JP's Comments 12/12/24 12/16/24

Negative effects of redundant targets

Alex L. White*[1], John Palmer[2], Genevieve Sanders[1], Jannat Hossain[1] & Zelda B. Zabinsky[3]

[1] Department of Neuroscience & Behavior, Barnard College, Columbia University.

[2] Department of Psychology, University of Washington

[3] Department of Industrial & Systems Engineering, University of Washington

* Corresponding author: Alex L. White, alwhite@barnard.edu

76 Claremont Ave, New York, NY 10031

[redacted] stimuli are processed in parallel. This question is particularly important for situations that involve visual search (e.g., locating a friend in a crowd) and for complex tasks such as reading, when many relevant stimuli (e.g., words) are presented simultaneously. Depending on several sensory and cognitive factors, multiple stimuli might be identified simultaneously just as well as a single stimulus can be, or with hindered performance due to a processing capacity limit. There are gradations of limited-capacity parallel processing, due to finite cognitive resources being shared between stimuli that might compete or interfere with each other. At the extreme, multiple stimuli might be processed serially; for instance, if there is a central bottleneck in the brain for making a certain type of judgement.

One of the most straightforward tools for evaluating parallel processing capacity is the redundant target paradigm (e.g., van der Heijden et al., 1983). In this paper, we revisit this paradigm and develop generalizations of three classes of models that predict task performance: standard unlimited-capacity parallel, fixed-capacity parallel, and serial models.

We then test these predictions with three experiments that use written words as the target stimuli. These experiments assess how well two words can be recognized simultaneously. This is an important question because competing models of natural reading disagree as to whether multiple words are identified in parallel during each gaze fixation (Engbert et al., 2005; Reichle et al., 2006; Reilly & Radach, 2006; Snell & Grainger, 2019b). The redundant target experiments presented here differ from natural reading, but they characterize the fundamental capacity limits that readers cope with. Specifically, they assess how well participants can process two words at once when they are encouraged to try. The redundant target paradigm complements other approaches that use dual-tasks or measure spatial attention effects to investigate word recognition (Johnson et al., 2022; White et al., 2018, 2020; White, Palmer, et al., 2019).

Fundamentals of redundant target effects

The redundant target paradigm grew out of a larger visual search literature to investigate whether observers can process multiple stimuli presented simultaneously at different visual field positions (van der Heijden, 1975). The observer's task is to view a display and report the presence or absence of stimuli that belong to a target category. Non-target stimuli are termed "distractors." On some trials, one target is presented. On other trials, multiple targets are presented simultaneously – that is, the display contains redundant targets. The typical finding is a positive *redundant target effect*: a speeding of correct response times for multiple targets compared a single target. Such a "redundancy gain" is often taken as evidence that the targets were identified in parallel.

Studies that have used the redundant target paradigm can be divided into two broad categories. Studies in the first category seek to distinguish between a parallel model and a serial model (van der Heijden, 1975). They compare response times between displays that consist of two (or more) targets, versus displays that contain one target and no other stimuli. Positive redundant target effects have often been used to reject the serial model in favor of the parallel model. (Exceptions in the literature are reviewed below). Studies in the second category seek to distinguish between two flavors of parallel models: those with separate activations caused by each stimulus, versus those with interactive "coactivations" (C. W. Eriksen et al., 1989; J Miller, 1982; Mordkoff & Yantis, 1991). To do so, response times are compared between displays that contain two targets and displays that contain one target and one distractor. These "mixed" trials are not useful for testing the serial model, which is the focus of the present study.

Therefore, we focus on the comparison between displays with a single target presented alone and displays with two targets. [redacted]

[redacted]. As shown in **Figure 1**, contrasting predictions for correct response time arise from a standard unlimited-capacity parallel model and a standard serial model. They are called "standard" models because of strong assumptions about the independence of the processes for each stimulus. Both standard models assume that search is *self-terminating*: the observer responds as soon as they detect a target (Van Zandt & Townsend, 1993). In the General Discussion ~~section~~ we address how parallel and serial models can mimic each other when made more complicated (Algom et al., 2015). In this study, however, we focus on three relatively simple models with plausible assumptions, which we develop further by accounting for errors as well as response times. To make theoretical progress we then reject specific models that cannot account for performance in word recognition tasks.

The standard, self-terminating parallel model assumes that when two targets are present, they are independently processed in separate channels that race to produce the response. The completion time of each process is variable across trials. On two-target trials, the response time is determined by the faster of the two processes, so the observer is faster on average than when only one target is present (Raab, 1962). Thus, the standard, self-terminating parallel model predicts a positive redundant target effect: a speeding of correct responses.

The standard self-terminating serial model, in contrast, assumes that one stimulus is identified at a time (Townsend & Nozawa, 1995; van der Heijden, 1975). If the first target of two simultaneously presented targets is correctly identified, then the response time is on average the same as when only one target is present—the redundant target has no effect on performance. Previous descriptions of this serial model stop there; but as we explain in our new theory section below, a serial model that incorporates errors predicts a *slowing* of correct response times if the first target to be processed is misidentified, and search continues to process the second target correctly.

slightly different, such as requiring a choice response on each trial (Fournier & Eriksen, 1990; Grice & Reed, 1992; van der Heijden et al., 1983). One possibility is that letters are processed in parallel, producing a positive redundant target effect, but that effect can be masked by later decision- or response-selection processes when the response rule is more complicated. In other words, if conditions increase the degree of limited capacity, then it can reduce the redundant target effect.

[redacted]

We now turn the test case explored in the experiments below: redundant target effects for written words. Such effects can reveal the extent to which higher-level semantic or linguistic information about two stimuli can be processed in parallel. Of the handful of studies that have taken this approach, most – but not all – conclude that two words can be recognized in parallel. These studies have differed in four important respects: whether the redundant targets in a single trial are identical words; whether the single-target trials also contain a 'filler' stimulus; what the task is (semantic categorization vs. lexical decision), and how the subject responds (go/no-go vs. choice).

