# Serial processing of two words becomes parallel when they combine to form a compound word

Amritha Anupindi*, Liana R. Eisler*, Alex L. White
* co-first authors

Barnard College, Columbia University

Corresponding author:
Alex White, alwhite@barnard.edu
3009 Broadway
New York, NY 10011

**Acknowledgements**
NIH grant R00 EY02936 supported this work. We are indebted to Nicole Oppenheimer and Kimya Firoozan for assistance with data collection.

**Declaration of Interest**
The authors have no competing interests to declare.

**ABSTRACT**

According to one model of reading, words are recognized one at a time because of a bottleneck in the brain that cannot process two words simultaneously. Consistent with this serial model, we first show that participants can report lexical information about only one of two unrelated words that are flashed briefly above and below fixation and then masked. We then investigate whether two words that compose a single compound word (e.g., bottle + neck) can nonetheless be processed in parallel. The results demonstrate that indeed, under the same conditions in which two unrelated words cannot be recognized simultaneously, accuracy for recognizing either or both of two words that form a compound exceeds the prediction of the serial model. This result complicates theories of a serial bottleneck in lexical access, especially in the context of natural reading when words often form meaning jointly.

**Key words**

Word recognition; reading; visual attention; divided attention; language; parallel processing.

**HIGHLIGHTS**

- Skilled readers could not identify two unrelated words at the same time.
- Nonetheless, accuracy increased when two words formed a compound (e.g., pop+corn).
- Thus, compositional semantic relations between words facilitates parallel processing.
- Hypothesis 1: a serial process operates over linguistic units that span letter strings.
- Hypothesis 2: parallel processing is subconscious; conscious perception is serial.

**INTRODUCTION**

To understand how skilled readers read, we must determine how much information they can process at each glance at a page. That information intake is constrained by processing capacity limits inherent to the human visual system and linguistic functions in the brain. To read this page, for instance, you must precisely shift your attention down each line to process the text in small chunks (Rayner, 2009). That is primarily because words are not legible in peripheral vision. Nonetheless, the "span" of legibility in the central visual field is wide enough to allow multiple words to be read during a single gaze fixation (Legge, Mansfield, & Chung, 2001; Veldre, Reichle, Yu, & Andrews, 2023; Yeatman & White, 2021).

Does that mean that readers extract semantic information from multiple words simultaneously? On the one hand, some researchers argue for parallel processing of multiple words, citing evidence such as the parafoveal preview effect on fixation durations in natural reading (reviewed by Schotter, Angele, & Rayner, 2012). In experimental tasks, the sentence superiority effect, transposed word effect, and flanker effects have also been cited to support computational models that assume that readers distribute their attention to process multiple words in parallel during each gaze fixation (reviewed by Snell & Grainger, 2019).

On the other hand, serial models of reading propose that readers process one word at a time, via shifts of attention from word to word within each fixation (Pollatsek, Reichle, & Rayner, 2006). One series of experiments has provided support for the serial model by suggesting that it is not possible to recognize two English words at exactly the same time (White, Palmer, & Boynton, 2018, 2020). In these experiments, two randomly selected words were flashed briefly on either side of fixation and then replaced by post-masks (as in Figure 1 below). The key manipulation was that participants were either pre-cued to attend covertly to one side in order to semantically categorize just one word, or they were pre-cued to attend to both sides in order to categorize both words independently. The result was that accuracy in the divided attention condition dropped so far that it supported an 'all-or-none' serial model: participants could recognize one of the two words on each trial but had to make a random guess about the other. The post-masks were key to this result: they prevented participants from being able to process one word and then shift their attention, serially, to the other word.

Converging evidence is provided by several other studies that used the same approach to limit the time available to process pairs of words, but with variations on the types of stimuli and task requirements (Brothers, 2022; Campbell, Oppenheimer, & White, 2024; Johnson, Palmer, Moore, & Boynton, 2022). These behavioral results were also supported by fMRI evidence that the "visual word form area" in left ventral temporal cortex also responds to just one word at a time (White, Palmer, Boynton, & Yeatman, 2019).

In all of that prior research, the words presented simultaneously were randomly selected and unrelated (for an exception, see Brothers, 2022). In natural text, however, words make meaning jointly. To investigate whether compositional semantic relations facilitate parallel processing, we adapted the divided attention paradigm described above and introduced pairs of words that form compound words (e.g., stair + case). Our question is whether such words can be processed together even under conditions when two unrelated words cannot be recognized simultaneously.

In Experiment 1, participants first engaged in a divided-attention lexical decision task like those described above. Replicating prior results, divided-attention accuracy supported the serial model: participants could recognize only one word per trial. Then the same subjects performed a second task with nearly identical stimulus sequences, except the task was to report whether the two words formed a compound word. If only one of the two words could be processed on each trial, then accuracy for this "compound word judgment" should be at chance. As shown below, it was not.

In Experiment 2, we directly compared processing capacity for word pairs that do or do not form compounds with a novel full-report typing task. Again, two words were flashed briefly followed by post-masks, and participants were prompted to type in the word they saw at one post-cued location. Most word pairs were unrelated, but unbeknownst to the participants, some formed compound words (e.g., water + fall). Unlike in Experiment 1, these participants were not searching for compound words and were motivated to process each of the two words independently on each trial. Nonetheless, the data show that the relation between two words that form a compound can be detected, despite the apparent one-word-at-a-time processing bottleneck.

In both experiments, we placed the words just above and below the point of gaze fixation. This is not the format in which words are typically read, nor is it a typical way to present the two halves of a compound word. But we chose this arrangement for two reasons: first, the two words are both close to fixation and equally legible; second, this arrangement has provided strong evidence for serial processing of two unrelated words (White et al., 2020). Our experiments therefore provide a strict test of the hypothesis that compositional semantic relations between two words can facilitate parallel processing.

# EXPERIMENT 1

## Methods

*Participants*

10 volunteers (6 female, 2 male, 2 non-binary, ages 18-23 years) with normal or corrected-to-normal visual acuity participated in Experiment 1 in exchange for fixed monetary payment ($20/hour). Each subject gave informed consent in accordance with the Declaration of Helsinki and the Barnard College Institutional Review Board. All subjects were naïve to the purposes of the experiment and reported learning English before the age of 5. On the composite TOWRE-II Test of Word Reading Efficiency (Torgesen, Rashotte, & Wagner, 1999), the mean score was 114 (SEM = 3.7). Nine of ten participants scored above the norm of 100.

In the tradition of visual psychophysics, this study used a relatively small number of participants but many trials and multiple testing sessions per participant, with individually calibrated stimuli. The sample size was chosen ahead of data collection by a power analysis of a previous experiment that had a similar design (White et al., 2020). We estimated that at least 6 participants were needed to distinguish the fixed-capacity parallel and all-or-none serial models. We rounded up to 10 to be conservative and consistent with previous experiments. Each participant completed two tasks: a lexical decision task, and then a "compound word judgment task."

**Lexical decision task with unrelated words**

*Equipment and Stimuli*

We used custom MATLAB software (MathWorks) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) to present stimuli on a ViewPixx 3D screen (VPixx Technologies) with a 120 Hz refresh rate and 1920 x 1080-pixel resolution. The background brightness was set to the screen's maximum (100 cd/$m^2$). The stimuli consisted of a black fixation mark and black letter strings. The fixation mark was made of a black cross 0.38 degrees of visual angle (º) wide, with a 0.1º white dot at its center, and a thin black ring around it (0.38º diameter). The letter strings were written in Courier New font. The height of the letter "x" was 0.6º.

The stimulus set was composed of 820 real English words and 785 pronounceable pseudowords. Both categories were divided equally into strings of three, four, five, and six letters long. The real words came from all syntactic categories, ranging in lexical frequency from 2.4 to 6.3 Zipf (mean = 4.5). Zipf is a standardized measure of word frequency calculated through an adjustment of the log frequency per million words (Brysbaert & New, 2009; van Heuven, Mandera, Keuleers, & Brysbaert, 2014). The pseudowords were generated from the MCWord database (Medler & Binder, 2005) to have constrained trigram statistics, which makes them pronounceable and similar to real words. **Tables S2 and S3** at the end of the *Supplementary Materials* list all the words used in the experiment.

The post-masks were strings of non-letter characters drawn in a false font, BACS-2 Serif, which has several visual features matched to Courier New (Vidal, Content, & Chetail, 2017). Each post-mask was a real word from the stimulus set, of the same length as the word that preceded it, but presented in the illegible false font.

*Eye-tracking*

The subject's right eye gaze position was monitored at 500 Hz by an Eyelink 1000+ video-based eye tracker (SR Research). If the recorded right eye gaze position moved too far from the fixation mark during stimulus presentation, the trial was immediately terminated. Terminated trials were repeated at the end of the block, unless fewer than three trials remained. This applied to 10.7% (SEM = 2.4%) of trials on average in the lexical decision task. For more details, see the *Supplementary Materials.*

*Trial Sequence*

Each trial began with participants focusing on the fixation mark for a minimum of 200ms, followed by a 500 ms pre-cue. On divided attention trials, two green pre-cue lines, 0.16º long, appeared superimposed on the upper and lower arms of the fixation cross (**Figure 1A**). On focused attention trials, only one green pre-cue line indicated the side (above or below) to be post-cued. After a 500ms blank interval containing only the fixation cross, two letter strings appeared for 33ms, positioned above and below fixation and centered at 1.5º eccentricity, matching the stimulus positions used in the lexical decision task from Experiment 2 of White et al (2020). Each letter string had an independent 50% chance of being drawn from either of the two lexical categories (real word and pseudoword). The only constraints were that the letter strings on either side of fixation could not be identical, and neither string could have appeared in the previous trial.

The words were followed by an inter-stimulus interval (ISI), containing only the fixation mark. The ISI duration was set to each participant's threshold in the focused attention condition (mean ISI = 8 ms, SEM = 2.6 ms, range = [0 25]), as determined by a pre-test (see *Supplementary Materials).* After the ISI, two post-masks were presented for 250ms, at the same locations and with the same number of letters as the preceding words. Following another 100 ms blank interval (not shown in Fig. 1a), a post-cue appeared to indicate which word should be judged. The post-cue was a green line like the focused pre-cues. 500 ms after the post-cue appearance, a 25 ms click sound was played, which prompted the subject to press a key to report the lexical category (real or not real word) of the letter string on the side indicated by the post-cue. Keypresses before the click were not recorded. On focused attention trials, there was only one post-cue and one response (to the one word that was pre-cued). On divided attention trials, subjects were asked to judge both words in a random order. After the first post-cue, click, and keypress response, the post-cue reversed to the other side. 300 ms later, a second click prompted the second response.

Subjects pressed one of four keys (*a, s, d,* or *f*) with their left hand for words in the top location, or one of four keys (*m, <, >,* or *?*) with their right hand for words in the bottom location. For each hand, from left to right, the keys indicated "sure pseudoword," "guess pseudoword," "guess real word," and "sure real word."

Feedback about response accuracy (regardless of confidence level) was given with 100 ms auditory beeps (high pitch for correct, low pitch for incorrect). On focused attention trials, the feedback beep was played 350 ms after the participant's keypress. On divided attention trials, both feedback tones were played after both responses were recorded. After a 500ms inter-trial interval (ITI), the next trial began.

The cue condition (focused top, focused bottom, or divided) was blocked. Blocks of 20 trials each were run in sets of four: two divided attention and one of each focused attention condition (above and below), in a random order.

*Organization of experimental sessions*

The lexical decision task required four to five 1-hour sessions. In session 1, subjects completed the TOWRE test of word reading efficiency, received instructions, practiced the task, and ran a staircase procedure that adjusted the duration of the ISI to their 75% correct threshold in the focused attention condition (details in the *Supplementary Materials*). The main experiment began in session 2. The ISI was set to the staircase estimate and then adjusted as needed to maintain focused attention accuracy between 70 and 90% correct. Any run of 4 - 12 blocks with an ISI that was too high or too low was discarded and re-run. This exclusion applied to four blocks for two subjects, 8 blocks for one other, and 12 blocks for one other. The ISI did not differ between focused and divided attention conditions within each set of four blocks. Testing sessions continued until each subject had completed a total of 50 blocks (1,000 trials).

