item can capture
65 attention even without resembling the target. Combining bottom-up and top-down sources
66 of activation generates an activation map which generates a prediction of the order in
67 which stimuli are processed in visual search. Other extensions to these models have been
68 proposed, such as the Dimension Weighting Account, in which saliency weightings are
69 assigned to different target 'dimensions' (e.g. colour or shape), helping to explain
70 results where varying the target dimension within blocks of trials leads to longer
71 reaction times than where the dimension remains consistent within a block (?). Thus,
72 these models aim to produce a representation of the visual properties of the distractors
73 at each location in the visual field. However, these are predominantly qualitative models,
74 and thus it is difficult to use them to make specific quantitative predictions.

75
76
77 TCS falls under a class of models that take a different approach, in that they
78 focus solely on representing the difference between targets and distractors. For example,
79 in work on eye movement patterns, it has been proposed that performance in inefficient
80 (serial) visual search is mostly determined by the size of the 'func-

81 tional viewing field’, whose size varies as a function of target-distractor similarity
 82 (?). Similarly, work on attention has proposed the notion of ‘relative features’,
 83 where attention is tuned to feature relationships i.e. the appearance of the target
 84 relative to distractors in the environment (??). TCS also has features in common
 85 with other models that propose parallel identification of all items in a scene, with
 86 diffusion based mechanisms for identifying targets from distractors (??). How-
 87 ever, TCS (?) aims to provide a unifying framework that can make quantitative
 88 behavioural predictions for visual search based on this general assumption. As
 89 such, it is an attractive candidate model for a formal registered replication.

90 A key assumption of the TCS model is that behaviour is determined by com-
 91 paring the target template (held in memory) with every element present in the
 92 scene in parallel. This allows the visual system to reject peripheral non-targets
 93 quickly; the speed at which items are evaluated is determined by how different the
 94 item is from the template through an evidence accumulation process (formally,
 95 the slope of the logarithmic function is assumed to be inversely proportional to
 96 the overall magnitude of the contrast signal between the target and distractor).
 97 The model thus focuses on an initial, efficient processing stage of search; if suf-
 98 ficient evidence is not accumulated during this process, the model posits that a
 99 second stage is entered, requiring a sequence of eye movements to search for the
 100 target in a serial manner. TCS has been successful in predicting a number of em-
 101 pirical results, including search performance in heterogeneous scenes based on
 102 parameters estimated in homogeneous scenes, both with artificial stimuli (??) and
 103 with real-world objects visualised on a computer display (?). Table 1 provides an
 104 overview of studies investigating the TCS framework to date.

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

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

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

113 1.2. Rationale for proposed work

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

Reference	Overview
?	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.
?	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.
?	Logarithmic efficiency in efficient search cannot be explained by crowding in peripheral vision.
?	Logarithmic efficiency in efficient search cannot be explained by eye movements.
?	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.
?	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.
?	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.
?	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.
?	Extension of ? 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.

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 (?). However, when considering each par-
127 ticipant individually, it became clear that there was a high level of heterogeneity
128 in responses, with some participants performing close to optimally, and others ac-
129 tually performing worse than chance (?). Similarly striking variability has also
130 been reported in other search studies (??).

131 Taking search time distributions into account is also important for constrain-
132 ing theories of visual search (??): for example, they have been used to help dis-
133 tinguish between models that make similar predictions at the level of average re-
134 action times (??). Including subject and trial level data into our implementation
135 of the TCS will therefore further aid model development and assumption testing.

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

150 In the current manuscript, we focus on replicating and extending findings from
151 ?. In their study, participants searched for a target in a scene of homogeneous
152 distractors (see Figure 1). First, parallel search efficiency (measured by the log-
153 arithmic search slope) was estimated for cases where the distractors varied from

154 the target in one dimension: either colour (e.g. a cyan target being searched for
 155 in either yellow, blue or orange distractors) or shape (e.g. a semicircle target in
 156 either circle, diamond or triangle distractors). New participants then searched for
 157 the same targets in displays where the distractors were compounds, differing from
 158 the target in both colour and shape (e.g. searching for a cyan semicircle in either
 159 blue circles, orange diamonds or yellow triangles). The logarithmic search slopes
 160 in the initial experiments were then used to predict the logarithmic slopes and re-
 161 action times using a number of models. The authors found that the best model
 162 was a ‘collinear contrast integration model’ where the distinctiveness scores were
 163 summed along each attribute in the unidimensional experiments, creating an over-
 164 all contrast score that was used for compound stimuli predictions. In our regis-
 165 tered replication, we will attempt to verify the conclusions of ?, that the collinear
 166 contrast integration model does indeed offer the best characterisation of contrast
 167 signal combinations in visual search within the TCS framework.