We first summarize experiments that reported a positive redundant target effect for word recognition. Shepherdson and Miller (2014) used a semantic categorization task. Words were presented to the left and/or right of fixation in some experiments and above and/or below fixation in others. On each trial, the subject had to make a yes/no response to report the presence of a word belonging to a target semantic category (animals). Importantly, the "single-target" trials also contained a "filler" stimulus that was a pronounceable pseudo-word. The key result was that responses were faster in the redundant target condition compared to this modified target-filler baseline.

Interpreting this experiment is difficult. The critical question is what the standard self-terminating serial model predicts in terms of the RT difference between these two conditions. It predicts 0 difference only if we make an additional assumption about the "target-filler" trials (a single target paired with a pseudoword filler): the target stimulus is always processed first and the pseudoword is never processed at all (at least on correct trials). That seems unlikely, as the observer would have to first determine which stimulus is a real word before beginning to process it. Alternatively, if we assume that the serial process sometimes begins with the filler stimulus before moving on to the target word, then the serial model predicts slower responses on these target-filler trials than on two-target trials. The serial model thus predicts the same thing as most parallel models (a positive redundant target effect). Thus, we conclude that this experiment by Shepherdson & Miller (2014) does not strictly test the standard self-terminating serial model that we wish to test. Instead it is part of the broader set of experiments that use a mixed-pair baseline to test co-activation models.

We must go further back in time to find studies that did test the standard serial model for word recognition by comparing displays containing two words to display containing one word and nothing else. In the four experiments reported by Mullin and Egeth (1989), words were presented above and/or below fixation, and the participant's instruction was to make a go/no-go response to the presence of any target. Trials contained 1 distractor alone, 1 target alone, 2 distractors alone, or 2 targets alone. In two experiments with a semantic categorization task, the redundant target effect did not significantly differ from 0. The results varied surprisingly in two experiments that used a lexical decision task: targets were real English words and distractors were pseudowords. When the redundant targets were two identical copies of the same word, there was a significantly positive redundant target, consistent with parallel processing. A similar result was reported by Egeth et al., (1989). Other studies found consistent positive effects for lexical decision with identical words present to the left and right of fixation: Hasbrooke and Chiarello (1998) and Mohr and colleagues (Mohr et al., 1994, 1996).

Our goal in building this new theory is to compare the *qualitative* predictions of the various models: whether they predict positive, negative, or zero redundant target effects on response time and accuracy. Our goal is not to quantitatively fit models to our data; that is a different endeavor that we leave for the future (see also Cox & Criss, 2019). For now, it is sufficient to generate qualitative predictions that allow experimental data to rule out some models.

Figure 2 illustrates the typical range of predicted redundant target effects for each class of model that incorporates errors. The standard serial model ~~always~~ predicts negative effects. The standard unlimited-capacity parallel model ~~always~~ predicts positive effects. The standard fixed-capacity parallel model can predict either negative or positive effects. These results of the new theory are described in more detail in the following paragraphs.

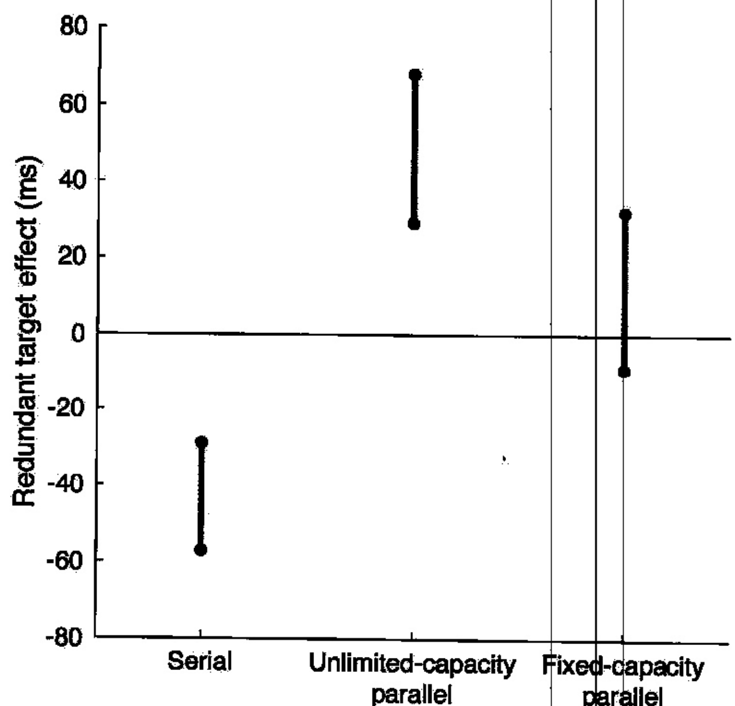

Figure 2: Model predictions. For each type of model, we plot the typical range of redundant target effects on correct response times. For all models we assume that for single-target trials, errors occur on 5% of trials and the mean correct response time is 800 ms. Each model's range encompasses the smallest and largest effects that we could generate with the parameters described at the end of the Appendix.

always positive: it is the difference between two products, and the first is always larger. This must be the case, as is clear when examining each part of the two products separately. First:

$$\left(\frac{1}{2-p_t}\right) \geq \left(\frac{p_t}{2-p_t}\right)$$

because $0 \leq p_t \leq 1$. Second:

$$E\left[\boldsymbol{D}_{t,correct}\right] > E\left[\min\left\{\boldsymbol{D}_{t_1,correct}, \boldsymbol{D}_{t_2,correct}\right\}\right]$$

because the mean difference between two identically distributed (non-negative) variables is always less than the mean of one of those variables alone.

