In the continuous-outcome source task, items can be similar in terms of the presentation context, either in the temporal domain (serial position in the presentation list) or in the spatial domain (angular distance between presentation angles). The items can also be similar in terms of features of the items themselves, and because we used words as stimuli, we identified the orthography and semantics of words as the most obvious relationships between words. A natural intuition is that increased similarity of any kind should result in more confusions between target and non-targets. Contrary to this, in Zhou et al. (2022), we found that neither semantic nor orthographic similarity between target and non-target words affected the probability of the non-target word intruding in the source retrieval task, although spatial and temporal similarity did.

One explanation for this null result lies particular words used as stimuli in Zhou et al. (2022). Word lists were constructed without regard to the semantics or orthography, so word pairs with high orthographic or semantic similarity were uncommon. As a result, any effect of item similarity may have been too small to exert a noticeable influence on intrusions, being dominated by the spatiotemporal similarity. If this is the case, then we should expect to see stronger evidence for item similarity when word lists are specifically constructed to maximize the orthographic or semantic similarity across the list. If, on the other hand, orthography and semantics truly don’t matter, then manipulating the similarity of items on each list should have no effect on intrusion gradient, which will continue to be determined by spatiotemporal similarity.

# Method

## Participants

Ten participants were recruited using Prolific, an online participant recruitment platform. For each session they completed, participants were paid at a rate of 6.50 GBP/hour. Nine of the ten participants completed ten sessions, while the remaining participant completed nine sessions. Participants were provided with plain language statements and consent forms and gave informed consent prior to the start of the first session of the experiment.

## Stimuli and Apparatus

The experiment was run online and presented in the browsers of participants’ computers. Software written in JavaScript using the jsPsych library (de Leeuw, 2015) controlled stimulus presentation and recorded responses. Participants were instructed to use the same display between sessions, and to keep the browser in fullscreen mode for the duration of each session, so that while hardware varied across participants, the experimental conditions for each participant was consistent across sessions. Words were displayed in 24 point Courier New white font positioned in the center of a uniform gray field. The use of a monospaced font and the restriction to six letters ensured that stimuli always occupied the same amount of space relative to the size of the screen.

Stimuli were six-letter words drawn from the SUBTLEXus database (Brysbaert & New, 2009). Word frequencies ranged from 10 and 500, which represents the number of times per million each word appears in the corpus of 51 million words. From this pool of words, study lists were constructed according to one of three experimental conditions. In the orthographic condition, lists of words were chosen that minimized the Damerau-Levenshtein distance between all the words in the list. The Damerau-Levenshtein distance is a measure of the minimum number of substitutions of single letters or transpositions of two adjacent letters needed to transform one word into another (because all words were six letters long, insertion or deletion of letters was not applicable). To compromise between the inter-list similarity, the number of items on each list, and the total number of lists, the maximum allowable Damerau-Levenshtein distance between any pair of words in each list was three edits. In total, there were 14 lists of 16 orthographically similar words.

To construct the stimuli for the semantic condition, we filtered out all words that appear on the orthographic lists from the word pool, and then followed a similar process that maximized inter-list semantic similarity from the remaining words. In contrast to free association approaches used to construct lists of words with critical lures (such as the DRM paradigm) which maximize the similarity between one unstudied word (the critical lure) and the rest of the list, we required all pairwise relationships between words on the same list to be above a threshold level of semantic similarity. To achieve this, we used vector representations of each word, with each vector representing 300 internal dimensions, obtained from a *word2vec* model that was pre-trained on multiple corpora of natural text (Mikolov et al., 2017)[[1]](#footnote-1). Semantic similarity was then defined as the cosine similarity between these vector representations. On a scale between 0, which means entirely orthogonal vectors (i.e. minimal semantic similarity) to 1, which means identical vector representations (i.e. maximal semantic similarity), all pairwise associations between words in the semantic condition were above 0.3. As with the orthographic lists, there were a total of 14 lists of 16 semantically related words that met these criteria. The code used to filter and construct the word lists are provided [REPO LINK], and the word lists themselves are provided as supplementary material.