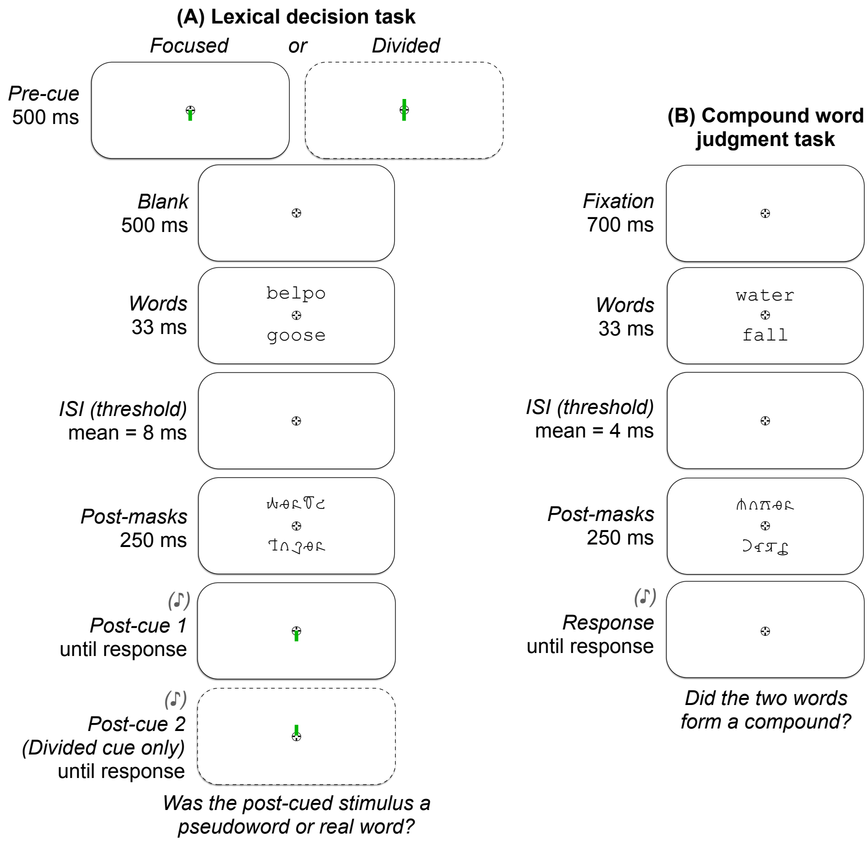


***Figure 1.*** Example trial sequences. (**A)** Lexical decision task. The example shows either a focused attention trial with the target on the bottom side, or a divided attention trial (dashed outlines). The focused cues were equally likely to point to the top side or the bottom side. Not shown is a 100 ms blank between the post-masks and the post-cue. The musical notes (♪) indicate clicks that were played 500 ms after each post-cue onset, prompting the participant to respond. (**B)** Compound word judgment task. On a minority of trials (not represented here), the ISI was set to 400 ms. The musical note (♪) indicates a click that was played 300 ms after the post-masks offset, prompting the keypress response. In both tasks, feedback for each response's accuracy was delivered with a beep.

*Analysis*

We used the confidence ratings to compute accuracy in units of area under the receiver operating characteristics (ROC) curve, $A_g$, a bias-free measure of accuracy in each condition (Pollack & Hsieh, 1969). $A_g$ ranges from 0.5 (chance) to 1.0 (perfect). One can think of $A_g$ as an unbiased estimate of proportion correct.

**Compound word judgment task**

The same 10 participants then completed the following compound word task. All methods were the same as in the lexical decision task except as indicated reported below.

*Stimuli*

The stimulus set (see Supplementary **Table S3)** was constructed from 193 English compound words divided into their constituent words (e.g., waterfall → water + fall). These compound words were defined as whole words that often appear in written form without a space between the two constituents. The set included both semantically transparent (e.g., arrowhead) and opaque (e.g., deadline) compound words. There were 293 unique constituent words, each of which appeared on average 2.6 times within the stimulus set and 3.3 times in the experiment for each subject. The constituent words came from all syntactic categories, ranging in length from three to six letters and ranging in lexical frequency from 2.8 to 6.3 Zipf (mean = 4.6). The whole compound words ranged in lexical frequency from 1.8 to 6.2 Zipf (mean = 3.1). 14 of the 193 compound words did not appear in the SCOPE SUBTLEX-US Zipf database, but all were verified as being familiar compounds by five research assistants.

*Trial Sequence and Procedure*

On each trial, two real words were presented simultaneously above and below fixation (**Figure 1B**). The word above fixation was always the first constituent of a compound word (e.g., "stair" in staircase), and the word below fixation was the second constituent of a compound word (e.g. "neck" in bottleneck). On half of all trials, the two constituent words were paired correctly to form a compound word together. On the other half of trials, the two constituent words came from different compounds. A given pair of words appeared on average 1.3 times in the experiment for each subject. Subjects completed 500 trials total in this experiment (20 blocks of 25 trials each).

There were no pre- or post-cues. Each subject's word-to-mask ISI duration was matched to the last ISI used for that subject in the lexical decision task. The ISIs across subjects ranged from 0-17ms (mean = 4ms, SEM=1.6ms). On 33% of all trials, the ISI was set to 400 ms instead of the threshold ISI. These provided the subjects with some "easy" trials and allowed us to ensure that the task was easy when subjects were given more than enough time to process both words (see Supplementary Figure S2).

300 ms after the offset of the post-masks, a beep prompted the participant's response. Responses before the beep were not recorded. Participants pressed one of four keys (*a, s, d,* or *f*), which indicated, in order left to right, "sure not compound word," "guess not compound word," "guess compound word," and "sure compound word." Auditory feedback was given as in the lexical decision task. On average, 9.7% (SEM = 2.2%) of trials were terminated due to fixation breaks.

This compound word task required one 1-hour session. Subjects received instructions and practiced the task before completing 20 blocks (25 trials per block).

## Results

*Accuracy in the lexical decision task supports the all-or-none serial model*

Lexical decision accuracy, reported in units of area under the ROC curve ($A_g$), was on average 0.82 (SEM = 0.008) in the focused cue condition. It was considerably lower in the divided cue condition, by a mean difference of 0.18 (SEM = 0.05, $t(9) = 10.79$, $p<10^{-5}$; BF = 8826). To compare this large deficit to the predictions of three different models, we plot our data on attention operating characteristics (AOCs; Sperling & Melchner, 1978). The mean AOC, shown in **Figure 2A**, plots accuracy for words above fixation against accuracy for words below fixation. The two focused-attention accuracies are pinned to their respective axes. The accuracy with divided attention forms a single point (open circle) in that 2-D space, which we then compare to the predictions made by three quantitative models of capacity limits (Bonnel & Prinzmetal, 1998; Scharff, Palmer, & Moore, 2011; Sperling & Melchner, 1978; White et al., 2018). Note that these are not all possible serial and parallel models, but they provide clear benchmarks against which to compare task performance. The three models are:

1. *Independent parallel model*: Two stimuli can be processed simultaneously just as well as single stimuli with focused attention. Thus, there is no cost to dividing attention. In the

AOC, this model predicts that the divided attention point will fall at the intersection of the dashed lines.

2. *Fixed-capacity parallel model*: The perceptual system can extract a fixed amount of information from the entirety of the stimulus display on each trial. Processing resources are devoted to just one word in the focused cue condition but shared between two words in the divided attention condition, thus, there is a modest deficit. This model traces the black curve in the AOC.
3. *All-or-none serial model*: The meanings of words are processed one at a time. Because the post-masks appear after an interval set to each participant's threshold, there is just enough time to recognize one word with focused attention. In the divided attention condition, participants can therefore process one word fully. If they then try to process the second word, it has been replaced by the post-masks, which removed any sensory memory trace. Therefore, only one word can be processed per trial, with equal accuracy as in the focused attention condition. When participants are asked to judge the word they did not process, they make a random guess. This model traces out the diagonal black line in the AOC.

As shown in **Figure 2A,** mean accuracy for the lexical decision task was best predicted by the all-or-none serial model. To test each model statistically, we constructed AOCs for each participant (see Supplementary **Figure S1**) and then computed the Euclidean distances between the divided attention point and the nearest points on the serial model's prediction line and the fixed capacity parallel model's prediction curve. Points below the predictions were assigned negative values. **Table 1** lists statistics on the means of these two distances. Divided attention accuracy significantly below the fixed-capacity model's prediction but was indistinguishable from the serial model's prediction. Evaluated individually, all but one participant was best predicted by the serial model (**Figure S1**).

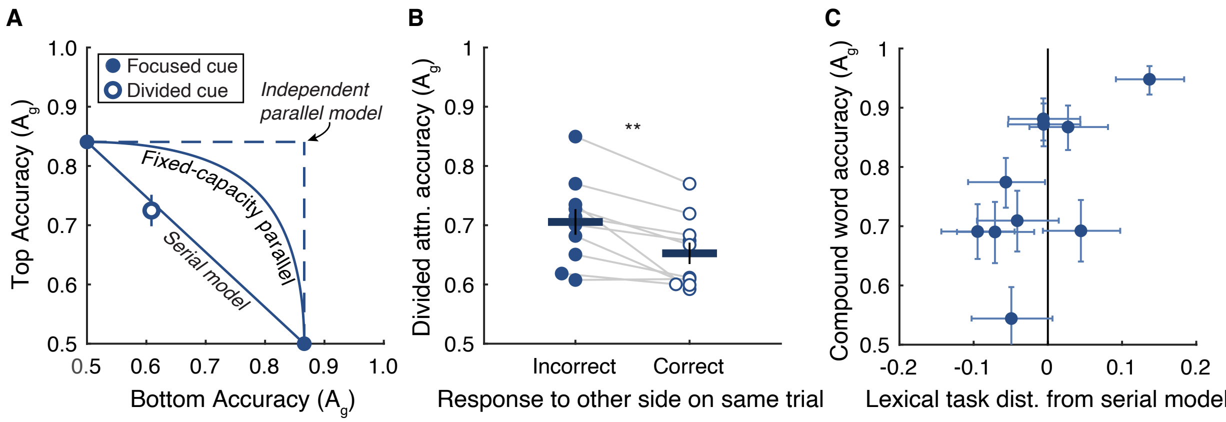


**Figure 2.** Results of Experiment 1. **(A)** Mean attention operating characteristic (N=10) for the lexical decision task. Solid points pinned to the axes represent focused attention accuracy (in units of area under the ROC curve, $A_g$). The open point represents divided attention accuracy. Error bars are ±1 SEM. Divided attention accuracy is closest to the all-or-none serial model's prediction. Individual subject AOCs are plotted in the Supplementary Material. **(B)** Stimulus processing tradeoff effect on accuracy ($A_g$) in the divided attention condition of the lexical decision task. Horizontal bars show across-subject means. Each dot is an individual participant's data, with each participant's two points connected by a thin gray line. The asterisk indicates $p < 0.01$. **(C)** Individual participants' accuracy ($A_g$) in the compound word judgment task plotted against their result from the lexical decision task; specifically, the x-axis is the distance of each participant's divided attention point to the all-or-none serial model (see Table 1). Both horizontal and vertical error bars represent 95% bootstrapped confidence intervals.

Thus, the AOC analysis allows us to reject both parallel processing models and supports the serial model: participants could fully process only one of the two words (or pseudowords) presented on each trial. These results match what has been reported before in a similar study with unrelated word pairs (White et al., 2020).

| Model of Comparison | Mean | SEM | $\lvert t \rvert$ | $p$ | 95% Confidence Interval | Bayes Factor |
|---|---|---|---|---|---|---|
| Fixed-capacity parallel | -0.105 | 0.02 | 4.798 | 0.001 | [-0.138 -0.051] | 42 |
| All-or-none serial | -0.012 | 0.02 | 0.542 | 0.601 | [-0.044 0.036] | 0.35 |

**Table 1.** Lexical decision performance: Mean distances of the divided attention point from the all-or-none serial and fixed-capacity parallel processing models. Statistics were computed by two-tailed t-tests, with 95% confidence intervals and Bayes Factors.

*A stimulus processing tradeoff also supports the serial model for the lexical decision task*

The all-or-none serial processing model further assumes that there is a trial-by-trial tradeoff between the two words, because participants can process only one per trial. This predicts that accuracy for each side (top word or bottom word) should be lower on trials when the response to the other side was *correct* than on trials when the response to the other side was incorrect (Braun & Julesz, 1998; Lee, Koch, & Braun, 1999; White et al., 2018, 2020). For example, if the response to the top word is correct, the top word was probably processed and therefore the response to the bottom word is less likely to be correct.

We tested that prediction by separating all responses on divided attention trials into two sets: (1) the response to the other side on the same trial was correct, and (2) the response to the other side on the same trial was incorrect. Within each set, we computed accuracy ($A_g$). These data are plotted in **Figure 3B.** The horizontal bars are the means and each participant's two data points are connected by a thin gray line. Accuracy was significantly worse when the response to the other side was incorrect than correct (mean difference = 0.053, SEM = 0.012; $t(9) = 4.06$, $p = 0.003$; BF = 17.3). This negative correlation between accuracies for the two responses is rare for dual-task performance, and, like the AOC analysis, rejects the two parallel models described above and supports the serial model.

*Accuracy in the compound word task nonetheless exceeds the serial model's prediction*

In the compound word judgment task, participants also viewed pairs of words flashed at the same positions and masked after the same brief interval. The task was to report whether the two words together formed a compound (like stair + case or grand + father), which was true on a random 50% of the trials. On the remaining trials, the same individual words were mismatched to not form compounds (e.g., stair + father). In one sense, this is also a lexical decision task: to report whether the two letter strings together form a single word or not.

A strict version of the serial model would assume that participants in this task are still able to identify only one of the two letter strings per trial. This predicts that accuracy in the compound word judgment task should be at chance ($A_g = 0.5$, or 50% correct), because identifying any one word provides no information as to whether the other word formed a compound with it or not.

However, accuracy on trials with the threshold-level ISI greatly exceeded the chance level (mean $A_g$ = 0.767, SEM = 0.037, 95% bootstrapped CI = [0.19 0.33]; $t(9) = 6.84$, $p < 0.0001$; BF = 362). This rejects the strict all-or-none serial model's prediction. (See Supplementary **Figure S2** for an analysis of accuracy in units of $d'$).

**Figure 2C** further demonstrates how the serial model fails to account for compound word accuracy at the level of individual participants. Each participant's accuracy in the compound word judgment task (y-axis) is plotted as a function of their distance from the serial model in the divided-attention lexical decision task (x-axis). These distance metrics are negative when lexical decision accuracy was lower than the serial model predicted.

All participants except one performed significantly above chance in the compound word task (in Fig. 2C their vertical 95% CIs exclude 0). That is true even when their lexical decision performance was near or worse than the serial model's prediction. (Only one participant, represented by the rightmost dot in Figure 2C, outperformed the serial model in the lexical decision task). This result should not have occurred if the participants could recognize only one half of each compound pair, as would be predicted by their lexical decision performance.