An illustration of this prediction is the middle bar in Figure 2. For this illustration, we used predictions of two specific models described in the Appendix. The upper point is for a diffusion model with parameters that generate large redundant target effects. The lower point is for a linear ballistic accumulator model with parameters that generate relatively small redundant target effects. While not strict limits, these model outputs illustrate the range of predictions from the standard, self-terminating, unlimited-capacity parallel model. They are always positive.

Predictions of Our Fixed-Capacity, Parallel Models

The parallel model that can most mimic a serial model is one that has limited capacity. The limited capacity slows processing when there are two stimuli and thus reduces and possibly eliminates the redundant target effect. Unfortunately, this model is so general that it does not make very specific predictions. Here, we consider a special case of the limited-capacity parallel model: the fixed-capacity, parallel model. The idea of fixed capacity is that a set of parallel processors extract the same total amount of information from multiple stimuli as they do from a single stimulus. ~~Thus~~ In other words, splitting a fixed set of 'resources' between multiple stimuli introduces a cost. Most of the prior work

Predictions about RT distributions

These models can also make predictions about the distributions of response times, including their overall variance. Parallel race models predict smaller RT variance for two-target than one-target trials, whereas the generalized standard serial model predicts the opposite. This serial model also assumes that correct RTs on two-target trials are a mixture of trials when one target is identified and processing stops, and other trials when one target is misidentified and the response is made only after the second target is correctly identified. Thus, the serial model also predicts a particular shape of the RT distributions.

These model predictions are fruitful ground for future investigations, but were not explored further in the present article, primarily because the experiments below were not designed to test them. They were conducted online with relatively large sample sizes and relatively few trials per participant, which prevents detailed analyses of RT distributions.

Predictions About Errors

Our last result concerns the usefulness of models that incorporate errors. All of the models described above predict a positive redundant target effect on accuracy (that is, fewer errors on trials with 2 targets than on trials with 1 target). This is not a surprise for typical parallel models that have been investigated in summation experiments of accuracy alone (Graham et al., 1978). What is new is that this result also occurs for our serial model, even though that model predicts slower response times for two targets. The reason is that when two targets are present and processed sequentially, there are two chances to correctly detect target presence, so accuracy increases compared to when only 1 target is present – even though doing so takes more time on average. While this result for errors does not distinguish between the serial and parallel models, it introduces a result that is specific to errors and that is not accounted for by pure response time models.

published case of a negative redundant target effect was in a lexical decision task with non-identical redundant targets (Mullin & Egeth, 1989).

[redacted] the design of the prior study that provides the strongest foundation for the present investigation (Mullin & Egeth, 1989). This clearly differs from the standard format of reading English from left to right, but it allows both words to be close to fixation and easily legible. We also sought to avoid a strong difference in performance across the two locations, which is known to be significant for lateralized presentations of words left and right of fixation (reviewed by Yeatman & White, 2021).

In each experiment, we also measured performance in three different tasks (color detection, lexical decision, and semantic categorization). The *color task* required participants to judge a low-level visual feature of the words, and served as a positive control condition for which we expected positive redundant target effects. [redacted] find a positive effect in this task could indicate that in these displays there are sensory interactions between two words presented at once, or that participant cannot attend to both locations equally, both of which could mask a true redundancy gain. As shown below, those concerns are minimized by the positive result in the color task. The *lexical decision task* requires the subject to distinguish real English word targets from pseudoword distractors. The *semantic categorization task* requires categorizing words either as targets that belong to a category of "living things" and distractors that belong to a category of "non-living things". The semantic and lexical tasks might tap into different levels of linguistic processing, and have both been used in prior redundant target studies (Egeth et al., 1989; Mullin & Egeth, 1989). In sum, within each of our three experiments, we carry out a side-by-side comparison of redundant target effects that arise in three tasks using the same stimuli. The tasks differ in which they require low-level color feature detection, lexical access, or semantic categorization. Altogether, this study includes over 184,000 trials of data from a total of 257 participants.

[I suggest moving this point into methods] overlapping, with means 19.7 and 14.5, respectively. Each word was repeated on average 3.7 times within the experiment.

Analysis: We computed two measures of performance in each condition: the mean response time on correct trials, and the percent of trials with incorrect responses (errors). In most cases we focus on trials with targets, because only those measures test our models that assume self-terminating search for targets. For both measures, we compared the means on trials with two targets to trials with one target with paired t-tests. All t-test p-values were corrected for false discovery rate across the 9 tests done for each measure in the entire study (Benjamini & Hochberg, 1995). We also used bootstrapping to get a 95% confidence intervals (CI) of each mean difference. Lastly, we supplement our pairwise tests with Bayes factors (BFs), which quantify the strength of evidence (Rouder et al., 2009). The BF is the ratio of the probability of the data under the alternate hypothesis (that two means differ) relative to the probability of the data under the null hypothesis (that there is no difference). A BF of 10 would indicate that the data are 10 times more likely under the alternate hypothesis than the null. BFs between 3 and 10 are regarded as substantial evidence for the alternate hypothesis, and BFs greater than 10 as strong evidence. Conversely, BFs between 1/3 and 1/10 are considered substantial evidence for the null hypothesis, etc. We computed BFs using the bayesFactor toolbox by Bart Krekelberg (https://doi.org/ 10.5281/zenodo.4394422).

insert from below

Estimates of the redundant target effect are complicated by the possible differences the judgment of the two single-target conditions (Mullin, Egeth & Mordkoff, 1988). Such differences are the rule in multisensory (e.g. auditory-visual) experiments, but even occur with two visual stimuli at isoeccentric locations. For example, in the current experiments, single words at the top location were judged more accurately and quickly (see results below).

Without such a difference between locations, the redundant target effect can be simply estimated by taking the difference between the two-target condition and the average of the one-target conditions (the "averaging baseline" approach). But with a known difference between locations, a better estimate might be the difference between the two-target condition and one-target condition from just the location that produces *faster* responses in the mean. This approach, with the fastest-location baseline, reduces the size of the estimated redundant target effect. Conversely, using the averaging baseline can overestimate the size of the redundant target effect. Many authors use the fastest-location baseline as being more conservative for the purpose of estimating the positive redundant target effects predicted by many parallel models. Some authors use a yet more conservative baseline of the average of the fastest response for each subject separately (Miller & Lopes, 1988).