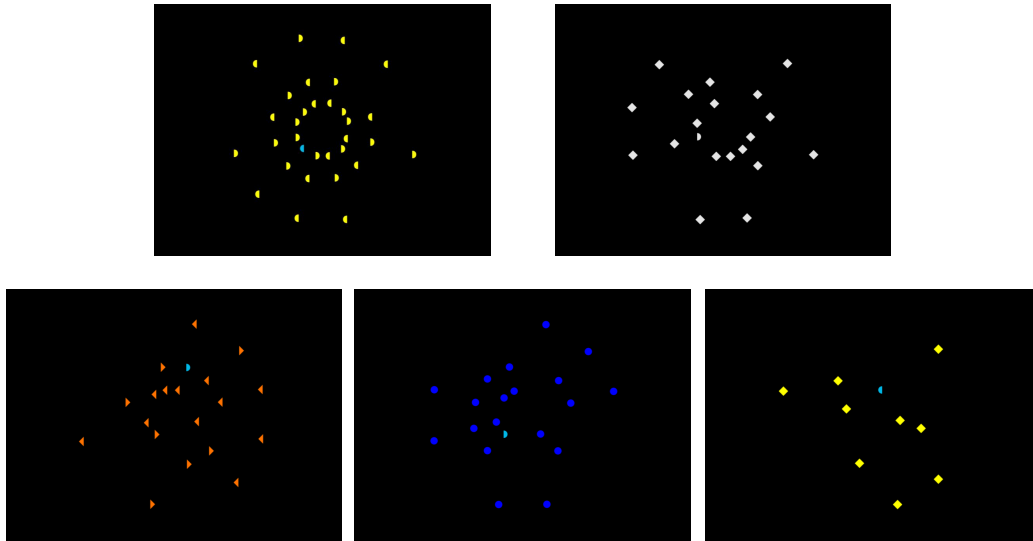


Figure 1: Example stimuli from ?. Top left: Expt 1A. Here, the target is a blue semicircle within a set of homogeneous (yellow semicircle) distractors. Top right: Expt 1B. The target is a grey semicircle in circular grey distractors. Bottom left: Expt 2A. The target is a blue semicircle in orange diamond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

168 We begin by verifying the analysis of ?. We then describe our proposed repli-

169 cation study, showing with pilot data how we are able to extend their model of
170 how multi-dimensional contrasts are calculated, both by incorporating a multi-
171 level design to predict within-subjects effects and by utilising a Bayesian gen-
172 eralised linear model framework to better represent the distribution of responses
173 (e.g. avoiding predicting negative reaction times, accounting for uncertainty in
174 model predictions).

175 2. The Target Contrast Model

176 We first describe the original Target Contrast Model, as presented in ? and ver-
177 ify that we can successfully replicate the original analysis (both using frequentist
178 modelling and Bayesian modelling; see *Supplementary Materials*).

179 2.1. TCS modelling overview

180 In Experiment 1a of ?, participants searched for a cyan semicircle target among
181 blue, yellow or orange semicircular distractors i.e. they searched for a target that
182 differed from the distractors by a *single feature* (colour). The experiment was then
183 repeated (1b) using a different single feature (shape, with participants searching
184 for the semicircular target within triangle, circle or diamond distractors). In Ex-
185 periments 2a, 2b and 2c, participants again searched for a cyan semicircle, but
186 this time, the distractors differed in both shape and colour. We will refer to these
187 conditions as *double features*. Note, unlike in standard conjunction searches, in
188 this paradigm, the distractors are all identical with respect to these features (i.e.,
189 orange triangles). Examples of all these stimuli are shown in Figure 1. ? also
190 carried out a replication of their basic results using slightly different target and
191 distractor stimuli (Experiments 3 and 4).

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

197 2.1.1. Calculating the intercept, a , and the logarithmic slope parameter, D_i

198 Experiments 1a and 1b and 3a and 3b were used to calculate the logarithmic
199 slope parameter D_i . In all experiments, the number of distractors varied, allowing
200 the data to be used to fit a log-linear model for reaction times, where reaction
201 times increase logarithmically with N_T , the number of distractors (see Equation
202 1). In the original model the error distribution was assumed to be normal. Thus

203 the results of Experiments 1 and 3 were used to calculate D_i , for each type of
 204 distractor. When colour varied, we will refer to D_c , for $c = 1, 2, 3$. Similarly for
 205 shape we will denote this (D_s), and the compound features are denoted as ($D_{c,s}$).
 206 Fitting the model specified in Equation 1 to the data, we obtain the values for
 207 D_c and D_s given in Table 2. As can be seen, the more similar the distractors are to
 208 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.