We may also make a prediction based on the independent parallel model. It assumes that both letter strings are processed simultaneously, with the same accuracy as when attention is focused on one word at a time. We may further assume that the probability of a correct response in the compound word task is the product of the probabilities of correct responses to the top and bottom words in the focused attention conditions of the lexical decision task. In other words, the probability of correctly determining that the two words form a compound is the joint probability of recognizing both words independently. The mean p(correct) predicted by this model is 0.659 (SEM=0.018), which was somewhat *lower* than the empirical compound word accuracy levels (mean difference = 0.087, SEM=0.129, 95% CI = [0.001 0.159], $t(9) = 2.03$, $p = 0.0724$; BF = 1.336). Thus, the prediction based on independent parallel processing of the two words comes close to *underestimating* the empirical compound word judgment accuracy.

One concern is that participants saw the words in this experiment more than once. Although each individual word was equally likely to appear in a compound pair or scrambled pair, it is possible that learning compound words boosted accuracy. However, we also analyzed the subset of trials when each word pair appeared for the first time (see Supplementary **Figure S2**). Accuracy still exceeded chance by a healthy margin. Therefore, the above-chance accuracy is not due to learning of compound word pairs.

*Interim Discussion*

For the lexical decision task with pairs of unrelated words, we reject two standard parallel processing models in favor of a serial model (one word processed per trial). The serial model accounts for both the cost of dividing attention and the negative stimulus processing tradeoff in accuracy, replicating prior experiments with unrelated word pairs (White et al., 2020). Nonetheless, the serial model cannot account for above-chance accuracy in the compound word recognition task, in which participants viewed very similar stimulus sequences as in the lexical decision task. This suggests that when two words combine to form a compound word, they can pass through any "bottleneck" in the word recognition circuitry together.

One interpretation of this result is that the serial bottleneck arises at the stage of lexical access: identifying a set of input letters with a concept in memory. The set of input letters could be divided into two strings with space between them (as in the experiments here). Thus, the compound word result is compatible with the serial model, if we assume the serial processing does not apply to single letter strings, but rather to single lexical items that can span multiple letter strings.

If this account is true, one may ask whether such holistic processing of multiple letter strings depends on the participants' voluntary effort to process them in that manner (which they were doing in the compound word judgment task). Experiment 2 was designed to answer that question.

## EXPERIMENT 2

Experiment 2 investigated whether the two words that happen to form a compound can be processed in parallel, under conditions when compound words are rare and unexpected, and when participants attempt to process the two words independently.

Subjects were presented with two real words on each trial and their task was to report one post-cued word by *typing* it into a text box (see **Figure 3**). In separate blocks, a pre-cue either instructed participants to focus attention on the top side or bottom side (because they would only be asked to report the word on that side), or to divide attention between both sides (because they could be asked to report either one of the two words). On 70% of all trials, the two words presented together were unrelated. Participants were always motivated to process the two words independently as best they could. On a random 15% of trials, the two words formed a compound word – for example, "cow" above fixation and "boy" below. Participants were not informed of the presence of compound words and almost never noticed them, as demonstrated by debriefing interviews after the experiment. The primary control condition that provides the baseline against which we compare accuracy used "scrambled compound" pairs: the halves of two different compound words were mismatched. We also included a "reversed compound" condition when the two halves of one compound word were in reverse order (the second word on top).

### Methods

All methods were the same as in Experiment 1 except for as described below.

#### *Participants*

13 volunteers (all female, ages 18-22) were recruited in the same manner and from the same pool as in Experiment 1, again with informed consent. The sample size was determined based on a power analysis of a pilot experiment with a similar design. On the composite TOWRE-II Test, the mean score was 112 (SEM=2.4), with all but one of the participants scoring above the norm of 100.

*Stimuli*

The stimulus set (which is reproduced in full in the *Supplementary Materials*, **Tables S4-S6***)* consisted of 918 English nouns from the MCWord database, ranging in length from three to six letters and in lexical frequency from 2.0 to 6.7 Zipf (mean = 4.4). 242 words were selected such that they could form 121 compound words when combined in pairs (e.g., "water" and "fall"). As in Experiment 1, this set included both semantically transparent and opaque compound words. 66.4% of the compound words also appeared in Experiment 1. The remaining 676 nouns that did not form compound word pairs were used for the "random pair" condition. The distributions of length, concreteness, orthographic neighborhood, and lexical frequency in the "compound" word set were roughly matched to the "random" set (see the *Supplementary Materials*).

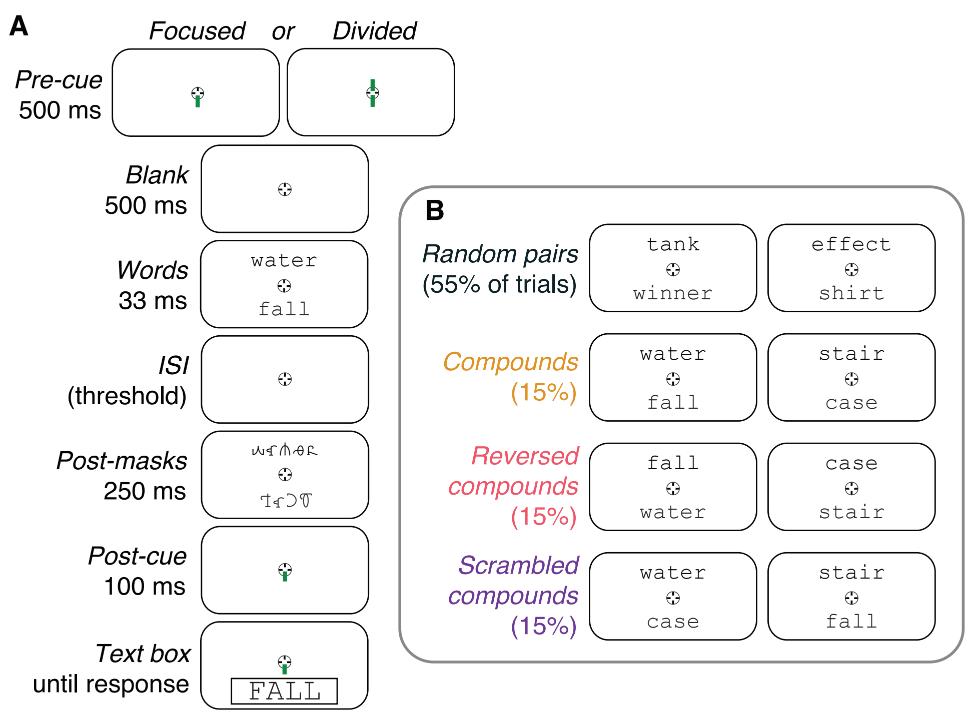


***Figure 3.*** Design of Experiment 2. (**A**) Example trial sequence, which could be a focused attention or divided attention trial, depending on the pre-cue. The mean threshold-level inter-stimulus interval (ISI) was 12 ms (range 0 to 33 ms). The participant's task was to type in the word they saw on the post-cued side (in this case, the bottom side). Not shown is a 100 ms blank interval between the post-masks and the post-cue. Central gaze fixation was enforced. (**B**) Examples of the four different word pair types, along with the percentage of trials in which they appeared.

*Trial sequence and procedure*

The trial sequence, which is illustrated in **Figure 3A,** was the same as the lexical decision task of Experiment 1 except as indicated here. At the end of the trial, a post-cue (green line) pointed to either the top or bottom side, followed 100 ms later a black-outlined text box appeared, centered 2.1º from fixation on the same side as the post-cue. (The post-cue remained visible along with the text box). The post-cued side matched the pre-cue on focused attention trials and was equally likely to be the top or bottom on divided attention trials. The response was prompted by the appearance of the text box, with no accompanying beep (unlike Experiment 1). The participant then had unlimited time to *type in* the word they had seen on the post-cued side. The letters they typed appeared in capital letters within the text box, and they could delete letters as necessary, then press the return key once they were satisfied. Feedback for correct or incorrect responses was delivered with beeps as in Experiment 1.

As shown in **Figure 3B**, there were four types of word pairs: (1) *Random pairs*: two randomly selected nouns from the set that do not form compounds. These pairs appeared in 55% of trials. (1) *Compound pairs*: the two words formed a compound word in the correct order, for example "sun + flower" and "grand + father." The first constituent word (e.g., sun in sunflower) appeared above fixation, while the second constituent (e.g., father in grandfather) appeared below fixation. (2) *Reversed compound*: two words could form a compound but were presented in reverse order, with the second constituent on top and the first constituent on bottom. For example, "flower + sun" and "father + grand." (3) *Scrambled compound*: the two words presented together were mismatched constituents from two different compound word pairs. The top word was the first constituent from one compound pair and the bottom word was the second constituent from another. For example, "sun + father" and "grand + flower".

The "non-compound" condition was presented in 55% of trials, while the other three conditions were presented in 15% of trials each (randomly intermixed). The participants received no explicit information about these conditions, were not aware that some word pairs would form compound words, and were solely instructed to enter the *single* post-cued word for each trial (even on divided attention trials). Importantly, the scrambled compound condition is the baseline against which we compare the reversed and compound conditions, as they all used the same set of constituent words.

The study required 2-3 one-hour sessions per participant. Following the instructions in the initial session, the participant practiced the task with slow stimulus presentation. Then we used a staircase procedure to estimate participant's threshold for the stimulus-mask ISI, as in Experiment 1 using focused cue trials only. The staircase converged on the 67% correct threshold. Then each participant completed 640 trials of the main experiment (in 32 blocks of 20 trials). The cue condition (focused top, focused bottom, or divided) was blocked, and the word pair types were randomly intermixed.

Throughout the experiment we adjusted the word-to-mask ISI to keep each participant's focused-attention accuracy between 65% and 85% correct. Any run of 4 to 12 blocks with an ISI that was too high or too low was discarded and re-run. The mean number of excluded blocks per subject was 4 (ranging from 0 to 16). The mean ISI on included trials was 12 ms (SEM = 2.5 ms, range across participants 0 to 33 ms).

*Analysis*

Our primary method of analyzing accuracy in this full-report typing task is to score a response as "correct" if the typed letter string exactly matches the post-cued word. This is a strict measure of how well the words were perceived, and it is the basis of the primarily analyses presented below. Specifically, *p(correct)* is the proportion of trials with an exactly correct response. To construct the AOCs from these data, we must specify the chance level of accuracy (accuracy achieved if the participant makes totally random guesses). This chance level forms the origin of the AOCs. To estimate the chance level, we assume that a participant with 0 information about the target word would type a word drawn randomly from the set of all 1,794 words that were in a set of unique words that includes the stimulus set and all words entered by all participants. Thus, the chance level is effectively 0: 1/1794 = 0.00057.

For these AOCs, we can compute the all-or-none serial model's prediction (diagonal line between the two focused cue accuracy points) and the independent parallel model's prediction (focused cue accuracy = divided cue accuracy). However, we lack the fixed-capacity parallel model for this task, as it has only been developed for 2-alternative forced-choice tasks like that used in Experiment 1. Therefore, we simply computed the minimum distance between each participant's divided-attention accuracy point and their own serial model's prediction.

An alternate method to compute accuracy is the "edit distance" between the participant's typed response and the target word that was presented. This metric could be sensitive to partial information about the words that the participant perceived. However, as briefly described in the discussion section below and reported at length in the *Supplementary Materials*, the analyses of the edit distance led to the same conclusions as the analyses of p(correct).

**Results**

*Accuracy is affected by the semantic relations between words in each pair*

**Figure 4A** plots the mean p(correct) in the focused and divided cue conditions, separated by word pair type. To analyze these data, we fit a linear mixed effects model to each participant's proportion correct, with fixed effects of cue and word pair type, and with random slopes and intercept by participant. First, overall accuracy in the divided attention condition was significantly lower than that in the focused attention condition ($F(1,96)=401$, $p<10^{-35}$). Mean accuracy on focused cue trials was 0.76 (+/- 1 SEM = 0.01), while accuracy on divided cue trials was 0.37 (+/- 0.02). Accuracy also varied across word pair types ($F(3,96) = 19.4$, $p<10^{-9}$), and word pair type interacted with cue condition: $F(3,96) = 15.4$, $p<10^{-7}$).

To interpret that interaction, within each cue condition we computed pairwise comparisons between each word pair type and the "scrambled compound" condition, which serves as our baseline. After correcting for false discovery rate (q=0.05), only one significant difference emerged: under divided attention, accuracy in the correctly ordered "compound" condition was greater than the "scrambled" compound condition (t(12=3.69, FDR-corrected p-value = 0.018, BF=15). **Figure 4B** plots that difference estimated from individual subjects' accuracy levels, for both focused and divided cue conditions. In the divided cue condition, all but one participant had higher accuracy on compound word trials than scrambled word trials.

This finding demonstrates that when participants divided attention because they could be asked to report either word, they were sensitive to the occasional semantic association between the two words as occurred when they formed a compound word together. That association improved the participants' ability to report either of the constituent words.