The goal of the current study is to investigate possible negative redundant target effects. In this context, the fastest baseline produces estimates that are biased to be more negative than the averaging baseline. Thus, for detecting negative effects, the averaging baseline is more conservative. We use this estimate throughout this article.

To compare across tasks across experiments, we also fit linear mixed effect (LME) models to single-trial data, with fixed effects of the task, the set size (number of words), random intercepts and slopes by participant, and random effects for individual stimulus items. All p-values for a certain test were corrected for false discovery rate across the 9 tests done in the study.

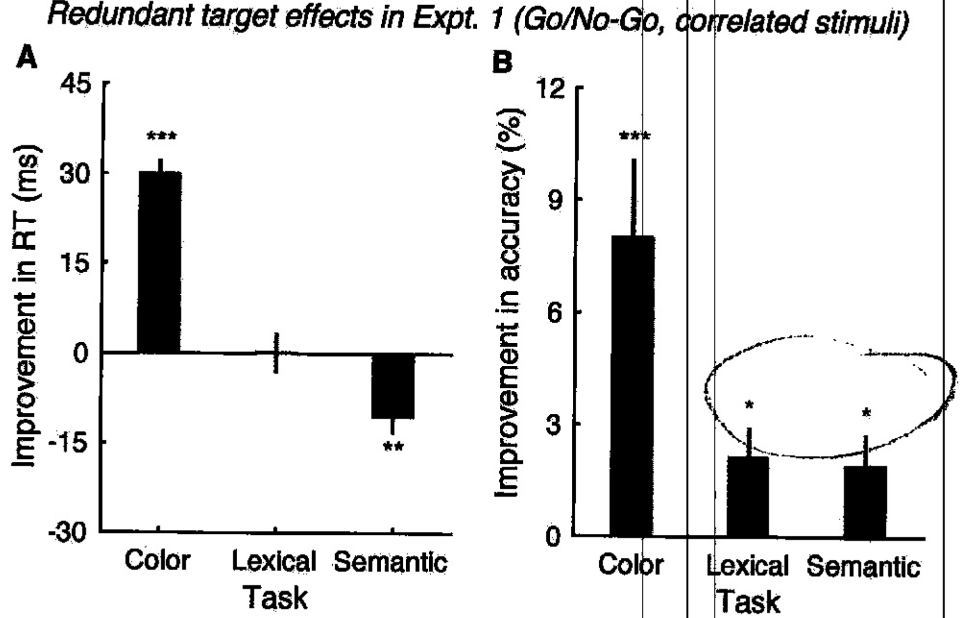

Figure 4: Redundant target effects in Experiment 1. These bar plots show the mean *improvement* in **(A)** mean correct response time (RT) and in **(B)** accuracy for 2 targets compared to 1 target. The mean performance levels from which these difference scores were derived are in Figure 5. Error bars are ± 1 SEM. Asterisks indicate that the mean effect is significantly different from 0 (***p<0.001, **p<0.01, FDR-corrected).

Results

<u>Response times</u>: **Figure 4A** shows that in the color task, there was a positive redundant target effect: a speeding of correct responses to two targets compared to one, by 32 ms on average. However, in the semantic task, there was a significantly negative effect (a slowing of responses) of 11 ms on average. In the lexical decision task, there was no effect of redundant targets. **Table 3** lists the statistics on each effect. The mean response times in each individual condition (rather than the differences between one and two targets) are shown in the top row of **Figure 5.**

To compare the redundant target effects across tasks, we also fit three linear mixed-effect models to single-trial correct response times (target-present trials only). Compared to the color task, both the lexical and semantic tasks had significantly different redundant target effects (both F>36, $p<10^{-8}$). The lexical and semantic tasks yielded effects that were marginally different ($F(1, 15475)=3.86$, $p=0.056$).

Task	Effect mean	Effect SEM	95% CI	*t*	*p*	BF
	Correct response time (ms)					
Color	31.98	2.49	[27 36]	12.67	6.35×10^{-12}	1.09×10^{10}
Lexical	-0.11	3.53	[-6 8]	0.03	0.98	0.20
Semantic	-10.67	3.53	[-17 -6]	3.60	0.002	27.69
	Errors (percent)					
Color	7.95	2.10	[4.54 13.23]	3.72	0.001	35.778
Lexical	2.12	0.80	[0.62 4.02]	2.59	0.017	3.22
Semantic	1.90	0.86	[0.49 4.12]	2.19	0.038	1.54

Table 3: Statistics on redundant target effects in Experiment 1, describing the mean *improvement* in response times and error rate, contrasting 1-target displays ([redacted] vs. 2-target displays. [redacted] The degrees of freedom for the t-tests was 27. For each measure (response time or accuracy), p-values are corrected for false discovery rate across all 9 comparisons including all 3 experiments in the study. BF = Bayes Factor.

Differences between top and bottom sides: on single-target trials, participants generally made faster correct responses and fewer errors for targets at the top location compared to the bottom location. Correct RTs were on average 19, 32, and 15 ms faster for the top location in the color, lexical, and semantic tasks, respectively (SEMs = 5 ms, all $p<0.01$). Error rates were on average 5, 8, and 2% lower at the top location in those three tasks, respectively (SEMs = 4, 2, and 1%, only significant in the lexical task, $p=0.0003$).

[redacted] redundant target effect would be the difference between trials with two targets and trials with one target at the top location only. This would shift estimates of the mean redundant target effect down, rendering even the lexical task's effect significantly negative (mean = -15 ms, p=0.001) while maintaining a significantly positive effect in the color task (mean = 24 ms, $p=2\times10^{-7}$). Given that we are most interested in detecting the negative effects predicted by our generalized serial model, in the primary analyses above (Table 3) we significantly different ($F(1, 15141)=2.61$, $p=0.11$). See the top row of **Figure 7** for mean correct response times in each condition separately.