## Procedure

Participants completed the experimental tasks over a maximum of 10 sessions. Each session consisted of 15 blocks, and each block consisted of 8 trials. There were a further 5 practice trials at the beginning of each session, the data from which was not included for analysis. We first describe the list manipulation at the level of the block, and then describe the structure of each individual trial. The composition of the word list was manipulated across blocks according to three conditions: 1) an orthographic condition where all words were drawn from the same orthographically related list, 2) a semantic condition where words were drawn from the same semantically related list, and 3) an unrelated condition where words were selected without constraint on their semantic or orthographic relationships. There were an equal number of blocks belonging to each condition, 5 each for a total of 15 blocks. At the beginning of each session, 5 orthographic and 5 semantic word lists were assigned to the orthographic and semantic blocks respectively.

In the orthographic condition blocks, all 8 trials used words drawn from the same preconstructed orthographic list. Similarly, words in blocks in the semantic condition were all drawn from the same semantic list. For the unrelated condition, words were drawn from across from the 18 word lists that were not used in the orthographic and semantic blocks, so that items on the same preconstructed list would not appear together. While incidentally similar pairs of words sometimes occurred, the average orthographic and semantic similarity of words in the unrelated condition was lower than the respective conditions. Figure 1 shows the pairwise orthographic distance and semantic similarity between words in each condition across the entire dataset.

Figure 1. Comparison of occurrences of pairwise orthographic (top) and semantic (bottom) similarity in orthographic, semantic, and unrelate

Chart, histogram

Description automatically generated

To begin each trial, participants were presented with a black marker positioned on a randomly generated angle on the outline of a circle, as well as a word positioned at the same angle as the marker, offset by a longer radius. The precise location of the word relative to the marker was determined by the sector the angle was in, with the word being offset to one of eight points on the bounds of the text box, corresponding to the middle of each of the four sides, and the four corners (i.e. in the North sector, the anchor was the bottom middle of the text box, while in the Northeast sector the anchor was the bottom left of the text box). The stimulus display remained visible for 1000 ms. Once the stimulus display time had elapsed, to ensure that participants attended to the source information, they were instructed to indicate the previous location of the cross on the blank target circle using a computer mouse. Responses made within π/8 radians of the true target location were classified as attended and advanced participants to the next item. There was no time limit for this response. Responses further away were deemed unattended and the words “TOO DISTANT” was displayed for 1000 ms, then the location was then re-presented and the verification task was repeated.

After studying each of the items for that block, participants were then instructed to complete a distractor task, which involved 30 seconds of arithmetic problems. These problems were presented as three single digit integers, which summed to a fourth number which would either be the correct sum, or a number that was one higher or lower than the actual sum. Participants indicated if the sum was correct by pressing the keys 0 (false) or 1 (true).

In the recognition task, the studied words (old words) were mixed with an equal number of unstudied foils (new words). When presented with each word, participants indicated whether the item was studied or unstudied using a 6-point confidence scale using the keys from 1-3 and 8-0 on the keyboard where 1 represented certainty that the item was new, and 0 represented certainty that the item was old.

Finally, in the source memory retrieval task, participants were cued with the words for 1500 ms, and then indicated the recalled location by a moving the mouse from the starting point in the center of the circle to a point on the circumference of the response circle. Response time was measured from the first movement of the mouse beyond a calibration marker, which was a circle with a radius of 8 pixels in the center of the screen. The cursor was required to be centered on this calibration marker to begin each trial. There was no time limit on the decision task.

1. Pretrained models were obtained from the fasttext.cc website, which were trained on the meta pages archive of English Wikipedia from June 2017, resulting in a text corpus of over 9 billion words in addition to news sources from statmt.org from 2007 - 2016, as described by Mikolov et al. (2017). Word2vec is an example of a semantic space model (see Jones et al., 2015 for a review) which differs from traditional approaches in that it relies on prediction during the course of training and negative sampling (Johns et al., 2019) and has been found to outperform models such as latent semantic analysis (LSA; Landauer & Dumais, 1997) in accounting for lexical decision latencies (Mandera et al., 2017). [↑](#footnote-ref-1)