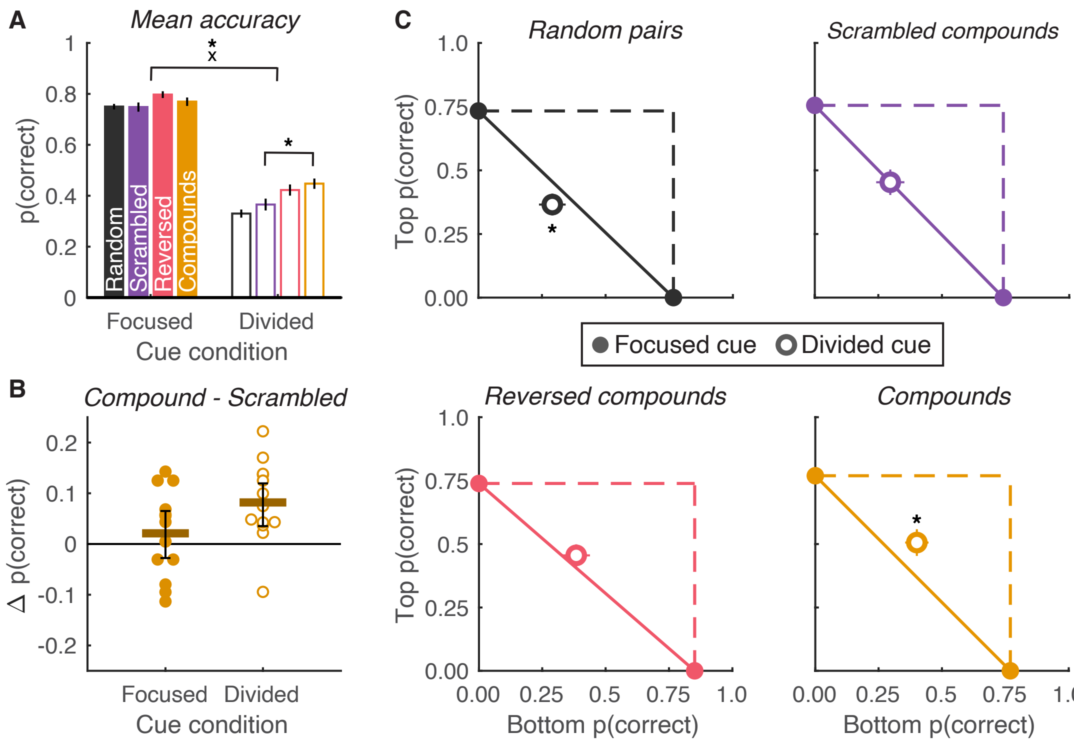


***Figure 4:*** *Results of Experiment 2.* (**A)** Mean accuracy: the proportions of trials with correctly reported words in each condition. Error bars are ±1 SEM. The central asterisk and "x" indicate the significant main effect of cue condition and the interaction between cue condition and word pair type. Within each cue condition, we compared all word pair types to the "scrambled compounds" baseline condition. The lower asterisk indicates the only significant difference: between compound pairs and scrambled compounds in the divided attention condition. **(B)** The *differences* in accuracy between trials with compound word pairs and scrambled compound pairs, separately for each cue condition. Each dot is an individual participant, and the horizontal lines are the means. Error bars in black indicate 95% *confidence intervals*. **(C)** Attention Operating Characteristics (AOCs) for each word pair type, constructed from mean accuracy levels. Error bars are +/- 1 SEM. Asterisks indicate that the minimum distance between the divided-attention accuracy point and the serial model's prediction line is significantly different from 0 (FDR corrected for the 4 comparisons).

The next step is to test the serial model with AOCs for each word pair type (**Figure 4C)**. For each word pair type, we constructed AOCs for each participant, computed the minimum distance between the divided-attention accuracy point and the serial model prediction, and then conducted a one-sample t-test on the 13 participants' distance measures. Those statistics are reported in **Table 2**. For the random pair condition, accuracy was significantly *worse* than the serial model predicts, highlighting the great difficulty in processing two unrelated words at the same time. For the scrambled compound and reverse compound conditions, accuracy was not distinguishable from the serial model's prediction (both BF<1). But for the compound word pairs, accuracy significantly exceeded the serial model's prediction, by about 9 percentage points in accuracy (BF=7.6). Thus, under the exact same conditions in which participants could only report one of two words presented in an unrelated pair, their accuracy improved when the two words happened to form a compound together. The AOC suggests that some parallel processing of the two words is possible in that condition, although with limited capacity.

| **Word pair type** | **Mean** | **SEM** | ***t*(12)** | **FDR-corrected *p*** | **BF** | **95% CI** |
|---|---|---|---|---|---|---|
| *Random pairs* | -0.073 | 0.020 | -3.51 | 0.0143 | 11.6 | [-0.110, -0.035] |
| *Compound* | 0.092 | 0.027 | 3.23 | 0.0143 | 7.64 | [0.037, 0.146] |
| *Reversed* | 0.046 | 0.032 | 1.39 | 0.2516 | 0.615 | [-0.018, 0.100] |
| *Scrambled* | -0.008 | 0.043 | -0.19 | 0.8543 | 0.283 | [-0.080, 0.083] |

***Table 2. Tests of the serial model in Experiment 2.*** For each word pair type (in rows), this table lists statistics on the distances between the divided attention accuracy point (in units of proportion correct) and the nearest point on the serial model's prediction line. BF = Bayes Factor; CI = confidence interval.

Participants were generally not aware that compound words were presented. Immediately after completing their last block of trials, the participant was asked, "Did you notice any relationship or pattern between the words?" All but two of the participants said no. Of these two participants, one noticed varying levels of difficulty among certain word pairs but remained skeptical that there was an underlying reason for it. The other reported noticing a single compound word (grass + hopper).

Nonetheless, there are two concerns to address: the first is biased guessing. The guessing hypothesis is that participants perceived only one word on the "compound" word pair trials, and when prompted to report the other side, they guessed a word that *would* form a compound with the word that they did perceive. This could explain the higher performance for compound word pairs, even if only one of the two words were recognized per trial. However, our data provide evidence to rule out this hypothesis. Specifically, we analyzed *incorrect* responses on scrambled compound trials in the divided attention condition. We counted how often participants reported a word that *did* form a compound with the word on the other side of the screen (which was not post-cued). 9 of 13 participants never did that. The remaining subjects guessed a compound on only 1 or 2 of these error trials. In the whole set of 628 error responses on scrambled-compound divided-attention trials, there were a total of 7 responses that did form a compound. Therefore, the relatively high accuracy for compound word pairs cannot be attributed to biased guessing.

The second concern is the occasional repetition of words. Each word was presented to each participant on average 2.3 times (including practice, staircase, and main experiment blocks). Each word was the post-cued target 1.7 times on average. Upon seeing a word for the second time during the experiment, the participant may process it faster, which could be especially beneficial on compound word trials. Overall, accuracy in the divided attention condition was slightly higher when a word repeated than when it appeared for the first time (mean benefit = 0.04, SEM=0.02). According to a linear mixed effect model, the increase in p(correct) was not quite significant ($F(1,4182)=3.21$, $p=0.073$), and it did not interact with word pair type ($F(3,4182)=0.26$, $p=0.85$).

We also constructed AOCs using only trials when *both* the words appeared for the first time (roughly 33% of the data). Accuracy for scrambled compound pairs was still close to the serial model's prediction (mean distance = -0.02), while accuracy for compound pairs was just as high above the serial model's prediction as in the main analysis of all trials (mean distance = 0.091). The standard error of that distance was larger (0.050 compared to 0.027), due to using a subset of the trials, and thus the t-test was not significant in this sub-analysis ($p=0.15$). Nonetheless, the slight benefit to overall accuracy when words repeated did not specifically benefit compound word pairs and cannot explain the main results reported above.

*Interim Discussion*

Experiment 2 confirms that when participants tried to recognize both of two *unrelated* words presented above and below fixation, they could report only one and made random guesses when asked about the other. Nonetheless, when the two words happened to form a compound word (as occurred unexpectedly on a minority of trials), accuracy rose modestly but significantly above the serial model's prediction.

Another unexpected result is that divided-attention accuracy for the "random pairs" was significantly *below* the serial model's prediction (upper left of Figure 4C). This suggests that participants could recognize only one word per trial, but they were slightly more likely to make an error in reporting it than in the focused-attention trials. This is not easy to explain, but it is the *scrambled* compound condition that provides the best control to compare against the compound word condition. The words in each scrambled pair were also unrelated to each other (like in the "random pairs" condition) and came from the same word set as those in the compound condition. In the scrambled condition, mean accuracy fell directly on top of the serial model's prediction.

The proportion-correct data analyzed above provide a strict test for how well participants could precisely report whole words. But might participants have perceived *partial* information about two words at once? If so, when they made an error, they may have reported a word that shares some letters with the target word. We also addressed this possibility by computing the "edit distance" between each target word and the word participant's type in, then normalized by the length of each word. This alternate measure of accuracy, which we call p(letters correct), can be considered the mean proportion of *letters* correctly reported.

As shown in **Supplementary Figure S4**, analyses of the p(letters correct) were largely consistent with the results reported above. For the random pairs and scrambled compound pairs, accuracy on the AOC was statistically indistinguishable from the serial model's prediction. For compound word pairs, accuracy was clearly higher than the serial model's prediction. Interestingly, for reversed compound pairs, accuracy was also significantly above the serial model, but by roughly half as much. Thus, there is some evidence that the association between the two halves of a compound influenced performance even when they were in the wrong order. Overall, the analysis of p(letters correct) demonstrates that the 'serial bottleneck' also applies to partial or sub-lexical processing of two unrelated words (Campbell et al., 2024), and confirms the significant benefit for words that form compounds together.

## GENERAL DISCUSSION

### *Summary*

In two experiments, we investigated capacity limits for processing pairs of words that form compound words, as compared to pairs of unrelated words. The data in both experiments demonstrate that a compositional semantic relation allows two words to pass simultaneously through the theorized "serial bottleneck" in word recognition (White, Boynton, & Yeatman, 2019).

First, the divided-attention lexical decision task in Experiment 1 demonstrated that two unrelated words could not be fully processed (to the point of lexical access) simultaneously. This is consistent with several previous studies (White et al., 2018, 2020; White, Palmer, et al., 2019). However, the same participants in the same time-limited conditions were able to perceive when the two words flashed together formed a compound word. That result suggests some parallel or holistic processing of the two halves of a compound word, even when presented on opposite sides of fixation.

The second experiment demonstrated that such parallel processing occurs even when participants do not attempt to combine the two letter strings into one word. In Experiment 2, we presented compound word pairs on a random minority of trials during a task that motivated participants to process two words independently. Compound words were rarely noticed by the participants. The serial model again accounted for the low accuracy when participants divided attention between two unrelated words, but it could not account for the rise in accuracy when the two words formed a compound. Thus, while the data for unrelated word pairs suggests that there is a "serial bottleneck" in word recognition, compound words seem to defy its constraints.

### *Theoretical implications*

We now consider two hypotheses to explain our results. The first we call the "high-level bottleneck" hypothesis: there is a serial bottleneck, but it lies at a late stage of processing when the orthographic input is matched to units of meaning that may encompass multiple letter strings, when they form linguistic structures together. Thus, the bottleneck pertains not to recognizing individual letter strings *per se*. The linguistic structure in a string of words determines how many of them can be processed at once. This hypothesis has been previously advanced to explain fixation patterns in the presence of English compound words (Cutter, Drieghe, & Liversedge, 2014) and developed as the "multi-constituent unit hypothesis" in the context of Chinese reading (Zang, 2019; Zang, Wang, Bai, Yan, & Liversedge, 2024).

This hypothesis relates to a more general question about compound words: does the language system automatically decompose the compound into its constituents and identify each of those before activating the meaning of the whole compound? That has been long debated (e.g., Andrews, 1986; Fiorentino & Poeppel, 2007; Ji, Gagné, & Spalding, 2011; Libben, Gagné, & Dressler, 2020; Sandra, 1990; Taft & Forster, 1976). Nonetheless, most models agree that the representation of the whole compound, as distinct from each of its constituents, is quickly accessed. The hypothesis advanced here is that such a whole-compound representation is the "unit of meaning" that is processed all at once.

This hypothesis still assumes that there is a serial process during reading, but it involves extracting meaning serially from chunks of text that may span multiple letter strings when those strings are meaningfully connected to each other (Snell & Grainger, 2019b). Unrelated words *cannot* be processed in parallel, according to this hypothesis. Thus, it may reconcile the apparent serial result for unrelated word pairs (White et al., 2020) and phenomena such as the "sentence superiority effect:" accuracy for identifying single words is higher when they appear in the context of a briefly flashed four-word sentence than in a string of words that do not form a sentence (Snell & Grainger, 2017).

Although we have no direct evidence of yet, other kinds of linguistic relations between letter strings – morphological, syntactic, and semantic – might facilitate parallel processing. However, there is one reason to be cautious before generalizing the hypothesis so far. Specifically, Brothers (2022) adapted the divided-attention paradigm (like used here) but for four-word sentences. Participants had to judge the *grammaticality* of two-word pairs on either the left or right side of fixation. The large drop in accuracy with divided attention supported the serial model, even with words embedded in sentence context. Thus, it may not be the case that all types of syntactic or grammatical structures support parallel processing of multiple words.

The second hypothesis we propose is the "unconscious parallel" hypothesis: two words are initially processed in parallel to the semantic level, even if they are unrelated to each other.

However, only one word is *consciously* perceived at a time. Due to the brief presentation and backwards masking in our experiments, participants are only aware of one of the two words presented per trial, and the tasks we employed tap into such conscious percepts. Nonetheless, relations between the two words can influence task performance. This could occur because of "cascaded processing" (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001): feedforward parallel processing across the visual field reaches a high level before any individual words (or perhaps even letters) are definitively identified (Wen, Snell, & Grainger, 2019). Conscious identification might occurs at a later stage, serially for one word at a time (Snell & Grainger, 2019a). This hypothesis allows that a participant in our compound word judgment task could recognize that the set of displayed letters form a semantic unit (a compound word) before being able to report what any of the individual constituent words were.