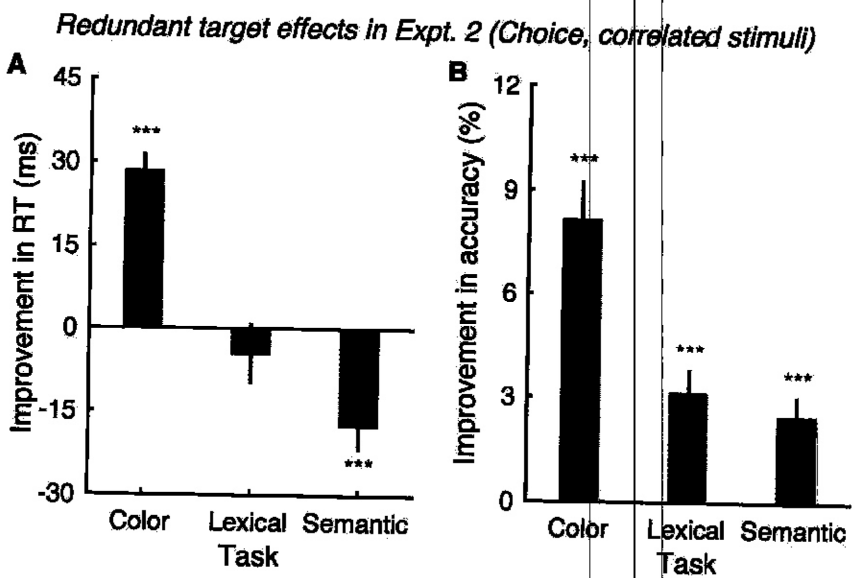

Figure 6: Mean redundant target effects in Experiment 2. Format as in Figure 4.

Accuracy. [redacted] **Figure 6B** plots the mean improvements in accuracy caused by redundant targets, which were significant in all three tasks (as also reported in Table 4). The mean percent errors (misses) in each condition are plotted in the bottom row of **Figure 7**. Compared to the color task, both the lexical and semantic tasks had smaller redundant target effects on accuracy (both $F>21$, $p<10^{-5}$). The redundant target effects did not differ significantly between the lexical and semantic tasks ($F(1, 16689)=0.05$, $p=0.82$).

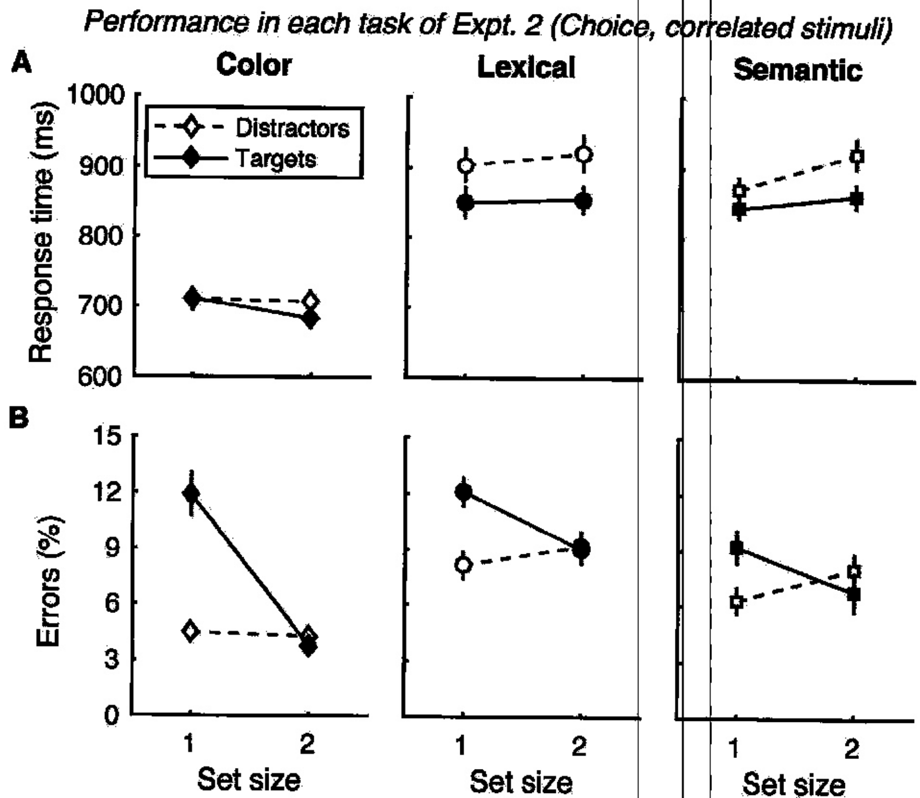

Figure 7. Mean performance in each task of Experiment 2. Format as in Figure 4.

Task	Effect mean	Effect SEM	95% CI	t	p	BF
	Correct response times (ms)					
Color	28.29	3.26	[22 35]	8.52	1.3×10^{-08}	4.1×10^{6}
Lexical	-4.51	5.51	[-15 6]	0.80	0.48	0.27
Semantic	-17.50	4.38	[-26 -10]	-3.92	0.001	57.8
	Errors (percent)					
Color	8.14	1.14	[6.04 10.85]	7.02	1.1×10^{-6}	1.2×10^{5}
Lexical	3.10	0.71	[1.71 4.70]	4.30	0.001	142.9
Semantic	2.45	0.60	[1.26 3.55]	4.03	0.001	74.8

Table 4: Statistics on redundant target effects in Experiment 2, formatted as in Table 3. The degrees of freedom was 28 for the color task and 27 for the others.