209 2.1.2. Estimating $D_{c,s}$, the logarithmic slope parameter for compound features

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

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

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

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

221 Finally, the collinear contrast integration model also assumes independence of
 222 feature dimensions, but assumes that while the visual features create a multidimensional space, the contrast between them is unidimensional. As D is assumed
 223 to be inversely proportional to contrast, the equation can be written as follows:
 224

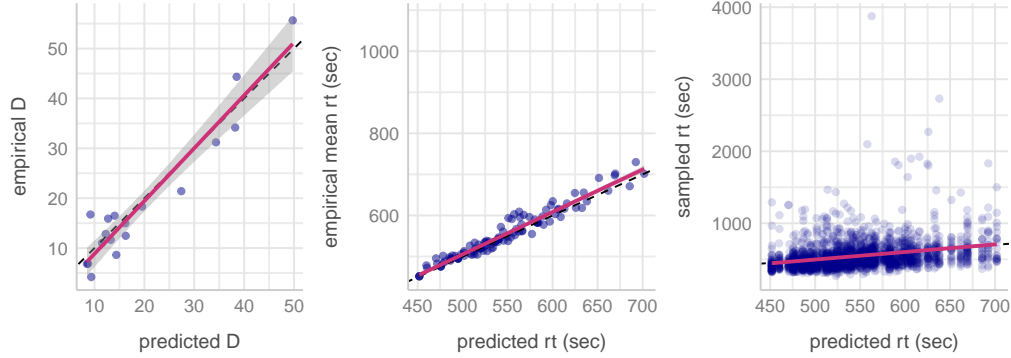


Figure 2: (left) The collinear method for calculating D offers a good prediction. (centre) Using the TCS to predict reaction times. (right) Each dot now represents a randomly sampled reaction time from an observer. Note that there is greater spread in the data points here, due to the fact that there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

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

225 ? found that with their dataset, the collinear contrast integration model was
 226 best able to predict $D_{c,s}$ from D_c and D_s , with $R^2 = 0.915$. We verified we were
 227 able to replicate this result using the dataset available on OSF (<https://osf.io/f3m24/>)¹
 228 and using the exclusion criteria originally applied; see Figure 2 (left panel) and
 229 *Supplementary Materials* for details. We show that we are able to do this using
 230 both the frequentist modelling approaches used in the original paper, and using
 231 Bayesian modelling.

232 2.1.3. Estimating a , the intercept parameter for compound features

233 As a is the intercept of the model, it represents how long observers take to find
 234 a target when $N_T = 0$, i.e., there are no distractors. As such, it should be inde-
 235 pendent of both shape and colour, and can be thought of as the role of non-search
 236 processes (such as motivation, motor preparation etc.) that influence reaction time.
 237 In ?, a was calculated for each sub-experiment. Here, we follow that method in
 238 order to replicate their results exactly.

¹downloaded on 28th August 2020

2.1.4. Estimating mean reaction times

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

2.1.5. 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 adapting TCS to make use of multi-level modelling techniques. Multi-level models allow us to take into account the hierarchical structure of the data (i.e. that each participant completes multiple trials) in a way that does not require averaging, meaning that we are able to model participant variability as well as group-level effects (?).

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. We will also test whether a slightly more complex extension of this model, the shifted lognormal model (which allows the distribution to be offset to the right i.e. mimicking the patterns seen in reaction time data, where valid responses begin at around 100ms) offers any improvement in model fit. Note that a Wald, or inverse Gaussian distribution, would also be a reasonable distribution choice for this data given that TCS is based on a diffusion process e.g. (?), and this distribution has been argued to be psychologically more plausible (e.g. ?, though see ?): we chose not to use this distribution as it often leads to computational issues, which would make it harder for others to reproduce or build on our approach later.

275 3. Hypotheses

276 We plan an experiment to test the extent to which the original results in ?
277 replicate and generalise, using our new modelling approach.