The key component of this hypothesis is that when two words form a compound word, the parallel processing allows them to *facilitate* one another. Thus, accuracy for reporting either one is increased. Such a facilitation between words is reminiscent of priming (Dehaene et al., 2001; Marcel, 1983; Neely, 1976). Priming usually occurs when one word that is presented first facilitates another word that is presented later, but similar effects can also occur when two words are presented simultaneously (Meyer & Schvaneveldt, 1971) or even with the prime appears after the target (Kiger & Glass, 1983; Logan & Schulkind, 2000). The spreading activation theory suggests that a network of semantically related concepts is activated by a 'prime' word (Collins & Loftus, 1975). Our hypothesis here is that something similar occurs for simultaneously processed words that can combine to refer to a single concept: each 'primes' the other, such that both are processed faster and are more robust to the post-masks.

According to the unconscious parallel hypothesis, there may be some level of parallel semantic processing even for a pair of words that are unrelated to each other. That parallel semantic activation is not revealed in task accuracy because such pairs of words do not facilitate each other, and only one can be consciously reported. Such a proposal is a stark departure from the simpler, serial-processing explanation of performance in such divided-attention word recognition tasks (e.g., White et al., 2020).

One challenge for both hypotheses is that accuracy was as high as predicted by the independent parallel model in the compound word task of Experiment 1, but accuracy was only modestly above the serial model's prediction for compound word pairs in Experiment 2. The "high-level bottleneck" hypothesis predicts the first finding, and the "unconscious parallel" hypothesis predicts the second, if it further assumes that parallel processing has limited capacity and the between-word facilitatory effects are rather weak. Both hypotheses have an easier time explaining both experiments if we further propose that explicit knowledge of the presence of compound words, and their task-relevance, strengthens the parallel or holistic processing of both constituents (especially when they are split up on opposite sides of fixation). That type of attentional effect would account for the higher accuracy in the compound word judgment task (Experiment 1) than for compound pairs in the single-word report task (Experiment 2).

*Limitations & Future Directions*

This study is the first to introduce semantic relations between words in the study of how well two words may be processed at once with divided attention. There are several limitations that may be addressed in the future. The first concerns the positions of the words above and below fixation. As explained in the Introduction, we chose those positions because prior studies found strong evidence for serial processing of two unrelated words positioned there, and because it allows for all the letters to be relatively close to fixation to maximize legibility of all of them. Presenting words to the left and right of fixation introduces an asymmetry, with words generally better processed in the right visual field (Mishkin & Forgays, 1952; Yeatman & White, 2021). Nonetheless, an important next step will be to investigate these questions with words arranged horizontally, more like nature English reading.

Second, it will be important to explore other forms of word relations beyond compound words. For instance, words may be related because of high co-occurrence frequencies in natural language or their proximity in semantic space (e.g., cows + horses). Words also combine to form meaning via syntactic relations, such as in two-word phrases (e.g., blue + sky, quiet + please, she + spoke). It remains to be explored how such other types of linguistic relations between words influence the ability to process them in parallel.

*Conclusion*

The debate about parallel vs. serial processing in reading often focuses on single words as discrete units, i.e., letter strings separated by spaces. The results reported here suggest a shift towards evaluating units of meaning that may encompass multiple letter strings (Cutter et al., 2014; Zang, 2019). Specifically, the data demonstrate that two words may be processed in parallel if they compose a compound word, even under the conditions when two unrelated words are processed serially. Thus, semantic relationships between words facilitate faster, more holistic parallel processing. That emergent parallel processing may be a key component of reading skill.

We suggest two potential explanations: 1) that the serial bottleneck can allow multiple words to be semantically processed simultaneously if they are recognized to be part of one linguistic structure or one unit of meaning; 2) that conscious perception is limited to one word at a time, but multiple words are fully processed in parallel in a manner that allows related words to facilitate each other. Further investigation is required to test these hypotheses.

## DATA AVAILABILITY

Upon publication, all raw data and analysis code will be made available on the Open Science Framework.

## REFERENCES

Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects? *Journal of Memory and Language*, *25*, 726–740.

Bonnel, A.-M., & Prinzmetal, W. (1998). Dividing attention between the color and the shape of objects. *Perception & Psychophysics*, *60*, 113–124.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 443–446.

Braun, J., & Julesz, B. (1998). Withdrawing attention at little or no cost: Detection and discrimination tasks. *Perception & Psychophysics*, *60*, 1–23.

Brothers, T. (2022). Capacity limits in sentence comprehension: Evidence from dual-task judgements and event-related potentials. *Cognition*, *225*, 105153.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.

Campbell, M., Oppenheimer, N., & White, A. L. (2024). Severe processing capacity limits for sub-lexical features of letter strings. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-023-02830-1

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, Vol. 108, pp. 204–256.

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2014). Preview benefit in english spaced compounds. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*, 1778–1786.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. Le, Mangin, J. F., Poline, J. B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, *4*, 752–758.

Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive Processes*, *22*, 953–1000.

Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, *65*, 406–430.

Johnson, M. L., Palmer, J., Moore, C. M., & Boynton, G. M. (2022). Evidence from partially valid cueing that words are processed serially. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-022-02230-w

Kiger, J. I., & Glass, A. L. (1983). The facilitation of lexical decisions by a prime occurring after the target. *Memory & Cognition*, *11*, 356–365.

Lee, D. K., Koch, C., & Braun, J. (1999). Attentional capacity is undifferentiated: Concurrent discrimination of form, color, and motion. *Perception & Psychophysics*, *61*, 1241–1255.

Legge, G. E., Mansfield, J. S., & Chung, S. T. L. (2001). Psychophysics of reading XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Research*, *41*, 725–743.

Libben, G., Gagné, C. L., & Dressler, W. (2020). The Representation and Processing of Compound Words. In V. Pirelli, I. Plag, & W. U. Dressler (Eds.), *Word Knowledge and Word Usage* (pp. 336–352). De Gruyter.

Logan, G. D., & Schulkind, M. D. (2000). Parallel memory retrieval in dual-task situations: I. Semantic memory. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1072–1090.

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, *15*, 197–237.

Medler, D. A., & Binder, J. R. (2005). MCWord: An on-Line orthographic database of the English language. Retrieved from http://www.neuro.mcw.edu/mcword/

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.

Mishkin, M., & Forgays, D. G. (1952). Word recognition as a function of retinal locus. *Journal of Experimental Psychology*, *43*, 43–48.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition*, *4*, 648–654.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Pollack, I., & Hsieh, R. (1969). Sampling variability of the area under the ROC-curve and of d'e. *Psychological Bulletin*, *71*, 161–173.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Attention to one word at a time in reading is still a viable hypothesis: Rejoinder to Inhoff, Radach, and Eiter (2006). *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1496–1500.

Sandra, D. (1990). On the Representation and Processing of Compound Words: Automatic Access to Constituent Morphemes Does Not Occur. *The Quarterly Journal of Experimental Psychology Section A*, *42*, 529–567.

Scharff, A., Palmer, J., & Moore, C. M. (2011). Extending the simultaneous-sequential paradigm to measure perceptual capacity for features and words. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 813–833.

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, and Psychophysics*, *74*, 5–35.

Snell, J., & Grainger, J. (2017). The sentence superiority effect revisited. *Cognition*, *168*, 217–221.

Snell, J., & Grainger, J. (2019a). Consciousness Is Not Key in the Serial-versus-Parallel Debate. *Trends in Cognitive Sciences*, Vol. 23, pp. 814–815. Elsevier Ltd.

Snell, J., & Grainger, J. (2019b). Readers Are Parallel Processors. *Trends in Cognitive Sciences*, *23*, 537–546.

Sperling, G., & Melchner, M. J. (1978). The Attention Operating Characteristic: Examples from Visual Search. *Science*, *202*, 315–318.

Taft, M., & Forster, K. I. (1976). Lexical Storage and Retrieval of Polysyllabic Words. *Journal of Verbal Learning and Verbal Behavior*, *15*, 607–620.

Torgesen, J., Rashotte, C., & Wagner, R. (1999). TOWRE-2: Test of Word Reading Efficiency, 2nd Ed.

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of*

*Experimental Psychology*, *67*, 1176–1190.

Veldre, A., Reichle, E. D., Yu, L., & Andrews, S. (2023). Lexical Processing Across the Visual Field. *Journal of Experimental Psychology: Human Perception and Performance*, *49*, 649–671.

Vidal, C., Content, A., & Chetail, F. (2017). BACS: The Brussels Artificial Character Sets for studies in cognitive psychology and neuroscience. *Behavior Research Methods*, *49*, 2093–2112.

Wen, Y., Snell, J., & Grainger, J. (2019). Parallel, cascaded, interactive processing of words during sentence reading. *Cognition*, *189*, 221–226.

White, A. L., Boynton, G. M., & Yeatman, J. D. (2019). You Can't Recognize Two Words Simultaneously. *Trends in Cognitive Sciences*, *23*, 812–814.

White, A. L., Palmer, J., & Boynton, G. M. (2018). Evidence of serial processing in visual word recognition. *Psychological Science*, *29*, 1062 –1071.

White, A. L., Palmer, J., & Boynton, G. M. (2020). Visual word recognition: Evidence for a serial bottleneck in lexical access. *Attention, Perception, and Psychophysics*, *82*, 2000–2017.

White, A. L., Palmer, J., Boynton, G. M., & Yeatman, J. D. (2019). Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. *Proceedings of the National Academy of Sciences*, *116*, 10087–10096.

Yeatman, J. D., & White, A. L. (2021). Reading: The Confluence of Vision and Language. *Annual Review of Vision Science*, *7*, 487–517.

Zang, C. (2019). New perspectives on serialism and parallelism in oculomotor control during reading: The multi-constituent unit hypothesis. *Vision*, *3*. https://doi.org/10.3390/vision3040050

Zang, C., Wang, S., Bai, X., Yan, G., & Liversedge, S. P. (2024). Parafoveal processing of Chinese four-character idioms and phrases in reading: Evidence for multi-constituent unit hypothesis. *Journal of Memory and Language*, *136*, 104508.

**SUPPLEMENTARY MATERIALS**

**for Anupindi, Eisler & White:** ***Serial processing of two words becomes parallel when they combine to form a compound word***

__**Experiment 1: Lexical decision task**__

**Methods:**

***Stimuli:*** The lists of all words used in all experiments are provided in **Tables S2-S6** at the end of this document.

***Eye tracking.*** Throughout the presentation of stimuli, we recorded the right eye's gaze position at 500 Hz with an Eyelink 1000+ video-based eye-tracker (SR Research). Fixation was established during the ITI at the start of each trial. The trial only advanced if the estimated gaze position was within 1.0º horizontally or 1.5º vertically from that fixation position. If a fixation break occurred between the pre-cue offset and post-mask offset, the trial was immediately terminated. The subject had to press a button to continue the next trial. Terminated trials were repeated at the end of the block, unless fewer than three trials remained.

To be sure that we included no trials in which subjects may have looked directly at a word, we also analyzed the eye traces offline. First, for each trial in a block, we computed the median gaze position (across measurement samples) in the 400ms before the pre-cue onset (excluding intervals for blinks). Then we defined the "central gaze position" for the block as the across-trial median of those initial gaze positions. This analysis corrects for any error in the eye-tracker calibration by assuming that subjects were fixating correctly in the interval before the pre-cue, when only the fixation mark was visible.

Then, for each trial, we analyzed gaze positions in the interval between the onset of the words and the offset of the post-masks. We defined an "offline fixation break" as a deviation that was more than 0.7º horizontally or 1.0º vertically from the central gaze position and that lasted more than 30ms. In the analysis, we excluded all trials with offline fixation breaks.

***Staircase.*** The staircase was run in blocks of 20 trials, alternating between the single-task above and below conditions (no dual-task trials in the staircase). During each run, the word-mask ISI in units of $\log_{10}$(seconds) was adjusted by a weighted 1-up/1-down staircase procedure controlled by the Palamedes toolbox (Prins & Kingdom, 2009). The step size down was always one-third of the step size up, which makes the staircase converge on the 75% correct threshold. Two staircases were randomly interleaved across trials, and blocks continued until both staircases had reversed direction ten times, and the threshold ISI was the mean value across all reversals.

***Analysis of Accuracy.*** For each condition of the lexical decision task, we used the participant's 4-level confidence ratings (from 'sure pseudoword' to 'sure real word') to construct a receiver operating characteristic (ROC). The ROC is a plot of hit rates (HR) as a function of false alarm rates (FR). "Targets" were real words at the post-cued location. To compute these rates from the participant's ratings, we varied an index i from 0 to 4. At each index level, we coded responses greater than i as "yes" responses. For each value of i, HR(i) is the proportion of "yes" responses on target-present trials and FR(i) is the proportion of "yes" responses on target-absent trials. For instance, when i = 3, only response ratings of 4 (highest confidence) on target-present trials are considered hits, and only response ratings of 4 on target-absent trials are considered false alarms. The five pairs of HR(i) and FR(i) trace out a curve, the area under which is a measure of accuracy: $A_g$.

***Bootstrapping***. Throughout the text, we report bootstrapped 95% confidence intervals (Cis) for average measurements. To compute these, we generated a distribution of 5,000 resampled means. Each of those is the mean of ten values sampled with replacement from the original set of ten subjects' means. The CI is the range from the 2.5$^{th}$ to the 97.5$^{th}$ percentile of the distribution of resampled means, with an "accelerated" bias correction (Efron, 1987).