[redacted] Correct RTs were on average 37, 37, and 32 ms faster for single targets the top location in the color, lexical, and semantic tasks, respectively (SEMs = 7, 10 and 6 ms, all $p<0.01$). Error rates were 8, 7, and 3% lower for

Remove underline

		Bottom word		
		None	Distractor	Target
Top Word	None	N/A	0.05	0.15
	Distractor	0.05	0.15	0.15
	Target	0.15	0.15	0.15

Table 5: The probability of stimulus pairings at the top and bottom locations in Experiment 3. The word at each location was either absent, a distractor, or a target. The green shading highlights conditions when 2 words were present. In this design, the two stimuli were uncorrelated and independent: the conditional probability of one stimulus being a target given that the other was a target was 0.5.

Analysis: We analyzed these data in the same way as Experiment 1 and 2, focusing on the comparison between trials with two targets and trials with a single target presented alone, which provide the best test of our self-terminating models of parallel or serial processing. The mean percentages of trials excluded for sluggish response times (>4 SDs above each participant's mean) in the color, lexical and semantic tasks were 0.48%, 0.56%, and 0.61%, respectively.

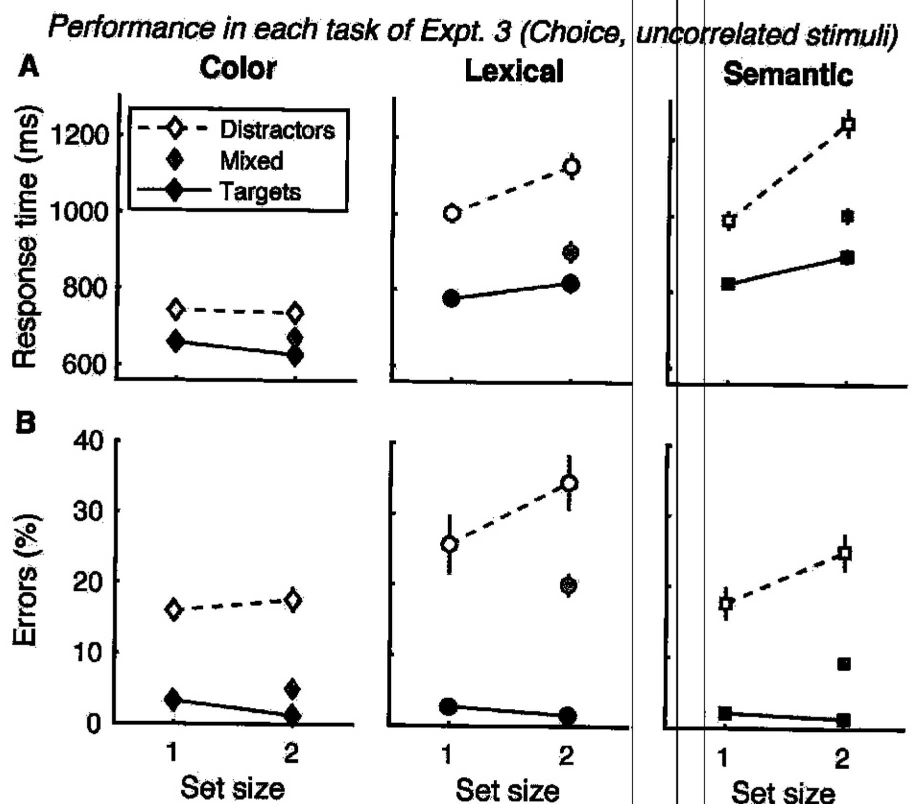

Figure 9. Mean performance in each task of Experiment 3. Format as in Figures 5 and 7, except that here we add the 'mixed' trials that contained 1 target and 1 distractor. That mixed-pair condition is represented by the single diamond with medium fill in each panel.

Accuracy: [redacted] **Figure 8B** plots the positive effects of redundant targets on accuracy, which were significant in all three tasks (see Table 6). None of the pairwise comparisons of these effects across tasks were significant after correcting for multiple comparisons (color vs. lexical: F(25111)=3.45, p=0.095; color vs. semantic: F(25105)=4.10, p=0.077; lexical vs semantic: F(25114)=0.10, p=0.821). The mean percentage errors in each condition are shown in the bottom row of **Figure 9.** [redacted]

Mention Table 6 for accuracy
— explain this table for misses

[redacted]

Task	Effect mean	Effect SEM	95% CI	t	p	BF
	Correct response time (ms)					
Color	33.07	9.73	[21 72]	3.34	0.0032	15.2
Lexical	-41.53	9.67	[-71 -28]	4.22	0.0006	116.5
Semantic	-73.90	8.74	[-93 -58]	8.30	2×10^{-8}	1.9×10^{6}
	Errors (percent)					
Color	2.06	0.38	[1.34 2.76]	5.35	5.3×10^{-5}	1.8×10^{3}
Lexical	1.11	0.27	[0.62 1.58]	4.08	0.001	83.5
Semantic	0.77	0.26	[0.26 1.34]	2.87	0.010	5.6

Table 6: Statistics on redundant target effects in Experiment 3, formatted as in Table 3. This table only reports performance on trials with 2 targets as compared to trials with 1 targets (trials with distractors are not included here). The degrees of freedom for the t-tests was 27.

[redacted]: Correct RTs were on average 15, 41, and 24 ms faster for single targets at the top than bottom location in the color, lexical, and semantic tasks, respectively (SEMs = 6, 15 and 6 ms, all $p<0.02$). Error rates were on average 2.7, 0.3, and 0.9% lower for single targets at the top location in those three tasks, respectively (SEMs = 0.5, 0.4, and 0.3%, significant in the color and semantic tasks, $p<0.01$).