278 3.1. Proposed Modifications to Experimental Design

279 In order to better test the above, and increases sensitivity, we propose to make
280 the following changes to the experiment described in ?:

- 281 1. **Within-subjects design.** This modification should give us greater power to
282 detect differences between different models, as well as allowing us to in-
283 vestigate how individual differences in the single-feature task might explain
284 differences in the double-feature task.
- 285 2. **Increase target-distractor similarity.** If the distractors are a very differ-
286 ent colour from the target, they may not distinguish well between different
287 contrast models. We will therefore run a version of the experiment where
288 the target is a red semicircle, with distractors being either orange, purple or
289 pink.
290
291

292 3.2. Registered Hypothesis

- 293 1. **Shifted lognormal model.** We hypothesise that a shifted lognormal model
294 will give the best fit to our single-feature data, when compared to a lognor-
295 mal and a normal model.
296
- 297 2. **Log-linear effect of N_T .** We will test the TCS model assumption that N_T
298 has a log-linear effect by testing models with and without the log of this
299 term. We expect that this will confirm the results previously seen in papers
300 testing TCS i.e. that the log-linear approach will be best.
301
- 302 3. **Contrast model comparisons.** We will test the hypothesis proposed by
303 (?): specifically, that the *collinear contrast integration model* outperforms
304 the *best feature guidance*, and *orthogonal contrast combination models* for
305 the calculation of D , by calculating and comparing the mean absolute pre-
306 diction error for each model.
307

308 **4. Reaction time predictions.** We will further test the hypothesis proposed by
309 (?) by testing which model gives the best prediction at the trial-by-trial RT
310 level.

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

321 3.3. *Planned Explorations*

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

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

337 3.4. *Pilot Experiment*

338 Full details of a pilot experiment with $n = 4$ participants (960 trials each) using
339 our proposed analyses can be found in *Supplementary Materials*. This suggests
340 that even with a small sample, we can convincingly demonstrate H1 and H2. How-
341 ever, more data will be required to discriminate between the models, particularly

for H4. Given that our methods are within-subject, we have reduced the number of trials per condition compared to ? (12 in our proposed study, compared to 40 in theirs). It is therefore possible that the increased noise in our estimated D single-feature parameters will make it more difficult to predict double-feature Ds accurately. However, we think this is unlikely to be the case for several reasons. Firstly, we can see that even in a small amount of pilot data, we can verify H3, with the collinear model having the lowest mean absolute prediction error. Secondly, while the pilot data does not give a strong weighting to the collinear model in H4, it is still the most probable model. Finally, we have carried out simulations to estimate our measurement error....

4. General Methods

4.1. Sample Size: Participants and Trials

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

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 36 different target positions in total, in three 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 ?, 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 (purple, 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 ? (1, 4, 9, 19 and 31). We will also run target-only 'zero distractor' trials (60 in total, with 12 being the white semicircle target and the remainder the red semicircle target).

377 The experiments were programmed in PsychoPy and Pavlovia (?). Stimuli
378 were pre-made to generate search array images with 1920×1080 resolution.

379 4.3. Procedure

380 Participants will complete the experiment in the laboratory, sitting at a view-
381 ing distance of 45cm from the screen (viewing distance will be fixed by using a
382 chin rest). They will view a fixation cross before viewing a search array: they
383 will press the space bar to continue to the trial. Participants will be told to search
384 for the target among distractors (either a red semicircle or a white semicircle, de-
385 pending on the block) and report if the semicircle target points to the left or right,
386 by pressing either the left or right button on a button box (Cedrus RB-540). They
387 will first complete 16 practice trials where they will receive feedback immedi-
388 ately after completing each trial. In the real experimental trials, participants will
389 receive feedback on their average accuracy and reaction time after each block of
390 120 trials. Participants will complete 8 blocks of trials (960 trials overall i.e. 192
391 trials in each of 5 experiments, consisting of 5 set sizes x 3 distractor conditions x
392 12 repeats + 12 zero distractor trials). The trials where the distractors are single-
393 feature shapes (i.e. the target is a white semicircle - Experiment 1b in ?) will all
394 appear in one block (which will appear at a randomly selected position within the
395 experiment). All other trials (where the target is red semicircle) will be fully ran-
396 domised i.e. all different conditions will be completely intermixed. This approach
397 will be taken as TCS requires the participant to have a well-defined target template
398 in mind in order to compare this to the stimuli in the display. Thus, participants
399 will be cued to search for the relevant target at the beginning of each block.

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

402 4.4. Data Pre-processing

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

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

410 *4.5. Analysis Plan*

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

414 Please see the analysis of our pilot data for a full implementation of our anal-
415 ysis pipeline, including all code (available on Github at https://github.com/scienceanna/TCS_Bayesian).
416

417 **5. Results**

418 *– blank –*

419 **6. General Discussion**

420 *– blank –*

421 **Conflict of interest**

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

423 **Acknowledgements**

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

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