***Bayes Factors***. Finally, we supplement our pairwise tests with Bayes Factors (BFs), which quantify strength of evidence. The BF is the ratio of the probability of the data under the alternate hypothesis (a distance is >0 or two conditions differ) relative to the probability of the data under the null hypothesis that there is no difference (Rouder, Speckman, Sun, Morey, & Iverson, 2009). A BF of 10 indicates that the data are 10 times more likely under the alternate hypothesis than the null. Typically, BFs between 1 and 3 are regarded as weak evidence for the alternate hypothesis, BFs between 3 and 10 as substantial evidence, and BFs between 10 and 100 as strong evidence. Conversely, BFs between 1/3 and 1/10 are considered substantial evidence for the null hypothesis, etc. (Kass & Raftery, 1995). We computed BFs for pairwise t-test ANOVAs using the bayesFactor toolbox by Bart Krekelberg (https://doi.org/10.5281/zenodo.4394422).

**Results:**

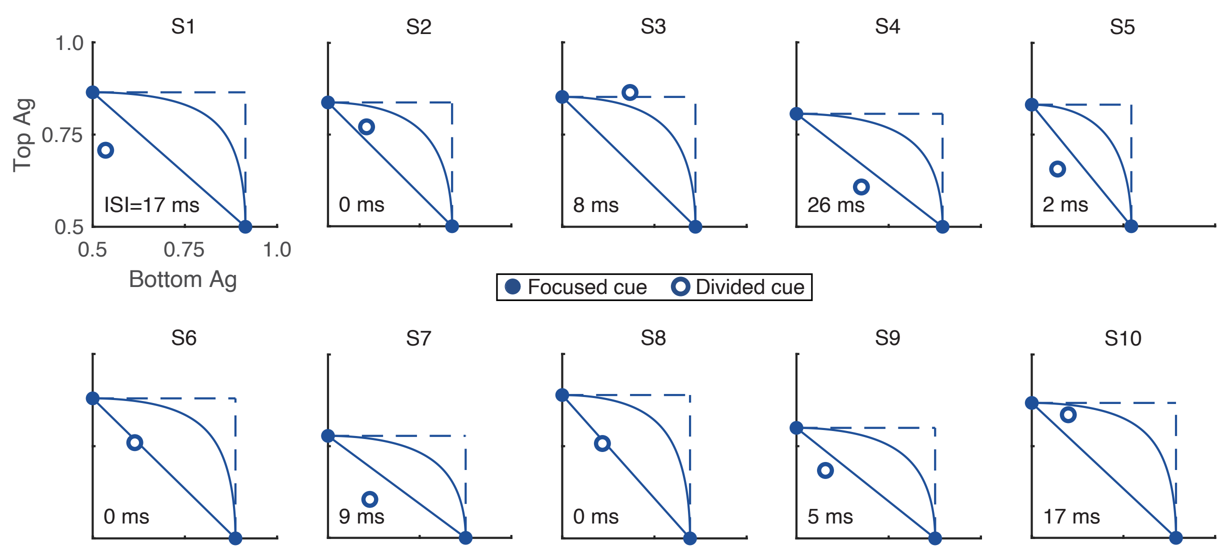


**Figure S1:** Individual participant attention operating characteristics in the lexical decision task of Experiment 1. Format as in Figure 2A. In each participant's graph, the number in the lower left corner is the mean duration of their threshold-level interstimulus interval (ISI) between the words and the post-masks.

## Experiment 1: Compound word judgment task

**Results:**

***Accuracy in units of d′: effects of ISI duration and word pair repetition***

To complement the analysis of proportion correct that is reported in the main text, we also calculate accuracy in units of *d′*, an unbounded and unbiased estimate of sensitivity - in this case, for detecting compound word pairs (Green & Swets, 1966). Here we separately analyze the trials when the words-to-mask ISI was set to each participant's threshold level, and the 32% of trials in which the ISI was lengthened to 400 ms. These trials were included to confirm that the task was easy when processing time was less constrained. Indeed, *d′* on these long-ISI trials was high: mean = 3.54 (SEM = 0.20), corresponding to 95% correct on average. *d′* was also significantly above chance (0 for this measure) on trials with the threshold-level ISI (mean = 1.58, SEM = 0.231; t(9) = 6.38, p = 0.0001; BF = 231).

Here, we also address the potential concern that the repetition of word pairs accounts for the above-chance compound word judgment accuracy. **Figure S2** plots *d′* separately for trials when each word pair appeared the participant for the first time, and trials when the word pair was repeated. That is done separately for trials with the threshold-level ISI trials and the 400 ms ISI. Word pair repetition had no effect on easy trials with 400 ms ISI (t<1, p=0.84, p=0.32). For the threshold-level trials, there was a significant improvement when the word pair repeated (mean increase = 0.56, SEM = 0.096; 95% CI = [0.399 0.778]; t(9) = 5.50, p = 0.0004; BF = 92). Importantly, however, *d′* was well above chance (*d′*=0) even on the 1st appearance trials (t(9) = 5.41, p = 0.0004; BF = 83). Thus, learning or inter-trial priming does not account for the high accuracy in the compound word judgment task.

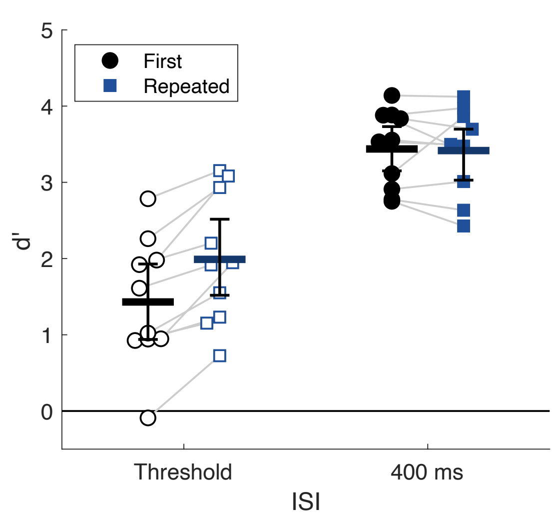


***Figure S2.*** Accuracy in units of *d′* for the compound word judgment task in Experiment 1. Each small circle or square represents one participant's *d′* in a particular condition. The x-axis separates trials in which the word-to-mask ISI was set to each participant's threshold from the lexical decision task (4 ms on average), or set to the "easy" level of 400 ms (which occurred on ⅓ of all trials). The black circles are for trials in which the word pair appeared for the first time in the experiment (70% of all trials), whereas the blue squares are for trials in which the word pair was repeated. The dark horizontal bars are the across-subject means, and within each ISI condition, the thin gray lines connect points from the same participant. In all four conditions represented here, *d′* was significantly above 0 (all *p*<0.001). Thus, even when the ISI was at threshold and a word pair appeared for the first time, accuracy exceeded the prediction of the serial model.

## Experiment 2

**Methods:**

***Stimuli.*** **Tables S5** and **S6** at the end of the Supplementary Materials list all the words used in the experiment. Words for the "random" set and the "compound" set were selected to have overlapping distributions of length, lexical frequency, orthographic neighborhood size, and concreteness ratings. The frequency and orthographic neighborhood size ratings were taken from MCWord (Medler & Binder, 2005). The concreteness ratings are from Brysbaert, Warriner, & Kuperman (2014). **Figure S3** below plots the distributions of those word properties.

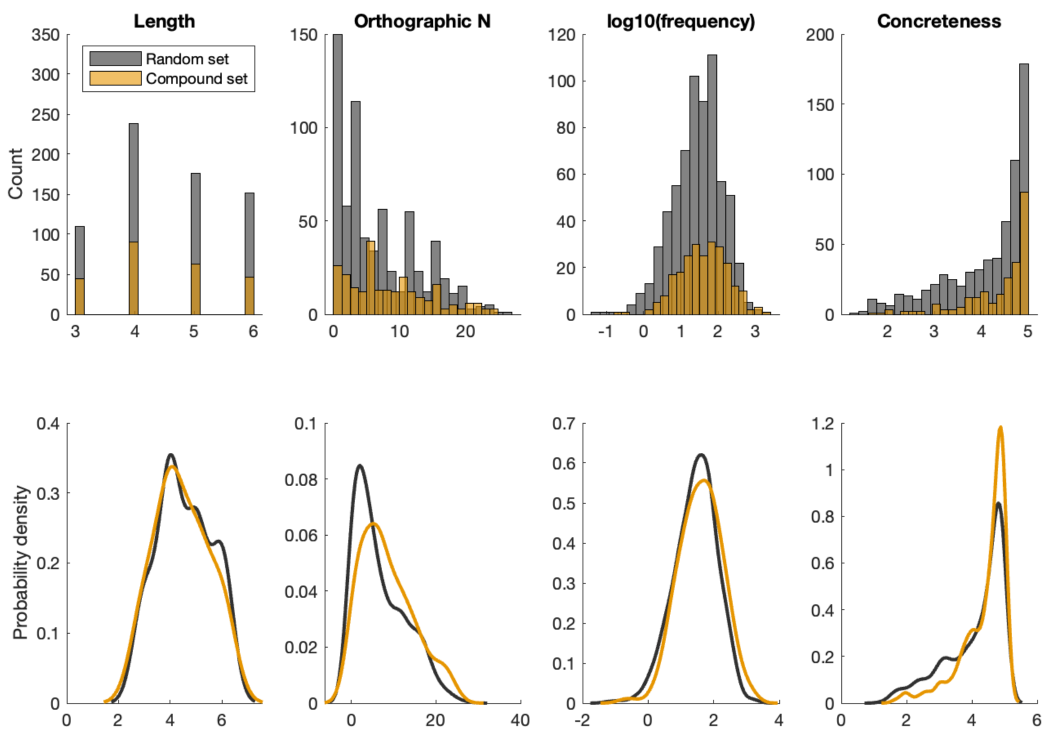


*Figure S3*: **Psycholinguistic properties of the words in Experiment 2.** The top row depicts histograms of word lengths, orthographic neighborhood size (N), lexical frequency in log units, and concreteness ratings. The overlapping distributions are for the set of words used in the "random pair" condition (in gray) and for the set of words that could form compounds (in gold). The latter set was used for the "scrambled", "reversed" and "compound" conditions. The bottom row depicts smoothed probability density distributions derived from the histograms above.

Because these distributions were not *perfectly* matched, we also analyzed the effects of these psycholinguistic word variables on accuracy in the task. We did so with general linear mixed-effect models fit to single-trial accuracy. Only orthographic neighborhood size (N) had a significant effect: words with greater N were *harder* to recognize. Notably, the mean N was higher in the compound set than the random set; thus, the small difference in mean N across the two stimulus sets cannot explain the higher accuracy for compound words. Furthermore, our key comparison in the main analysis is the comparison between the compound word trials and the scrambled compound trials, which used the exact same set.

### *Analysis of partial word accuracy*

An alternate metric of accuracy is the "edit distance" $E$ between the participant's typed response and the target word that was presented. $E$ is the smallest number of letter insertions, deletions, and substitutions needed to transform the reported string into the target string. We calculated $E$ with MATLAB's editDistance function. $E$ is biased to be larger for longer words, however, and there was a range of word lengths in our data set. To normalize $E$ by the length $L$ of each target word, we computed $E_{norm}$ on each trial as follows :

*If E<L*

$E_{norm} = (L-E)/L$

Otherwise,

$E_{norm} = 0.$

where $L$ is the number of letters in the target word. Thus, the mean $E_{norm}$ in each condition can be considered the average *proportion of letters* in the target words that were correctly reported. It ranges from 0 (total guess) to 1 (perfect response). This is why in the graphs of mean $E_{norm}$ below, the y-axes are labeled p(letters correct).

To build the AOCs and compare accuracy to the serial model prediction, we needed an estimate of what $E_{norm}$ would be when the participant has no information about the post-cued word and randomly guesses. To this end we conducted a simulation by collecting a set of words including all of the words in the stimulus set as well as all of the letter strings reported by all the participants in the actual experiment. Subsequently, the simulation randomly selected one of these words as its "response" to be used in the calculation of $E_{norm}$ for stimulated trials. On average, $D$ was 0.0729 on these simulated guessing trials. Thus, 0.0729 is the chance accuracy level that forms the origin of the AOCs shown in **Figure S4C.**

We analyzed $E_{norm}$ because it could be sensitive to partial information about the words that the participant perceived. In theory, even if the AOC computed from the strict binary measure of accuracy (p(correct)) shows a serial result, an AOC constructed from $E_{norm}$ could reveal some evidence of parallel processing of the two words – perhaps at a sub-lexical level.

**Results:**

The results of our analysis of $E_{norm}$, also labeled p(letters correct) are in **Figure S4**. The results are largely consistent with the main analysis of whole-word p(correct) in the main text Figure 4. In the divided attention condition, accuracy was higher for compound word pairs than scrambled compound pairs (Fig S4A-B). On the AOCs (Fig. S4C), accuracy for random word pairs was just slightly above the serial model's prediction but indistinguishable from it. Accuracy for compound pairs was significantly above the serial model's line, as it was for reversed compounds, but by roughly half as much. Accuracy for scrambled compounds was on average modestly above the serial model's line, but not significantly so. Statistics for these distances are listed in **Table S1**.

As a second alternate measure of accuracy, we scored each response as correct if the editDistance $E$ was less than or equal to 1, and incorrect otherwise. In other words, we allowed a response to be considered "correct" even if there was a mistake of one letter. We then analyzed these binary correct/incorrect data in the same way as the primary analyses of p(correct) in the main text. The results were essentially the same as what is shown in Figure 4.