Discussion

The results of Experiment 3 were consistent with both prior experiments: there is a positive redundant target effect on response times only in the color task. One new result in this experiment was that the lexical task, as well as the semantic task, yielded a significantly *negative* redundant target effect (slowing of responses). These results again reject the serial model for the color task only. The serial model can account for lexical and

These data are all consistent with the hypothesis that the low-level features, such as color, of multiple stimuli are processed in parallel, but [redacted]. The serial model will always predict the negative redundant target effect for semantic judgements, as long as there are some errors in classifying targets. Some fixed-capacity parallel models with additional specific assumptions built into them can also account for the negative effect. The lexical decision and semantic categorization tasks were two ways to assess word recognition: the lexical decision task requires the participant to judge the familiarity of each letter string, and the semantic categorization task further requires further judgment of the word's meaning.

Differences in difficulty across tasks

Another clear result in all three experiments was that RTs and error rates were overall lowest in the color task. Could that explain why only the color task yielded a positive redundant target effect, while the semantic task yielded a negative effect? We think this is unlikely, for two reasons. First, the individual participants in the color task who had the slowest RTs or highest error rates (in line with the means in the semantic task) still had positive redundant target effects on RTs. Similarly, the best-performing subjects in the semantic task all had negative redundant target effects on RTs (except for one participant with perfect accuracy in Experiment 1). Second, we separately analyzed low- and high-frequency words in the semantic task. Responses to higher-frequency words were faster and more accurate overall, but still produced a negative redundant target effect in all three experiments. Thus, the qualitatively different redundant target effects across the different tasks are not explained by the color judgment being somewhat easier on its own.

In any case, the literature is dominated by reports of positive redundant target effects (Townsend, 1990; van der Heijden et al., 1983). In this study the color task serves as a positive control condition to demonstrate the expected redundancy gain for a simple visual feature task. What we highlight is the existence of a negative redundant target effect with the same stimuli in the semantic task. This is quite rare and can be explained by our updated models which incorporate errors.

[Insert new section on sensory interactions.]

Relation to previous redundant target studies of word recognition

In contrast to our key result, some prior studies have reported positive redundant target effects for word recognition. However, several of those experiments presented two copies of the same word on redundant target trials (Egeth et al., 1989; Hasbrooke & Chiarello, 1998; Mohr et al., 1994, 1996; Mullin & Egeth, 1989, Experiment 1-2). The resulting redundant target effects might be explained by facilitation or co-activation at the stage of letter or syllable processing, for example, rather than actual semantic recognition. The only study to report a positive redundant target effect that did not present identical pairs of words was by Shepherdson & Miller (2014). They found that semantic categorization judgments were faster for two targets than for a single target paired with a pronounceable non-word. This could be interpreted as a positive redundant target effect and evidence for parallel processing. However, the result can be explained by the standard serial model if we assume that on some one-word trials, the participant processes the non-word before they process the target. Therefore, we consider the strict test of our serial model to be the contrast between trials with two targets (which are two different words) and trials with a single target presented alone.

Thus far, all experiments that made this strict test of the serial model with word recognition tasks lead to the same conclusion. They were conducted by us in the present study and by Mullin & Egeth (1989). The experiments in that prior study were like our

Not all studies agree, however (Snell & Grainger, 2019a). Varieties of the "flanker paradigm" (Eriksen & Eriksen, 1974) have demonstrated that judgments of one target word are influenced by the characteristics of nearby, task-irrelevant words (Snell et al., 2017; Snell, Mathôt, et al., 2018; Snell & Grainger, 2018, but see Broadbent & Gathercole, 1990). That is true even when the whole display is flashed for 50 ms and then masked (Snell, 2024). It is important to note that these experiments differ from the redundant target experiments reported above in at least two key ways: (1) the target word was fixated directly and flankers were arranged horizontally to the left and right; (2) only one word was task-relevant. Serial models ~~may~~ might account for these results by serial switching from one stimulus to another, especially when the absence of post-masks leaves unconstrained the time available to process the display. Even if that is controlled, there is the possibility of selection errors: a flanker is processed instead of a target on some trials, which impairs accuracy only when the flanker is incongruent with the target. Such selection errors could produce flanker effects even under a serial model (see discussion in Snell, 2024).

Also relevant is the "sentence superiority effect": when the words that are flashed along with the target form a sentence, the target is reported more accurately than when the word order is scrambled (Snell & Grainger, 2017; Wen et al., 2019). One interpretation of these results is that the words in each display were all processed simultaneously (Snell & Grainger, 2019b). However, similar concerns about rejecting the serial model apply to these data as to the flanker effect studies summarized above.

Thus, prior research on the capacity for processing of multiple words has yielded some inconsistent results. The data reported in the present article add to that prior research in several ways. The redundant target paradigm complements the dual-task paradigm because the words do not have to be post-masked, and the observer needs not make independent judgements about two words simultaneously. Moreover, the effects on response time can be compared to specific quantitative models. Altogether, our results so far are consistent with the dual-task studies in supporting the serial model. It is important to note, though, that we have tested only displays with words placed above and below the fixation point, rather than in more naturalistic arrangements.

One last consideration is the possibility of parallel orthographic processing across multiple words, as is assumed by some models of reading (Snell, van Leipsig, et al., 2018). It is possible to assume such parallel processing of letters across the visual field even if a subsequent lexical or semantic stage is serial, word by word. That could explain the positive redundant target effects when the two targets both the same word, and thus benefit from parallel letter identification (Hasbrooke & Chiarello, 1998; Mohr et al., 1994; Mullin & Egeth, 1989). Such parallel orthographic processing could also result in *interference* between the processing of multiple words that contain different letters, perhaps explaining the negative effects of redundant targets in the experiments above. This explanation is essentially a non-standard parallel model with dependencies between each stimulus process. If it produces strong interference between neighboring words that contain different letters, then a real reader would be incentivized to adopt a serial strategy: attend to one word at a time to minimize that interference.

Relation to accounts of the equivalence of parallel and serial models

A number of theoretical studies have shown that some very general parallel and serial models can mimic one another (e.g., Townsend & Ashby, 1983). These were important results, which some have taken this to mean that it is impossible to distinguish all kinds of parallel and serial models (for a helpful discussion, see Townsend, 1990).