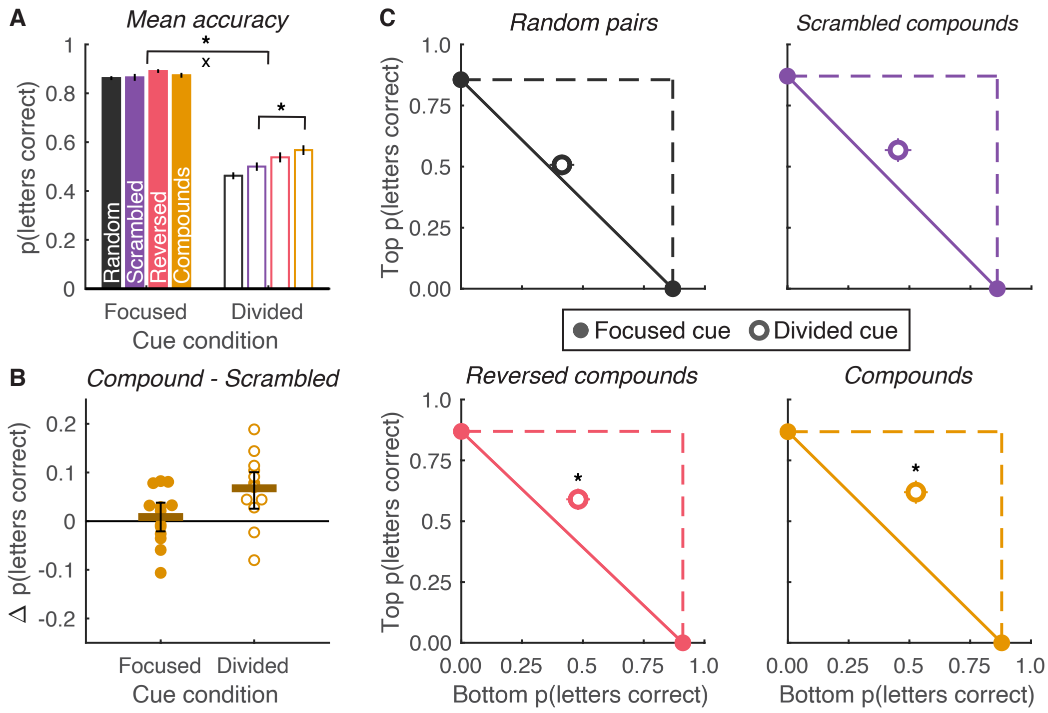


***Figure S4***: Analyses of accuracy calculated as $E_{norm}$, the mean proportion of letters within target words that were correctly reported. Format as in main text Figure 4.

| **Word pair type** | **Mean** | **SEM** | **t** | **FDR-corrected p** | **BF** | **95% CI** |
|---|---|---|---|---|---|---|
| *Random pairs* | -0.014 | 0.017 | -0.825 | 0.425 | 0.372 | [0.045 0.020] |
| *Compounds* | 0.138 | 0.025 | 5.275 | 0.001 | 160.6 | [0.09 0.188] |
| *Reversed* | 0.082 | 0.030 | 2.619 | 0.045 | 3.020 | [0.027 0.144] |
| *Scrambled* | 0.052 | 0.031 | 1.637 | 0.170 | 0.805 | 0.001 0.124] |

**Table S1:** Statistics on the mean distances from the serial model in Experiment 2, based on analyses of p(letters correct) (also called $E_{norm}$). BF = Bayes Factor; CI = confidence interval.

**SUPPLEMENTAL REFERENCES**

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911.

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, *82*, 171–185.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*, 773–795.

Medler, D. A., & Binder, J. R. (2005). MCWord: An on-Line orthographic database of the English language. Retrieved from http://www.neuro.mcw.edu/mcword/

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237.

***Table S2:*** *Real words used in Experiment 1*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| abuse | boat | court | fan | heel | lit | office | return | slave | tip |
| accept | body | cover | far | hero | little | oil | rice | sleep | tired |
| ache | bog | cow | fat | hey | load | old | rider | slide | title |
| across | bomb | craft | father | hid | lobby | older | rim | slow | toe |
| act | bone | creep | fed | higher | log | opened | ring | slowly | ton |
| action | bony | crept | fee | highly | longer | others | rip | sly | tool |
| acute | book | crew | few | hind | looked | ought | road | snake | tooth |
| adapt | boot | cross | fill | hint | lot | over | rob | snap | top |
| add | bottle | crow | finer | hip | loved | owe | rod | soap | town |
| afraid | bottom | crown | finger | hit | low | pact | roger | social | toy |
| after | bought | crude | fire | hive | lunch | pad | rome | sofa | trace |
| age | bow | crush | firm | hog | lust | paid | room | soft | trade |
| ago | bowl | cry | fish | hole | magic | pan | rope | sold | trail |
| agreed | box | cue | fit | holy | main | paper | rot | solve | trap |
| ail | boy | cult | fix | home | making | papers | rough | some | tray |
| aim | bra | cup | flag | honey | man | par | row | son | tread |
| air | brain | cut | flap | hood | map | pass | royal | sort | trip |
| alien | break | cycle | flies | hook | marsh | passed | rub | sound | trips |
| almost | bridge | dam | floor | hop | match | paste | rug | sow | truck |
| along | broken | dark | flow | horn | mate | pay | rum | space | truly |
| amid | brush | day | fly | horse | math | peer | run | spent | try |
| amount | bug | dead | foe | hot | matter | peg | rush | spider | trying |
| ample | bunk | death | fog | house | maze | pen | rusty | spoon | tub |
| anger | burger | degree | follow | hut | melon | people | rye | spy | tug |
| animal | burn | demand | food | ice | melt | pepper | sad | stack | type |
| ankle | bus | den | fool | icy | member | period | sail | stage | under |
| anyone | bush | dense | forced | idle | mere | pet | sake | stain | united |
| anyway | butter | desk | forces | ill | merely | petty | sale | stalk | urged |
| ape | buy | dew | former | income | mess | photo | sand | start | use |
| appear | cab | die | fox | indeed | met | pick | sat | state | van |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| areas | cake | diet | frank | ink | meter | pie | sauce | states | vast |
| arm | camp | dig | friend | inner | middle | pig | save | steak | vat |
| arose | candle | dim | frog | inside | mile | pile | saw | steal | visit |
| around | cannot | dimly | froze | italy | milk | pin | saying | steam | vote |
| arrow | cap | dinner | fry | itself | mill | pine | scare | steel | vow |
| art | car | dip | fuels | ivory | mint | pit | scene | step | waist |
| ash | card | direct | fun | jail | minute | places | scent | stick | wake |
| ask | care | dish | fur | jar | mix | plain | school | sting | walk |
| ate | carry | doctor | fuss | jaw | mob | plan | scoop | stone | wall |
| attack | cart | dog | gap | jelly | modern | plane | screen | stony | war |
| aware | case | doing | gaps | job | moment | plate | scrub | stool | wary |
| awe | cast | dot | garden | jug | money | play | sea | stop | washer |
| axe | castle | doubt | gas | junk | month | played | seeing | store | waste |
| baby | cat | dozen | gave | jury | months | plead | seemed | storm | watch |
| back | caught | drag | gay | keep | moon | please | seize | strap | water |
| bad | cave | draw | gift | key | moral | plus | series | stream | wave |
| bag | cease | dread | gin | kick | motel | pocket | serve | street | wax |
| ball | chance | dream | girl | kid | mother | police | set | string | way |
| ban | change | dried | giving | kind | motor | pony | sets | stroke | wear |
| band | chat | drift | glad | kit | mouse | pool | shade | suit | web |
| bar | check | drive | god | knight | move | pop | shall | sum | weight |
| base | cheese | dry | grand | lace | moved | pot | share | sun | wet |
| basket | chest | dug | grass | lady | moving | pounds | sharp | sunk | whale |
| bat | chief | dusk | grave | lamp | mud | power | sheet | super | wheat |
| bay | child | dwarf | grip | land | mug | print | shelf | sweet | whole |
| bean | chin | dye | ground | lap | myself | prize | shell | swim | width |
| beat | chip | dying | groups | laugh | nail | probe | ship | swung | wife |
| became | choice | ear | grove | law | naked | proof | shock | syrup | wig |
| become | chop | earl | growth | lawn | nap | pub | shoe | table | win |
| bed | church | earn | guard | lay | nasty | public | shone | tag | wind |
| bee | clerk | ease | guest | leap | nature | quick | short | tail | window |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| beech | cling | easily | gum | leapt | nearly | quit | shot | taken | wing |
| before | clock | eat | gut | learn | neat | radar | should | taking | wire |
| beg | closed | echo | guy | leaves | neck | rag | show | talk | wise |
| begun | cloth | egg | habit | led | net | raid | shut | tan | wish |
| behind | cloud | ego | hail | ledge | news | rail | shy | tap | wit |
| bent | club | either | hair | leg | nice | rain | sick | tape | within |
| berry | cod | elder | hall | lemon | night | ram | side | taste | woman |
| beside | code | elm | ham | length | nine | ran | simple | tax | won |
| better | coffee | end | handy | letter | nobody | rat | simply | tea | wood |
| beyond | color | enemy | happen | lever | nod | rate | sin | ten | word |
| bid | comb | energy | hardly | lid | noon | rather | single | tense | work |
| big | coming | enough | haste | lie | nor | raw | sink | thing | worked |
| bike | cone | envy | hat | life | note | read | sip | though | worm |
| bin | cool | era | hatch | lift | noted | really | sir | threw | wreck |
| bit | cop | eve | having | light | notice | reason | sit | thumb | wrist |
| black | corn | evil | hay | likely | number | recent | site | tidal | writer |
| blast | corner | exit | head | lily | nut | red | six | tide | wrong |
| blaze | cot | expect | heap | lime | nuts | reef | skill | tidy | yard |
| blood | count | eye | heart | line | oak | remain | skin | tie | yet |
| blue | couple | fall | heat | lined | ocean | report | sky | time | youth |
| board | course | family | hedge | lip | odd | rest | slab | tin | zoo |

***Table S3:*** *Pseudowords used in Experiment 1*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| abber | bant | covies | fid | haly | ler | ninds | rettle | ter | viead |
| aboks | baran | coys | figuge | hame | lere | nome | rew | terts | vig |
| abore | bas | cre | fike | han | leth | nop | rigat | tes | vight |
| abost | basee | cuce | fim | haner | levef | norked | rint | tew | vikel |
| aboth | bave | cun | fime | hao | lew | nox | rirse | thad | vite |
| aboue | bbi | dach | fimer | har | lex | noy | ris | thal | voo |
| abown | beald | dack | fimp | hasier | lich | nur | rith | thales | vorss |
| abrint | bealth | danded | finte | haso | lide | oas | romes | tham | wadin |
| abung | beange | dar | fis | hather | lifors | ocowd | roo | thap | waloy |
| acarle | beargh | das | fiter | hax | ligin | odo | ropt | thar | wam |
| accays | beart | dater | fith | hea | lilk | oftas | roren | thay | wat |
| accous | beay | datung | flang | heang | lill | ofted | rou | thed | waw |
| acrmen | beayed | decure | flers | hec | lim | oim | rown | theed | wead |
| acthes | bebond | deel | fod | hecker | lirst | ois | rur | theen | weang |
| actung | becore | demife | fom | hed | lis | oll | russ | thep | weem |
| ader | beew | derk | fomby | hee | lissow | ond | rutten | thepe | weet |
| adoply | beey | dery | fon | heem | lith | onok | sammer | thers | weng |
| afffal | beffed | detes | foo | hees | lito | onom | san | thes | wer |
| affond | beffur | dever | foso | heesp | lits | onor | sar | thest | werc |
| afrved | beick | dilled | fot | helt | lomer | onte | sarmed | thet | weso |
| aftat | beined | dir | fou | hene | lonch | onth | sas | thice | whair |
| aftes | beltow | dis | fouly | herx | lothic | onve | scets | thid | whan |
| aftio | bemer | dit | foung | het | lounke | ony | scoue | thild | wheam |
| agage | benain | doo | foused | hig | lourer | opoken | seach | thill | whees |
| agase | beosit | doon | fow | hil | loy | opom | sed | thind | whein |
| agele | beoual | doow | fown | hin | loys | opow | seese | thint | wheir |
| agybe | bepe | dor | foy | hix | mais | otere | seged | thioy | whem |
| aie | bephes | dould | fral | hmt | makled | othee | seld | thip | wheme |
| ais | ber | dow | frek | hody | mand | othel | sende | thir | whice |
| ait | berdy | drach | fren | hom | mapen | otial | ser | thist | whis |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| alary | bere | dreir | fres | homagh | marer | otich | sest | thit | widded |
| ald | bes | dren | frey | hong | marind | oub | sheal | thits | wiel |
| alere | bestet | drere | fron | hortes | marsed | oue | shedy | thok | wift |
| aleted | beved | drout | fror | hoys | mas | ougin | shen | thom | wike |
| alk | beving | drow | froust | huier | mave | oun | shice | thougs | wilf |
| almame | bew | druld | frow | hurlet | mee | ous | shich | tid | wim |
| almary | beyids | dur | frur | ict | meers | oush | shiled | tild | wimber |
| alst | bie | dut | fuich | ild | mem | ouy | sho | timy | wime |
| alugly | bim | eabing | fusted | ind | mench | owo | shoped | tir | wis |
| alwler | bir | ead | fut | inen | mer | pabil | shougs | tith | wist |
| amare | bis | eadat | futh | insing | mes | pable | sid | titin | wita |
| amouly | bodds | eame | gack | inth | mex | pantly | sig | tob | wito |
| andand | boe | ean | gan | irat | mey | pas | sigat | tobs | woift |
| andily | boiced | eand | gand | iskel | mil | pastle | sild | toick | woind |
| ane | boile | easeng | gandes | itates | mim | pathed | silld | toom | woits |
| aneder | bome | easut | gea | ite | mimply | pattor | sim | toow | wome |
| anione | bonton | eative | geaner | ity | mir | pava | slade | torx | wonch |
| anirse | bor | ecched | geants | jad | mommed | ped | slen | toso | wonnd |
| ank | bou | eduad | ged | jas | mont | phe | sler | tou | woong |
| anreng | boung | eet | ger | joo | moow | pim | smain | tre | wor |
| ansing | brars | ege | gew | jor | mork | pito | soan | trere | wost |
| anstey | brarse | elst | gide | jore | morker | plint | somar | trich | woule |
| antal | brity | ene | gith | jorld | mors | pommon | sor | tro | wouly |
| antang | bup | enougs | gixen | jou | mory | pon | sorms | troun | wouse |
| anyind | cany | enther | gixil | joue | mouds | ponch | sors | tuch | wover |
| anyion | cargls | eny | gofte | jow | mounge | ponger | souir | tud | wrom |
| apary | carsed | ereds | goms | juch | mour | por | sould | tuling | wrose |
| aps | cas | ers | gon | juck | mre | pould | soull | tullo | wrow |
| ara | catten | ert | gont | jull | mry | pount | sount | twa | yeang |
| araled | cax | evea | gop | jumer | mun | pown | spes | twere | yee |
| areape | ceened | evear | goss | keat | muns | printh | spi | ucil | yener |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| areest | che | evel | gou | kee | munt | proked | stalic | undan | yer |
| areond | chert | evenge | gould | ker | nade | proks | staven | upase | yeter |
| arey | chice | evese | goung | kig | neald | prost | stazy | upell | yex |
| arght | chinds | evew | gouth | kir | neane | prums | ster | upur | yod |
| arinte | chuder | facars | gow | kister | neb | puints | stild | usee | yoem |
| arom | cill | fand | graias | kith | ned | pur | stims | usey | yoll |
| aross | cinal | fanter | grald | lais | neen | puv | stoor | ush | yond |
| ars | clant | fap | grers | lalan | nees | quost | sull | usk | yont |
| ary | cleagy | farine | groid | lan | neey | rald | sune | vajos | yop |
| asrune | clunts | fas | grore | langhs | neft | rall | sur | vas | yorty |
| asuly | clure | fea | grores | lant | neing | rane | sut | vean | yot |
| atreng | colies | feem | grult | lar | neith | recers | sweal | veat | yove |
| atries | coneer | fene | gye | las | nen | recoum | taght | vede | yown |
| ats | coout | fer | hab | leara | nener | reem | tanin | veee | |
| auring | cordal | fere | hach | leary | ner | reess | targe | veey | |
| awach | corld | fes | hais | lel | nern | reford | tas | vell | |
| awat | couble | fese | hald | len | nes | ren | tast | vely | |
| bame | couge | fet | haldes | leoped | nex | rer | teawn | vene | |

*__Table S4:__ Compound words used in Experiment 1*

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| after | math | | clock | wise | | hay | wire | | night | fall | | stop | watch |
| after | noon | | corn | stalk | | heart | beat | | note | book | | stream | line |
| after | taste | | court | house | | heat | stroke | | note | card | | suit | case |
| air | craft | | cow | boy | | hedge | hog | | nut | job | | sun | burn |
| air | plane | | cow | girl | | home | town | | over | flow | | sun | screen |
| arm | pit | | cross | road | | home | work | | over | weight | | sun | tan |
| arrow | head | | cross | word | | honey | comb | | paper | cut | | super | hero |
| ash | tray | | cry | baby | | horse | play | | pass | code | | sweet | heart |
| back | ground | | cup | cake | | horse | power | | pay | check | | swim | suit |
| back | stage | | day | break | | hot | pot | | pepper | mint | | table | cloth |
| back | yard | | day | dream | | house | wife | | pick | pocket | | tail | bone |
| band | width | | day | light | | ink | pot | | pig | pen | | tape | worm |
| basket | ball | | dead | line | | jail | break | | pine | cone | | tea | cup |
| bed | bug | | desk | top | | jelly | bean | | play | mate | | thumb | nail |
| bed | room | | dish | cloth | | junk | yard | | police | woman | | tide | pool |
| bed | sheet | | dish | washer | | keep | sake | | pop | corn | | time | share |
| bee | hive | | dog | house | | key | hole | | rail | road | | tip | toe |
| blood | stain | | draw | bridge | | key | stone | | rain | boot | | tool | box |
| blue | berry | | draw | string | | knight | hood | | rain | bow | | tool | kit |
| blue | print | | drift | wood | | lady | bug | | row | boat | | tooth | brush |
| board | walk | | drive | way | | lamp | shade | | sail | boat | | tooth | paste |
| body | guard | | ear | ring | | land | fill | | sand | castle | | trail | head |
| book | shelf | | egg | shell | | land | slide | | sauce | pan | | truck | load |
| book | store | | eye | lid | | lap | dog | | scare | crow | | type | writer |
| boot | strap | | fan | base | | law | suit | | school | work | | under | cover |
| bottle | neck | | finger | nail | | lay | man | | sea | shell | | under | weight |
| boy | friend | | fire | ball | | leap | frog | | ship | wreck | | waist | band |
| brain | storm | | fire | fly | | life | guard | | shoe | lace | | wall | paper |
| bus | boy | | fish | bowl | | lime | stone | | show | room | | water | color |
| butter | fly | | fish | hook | | lip | stick | | side | kick | | water | melon |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| butter | milk | | fog | horn | | lunch | bag | | sky | light | | wave | length |
| candle | stick | | friend | ship | | main | stream | | snake | skin | | web | cast |
| car | pool | | girl | friend | | match | box | | soap | box | | web | site |
| cat | fish | | grand | child | | mile | stone | | some | thing | | wheat | grass |
| cave | man | | grave | yard | | motor | cycle | | sound | proof | | wind | mill |
| cease | fire | | gum | ball | | mouse | trap | | sound | wave | | wish | bone |
| cheese | burger | | hair | cut | | mug | shot | | spider | web | | wrist | band |
| cheese | cake | | ham | burger | | nap | time | | steam | boat | | | |
| chin | rest | | hay | stack | | news | letter | | step | stool | | | |

***Table S5:*** *Compound words used in Experiment 2.*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| air | craft | | door | step | | law | suit | | sauce | pan |
| arm | chair | | dragon | fly | | leap | year | | scare | crow |
| ash | tray | | draw | bridge | | lunch | box | | school | bag |
| back | ground | | drive | way | | main | stream | | screw | driver |
| basket | ball | | earth | quake | | mile | stone | | ship | wreck |
| bee | hive | | egg | shell | | motor | cycle | | shoe | lace |
| blood | stain | | eye | lid | | mouse | trap | | shop | keeper |
| blue | berry | | fan | base | | mug | shot | | silver | ware |
| board | walk | | finger | nail | | news | flash | | sky | rocket |
| body | guard | | fisher | man | | night | fall | | slide | show |
| book | shelf | | fog | horn | | note | card | | sound | proof |
| bottle | neck | | girl | friend | | paper | cut | | spider | web |
| brain | wash | | grand | mother | | pass | code | | spread | sheet |
| butter | milk | | grass | hopper | | pay | check | | summer | time |
| candle | stick | | grave | yard | | pepper | spray | | sun | shine |
| car | pool | | hair | brush | | pick | pocket | | super | charge |
| cat | fish | | hay | stacks | | pig | pen | | sweet | heart |
| cheese | burger | | heat | stroke | | pigeon | hole | | table | cloth |
| chin | rest | | hedge | hog | | pillow | case | | tail | bone |
| cloak | room | | home | town | | pine | cone | | tape | worm |
| clock | wise | | honey | comb | | play | mate | | tax | payers |
| copper | head | | hop | scotch | | police | woman | | tea | cup |
| court | house | | horse | power | | pop | corn | | thumb | prints |
| cover | slip | | jelly | bean | | praise | worthy | | tooth | paste |
| cow | boy | | kick | boxing | | rail | road | | type | writer |
| cry | babies | | knight | hood | | rain | boot | | waist | band |
| day | break | | lady | bug | | rattle | snake | | water | color |
| dead | line | | lamp | shade | | ring | leader | | wave | length |
| desk | top | | land | fill | | sail | boat | | work | flow |
| dish | washer | | lap | dog | | sand | castle | | rib | cage |
| | | | | | | | | | pot | pie |

***Table S6:*** *Words used in the "random pair" condition in Experiment 2.*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| act | cherry | end | icicle | mood | queen | snow | truth | net | port | wasp |
| actor | chess | energy | idea | moon | quiet | soap | turkey | pet | soil | zest |
| advice | chest | engine | income | motion | quill | sock | turn | aid | taxi | bank |
| affair | child | entry | injury | mouth | quilt | soda | twig | bar | text | bill |
| agency | church | error | ink | movie | rabbit | sofa | twist | bow | tour | film |
| alarm | circle | estate | insect | mud | rake | song | two | ego | urge | fire |
| amount | city | event | iron | muscle | range | sort | uncle | fig | acres | firm |
| anger | clam | exam | island | music | ratio | soup | union | fox | beard | half |
| angle | class | expert | jam | name | reason | space | unit | inn | birds | hand |
| animal | client | extent | jar | nation | recess | spark | user | job | fruit | pain |
| apple | clover | face | jeans | nature | recipe | speech | value | lad | image | park |
| arch | club | fact | jewel | need | record | sponge | van | par | money | path |
| area | coach | family | judge | needle | region | spoon | vase | row | nurse | plan |
| art | coal | farm | jump | nerve | reward | spot | vein | set | speed | site |
| aspect | coast | farmer | kettle | noise | rhythm | spring | verse | sum | woods | star |
| attack | coffee | fear | key | north | rice | square | vessel | tip | acid | task |
| aunt | coil | field | king | nose | riddle | stage | video | dot | age | |
| baby | collar | flag | kiss | number | river | stamp | view | fee | bulk | |
| badge | cook | flame | knife | nut | robin | start | virus | fur | coat | |
| bait | cookie | flavor | knot | ocean | rock | steak | voice | gas | crop | |
| basin | cough | flesh | lab | offer | role | steam | volume | gig | drug | |
| basis | county | flight | ladder | office | roll | steel | voyage | gut | eggs | |
| bath | cousin | flock | lake | oil | roof | stem | wall | ivy | four | |
| battle | crate | floor | laugh | orange | root | stew | war | mat | pair | |
| bear | crayon | flower | leg | order | rose | stitch | waste | wit | post | |
| bed | cream | fold | letter | oven | route | stop | watch | ape | rent | |
| beef | credit | food | level | owner | rule | store | waves | bum | wage | |
| bell | crib | foot | limit | page | run | story | wealth | cot | ant | |
| bike | crime | force | lip | pail | safety | straw | week | cue | bud | |
| bird | crook | fork | liquid | parcel | salad | street | weight | dew | cab | |
| bite | crowd | form | loaf | part | salt | string | wheel | eel | can | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| blade | crown | fowl | locket | party | sample | studio | wife | jug | dip | |
| blow | cub | frame | look | peace | scale | system | window | rag | dye | |
| bomb | curve | frog | loss | pear | scarf | tale | wine | spa | elk | |
| bonus | data | front | love | pencil | scene | talk | wing | tan | eon | |
| border | dealer | fuel | low | people | scent | tank | winner | tub | era | |
| brake | death | gate | lumber | person | seat | taste | winter | beet | foes | |
| branch | debt | geese | magic | pest | sector | team | wire | rack | gaps | |
| brass | deer | gene | mall | phone | seed | teeth | wish | date | gem | |
| bread | degree | glass | map | photo | self | temper | wood | feet | gems | |
| breath | depth | glove | marble | piano | sense | tennis | wool | game | ham | |
| brick | design | goal | mark | pickle | shake | tent | word | leak | hub | |
| bubble | desire | grade | market | pin | shape | test | worker | wild | jaws | |
| bucket | detail | grain | mask | pipe | sheep | thanks | world | yolk | jig | |
| bulb | device | grape | mass | pizza | shirt | theory | wound | fail | lips | |
| burst | dime | grip | match | place | shock | thing | wrench | fund | ode | |
| buyer | dinner | group | math | plane | side | thread | wrist | origin | owls | |
| cactus | dirt | growth | meal | plant | sign | thrill | yak | seven | doe | |
| cake | disk | guest | meat | plate | silk | throat | yarn | spirit | fob | |
| camera | dock | guitar | media | player | singer | throne | yoke | hue | kin | |
| camp | doll | gum | member | plot | sink | ticket | youth | hug | mob | |
| cannon | donkey | gun | memory | plough | sir | tiger | zebra | kid | pat | |
| canvas | drain | hall | menu | poetry | sister | tin | zinc | nap | pub | |
| cap | drama | hammer | metal | point | size | title | zipper | pod | tad | |
| care | drawer | harbor | method | poison | skate | toad | zoo | pup | zip | |
| cart | dress | hate | mice | policy | skill | toe | bus | tar | gang | |
| cast | drink | health | mind | porter | skin | tongue | fun | beds | jolt | |
| cattle | drop | height | mine | potato | skirt | topic | hat | dare | lawn | |
| cause | duck | hill | mint | powder | sleep | touch | hip | deck | loft | |
| cave | dust | hook | minute | price | slope | toy | ice | duke | pony | |
| cell | ear | hope | mist | profit | smash | trade | jet | gold | puck | |
| cellar | edge | hose | mitten | prose | smell | trail | sea | golf | roar | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| cent | editor | hot | mode | pull | smile | train | tie | knee | seal | |
| chance | effect | hotel | moment | pump | smoke | tree | bun | list | solo | |
| change | effort | hour | monkey | push | snail | trick | kit | lord | tusk | |
| cheek | elbow | humor | month | quartz | sneeze | trip | log | peak | vibe | |