In this article, we distinguish more specific versions of parallel and serial models. In particular, the three models considered here follow previous work in making independence and capacity assumptions. The independence assumptions concern the processing of each stimulus, which is independent of the other. The capacity assumptions are what distinguish the models (unlimited-capacity, fixed-capacity, and serial), which are still general in the sense that they assume no particular stochastic process or distribution. These "standard" models are not equivalent and *can* be distinguished. That is not inconsistent with the 'model mimicry' results. To make the parallel and serial models mimic each other, one must relax the independence assumptions and/or capacity assumptions (Little et al., 2015; Townsend & Ashby, 1983). Such very general models can mimic one another while the standard versions of the models cannot. Some parallel-serial mimicry can also be accomplished if certain models assume exhaustive search rather than self-terminating search, but the self-terminating assumption has been found more tenable in most conditions (Townsend & Nozawa, 1995, 1997; Van Zandt & Townsend, 1993).

In this article, we focus on redundant target effects and whether they are positive or negative. We also improve the models to include errors. Previous attempts to model redundant target effects have ignored errors and thus only considered positive (or 0) effects. They were also focused on distinguishing coactivation versus separate activation models (as reviewed in the Introduction). Nevertheless, it is very likely that equivalence can be found between the standard self-terminating serial model and standard parallel models with (very) limited capacity. Our standard *fixed-capacity* parallel model makes a related point because it can have either negative or positive redundant target effects.

Given model mimicry, what can we learn from testing specific models? Each time there is evidence against the standard parallel model, one must "patch" the model with some addition. The use of multiple behavioral paradigms, as reviewed above in the section on word recognition, can result in a need for several such patches. As more patches are required, the model becomes less plausible and less useful. Ultimately, the simpler model is preferred (in this case, the standard serial model for semantic judgements of written words).

In summary, we focus on identifying and testing relatively general standard models that have three advantages: they are straightforward to implement and interpret; they do not assume any particular stochastic process or response time distribution, and they include errors while also modeling respond time. Some of these models can be distinguished using the redundant target effects.

[redacted] condition, which our experimental design was not optimized for. The standard self-terminating serial model predicts particular changes in the shape of the RT distribution from one-target to two-target trials, primarily because two-target responses are the mixture of two distributions with increased variability. The experiments reported above did not include enough trials per participant per condition to make those tests, but they could be done in the future. Additional theoretical word would then need to derive corresponding predictions about RT distributions from various types of fixed-capacity parallel models.

Relation to previous theory on response time and accuracy

Most previous work has followed one of two paths. One is to develop a pure response time theory that ignores errors (e.g., Townsend & Ashby, 1983; Townsend & Nozawa, 1995). This work has also been general in not assuming particular stochastic processes or response time distributions. The second path is to assume a specific stochastic process such as the diffusion process or the linear ballistic accumulator (both described in the Appendix; see Luce, 1991, for an introduction). For example, Blurton, Greenlee & Gondan (2014) built on the diffusion process to model the redundant target effect. The strength of this path is the integrated treatment of response time and accuracy.

Here we sought to expand the general response time models to incorporate errors. The surprising result was that the standard, self-terminating, serial model with errors showed a *negative* redundant target effect. This is not predicted by the corresponding pure response time model.

This work complements other recent effects to generalize pure response time models. In particular, Little et al. (2022) extended part of the theory of systems factorial technology to include errors. They examine the prediction of the double-factorial paradigm to distinguish parallel and serial processes. They showed that the previous analysis of exhaustive search models was general to conditions with errors. However, they did not find a similar general result for self-terminating search models. Instead, they examined two special cases and showed that the analysis for pure response time did generalize to those cases. This is important progress, but it remains to be determined if this method of distinguishing parallel and serial models holds for *all* standard, self-terminating models with errors.

In summary, a critical development is the creation of general theories of both response time and accuracy. We have developed such a theory for the redundant target paradigm.

Conclusion

This study makes two contributions: first, we generalized three standard models of the redundant target effect, which yielded new ~~results~~ predictions. By accounting for errors as well as response time, the standard, self-terminating serial model predicts that redundant targets can slow correct responses, even when they increase accuracy. In contrast, the standard, self-terminating, unlimited-capacity parallel model always predicts positive redundant target effects, even when allowing for errors. We also developed specific examples of standard, self-terminating, fixed-capacity parallel models, some of which can predict negative redundant target effects.

Second, we presented experimental tests of these predictions for judgements of words. When the task required judgment of the letter colors, a positive redundant target effect rejected the standard self-terminating serial model. This is the result most commonly observed in the redundant target literature. But when the task required judgments of the words' meaning, a negative redundant target effect rejected the standard, self-terminating, unlimited-capacity parallel model and were instead consistent with either the standard self-terminating serial model or some variants of the standard fixed-capacity parallel model. This stands in contrast to most previous studies of word recognition that found positive redundant target effects and argued against the standard serial model.

In sum, the primary contribution of this work is the demonstration that negative redundant target effects do occur. We provide a simple serial model that can readily account for them, and we demonstrate that they are also consistent with some (but not all) ~~limited~~ fixed-capacity parallel models. ~~Further theoretical work must develop the fixed-capacity parallel models and devise ways to test them empirically.~~ Thus, the redundant target paradigm provides no evidence against a serial model of word recognition.

Acknowledgments: We are grateful to Nicole Oppenheimer for help with data collection. The National Eye Institute provided funding (grant R00 EY-029366 to ALW).

Conflict of interest statement: The authors have no conflicts of interest to declare.

Author contributions statement: Conceptualization, AW & JP; Investigation; AW, GS & JH; Methodology, AW & JP; Formal Analysis: AW; Theory development: JP and ZZ; Writing: AW, JP & GS.

Public data repository: All stimuli, data, and analysis code are available at: https://osf.io/a9kqj/.