UNIVERSITYᴼF BIRMINGHAM

University of Birmingham
Research at Birmingham

Remembering spatial words

Kumcu, Alper; Thompson, Robin L.

Document Version
Peer reviewed version

[Link to publication on Research at Birmingham portal](#)

Remembering Spatial Words: Sensorimotor Simulation Affects Verbal Recognition Memory

Alper Kumcu[1] and Robin L. Thompson[2]

[1] Department of Translation and Interpreting, Hacettepe University

[2] School of Psychology, University of Birmingham

Author Note

Correspondence concerning this article should be addressed to Alper Kumcu, Department of Translation and Interpreting, Hacettepe University, Turkey, 06800.

Email: alperkumcu@hacettepe.edu.tr, ORCID: https://orcid.org/0000-0003-0844-3562

Word Count: 12,338 (including abstract)

Abstract Word Count: 248

Abstract

Previous evidence shows that words with implicit spatial meaning or metaphorical spatial associations are perceptually simulated and can guide attention to associated locations (e.g., *bird* – upward location). In turn, simulated representations interfere with visual perception at an associated location. The present study investigates the effect of spatial associations on short-term verbal recognition memory to disambiguate between modal and amodal accounts of spatial interference effects across two experiments. Participants in both experiments encoded words presented in congruent and incongruent locations. Congruent and incongruent locations were based on an independent norming task. In Experiment 1, an auditorily presented word probed participants' memory as they were visually cued to either the original location of the probe word or a diagonal location at retrieval. In Experiment 2, there was no cue at retrieval but a neutral encoding condition in which words normed to central locations were shown. Results show that spatial associations affected memory performance although spatial information was neither relevant nor necessary for successful retrieval: Words in Experiment 1 were retrieved more accurately when there was a visual cue in the congruent location at retrieval but only if they were encoded in a non-canonical position. A visual cue in the congruent location slowed down memory performance when retrieving highly imageable words. With no cue at retrieval (Experiment 2), participants were better at remembering spatially congruent words as opposed to neutral words. Results provide evidence in support of sensorimotor simulation in verbal memory and a perceptual competition account of spatial interference effect.

Keywords: sensorimotor simulation, conceptual cueing, spatial interference, compatibility effect, recognition memory, grounded-embodied cognition

Remembering Spatial Words: Sensorimotor Simulation Affects Verbal Recognition Memory

In contrast to propositional and computational theories (Chomsky, 1957; Fodor, 1975; Pylyshyn, 1986) that propose abstract, amodal, and arbitrary symbols; grounded-embodied cognition theories assume that there are no strict delimitations between perception and cognition and that mental representations are sensorimotor in nature (Barsalou, 1999, 2008; Paivio, 1986; Pulvermüller, 1999). A grounded-embodied treatment of language suggests that upon reading or hearing a word, traces of previous perceptual experiences (e.g., visual, auditory, olfactory, gustatory or tactile) related to the meaning of the word are automatically reactivated and give rise to a *sensorimotor simulation* of its referent (Bergen, 2016; Zwaan & Madden, 2005; Zwaan & Radvansky, 1998). For example, reading a spatial word, that is a word with a spatial association such as *bird* which elicits a mental image, can give rise to a perceptual experience of an upward location based on the typical location of birds in the physical world.

Several studies have shown the involvement of sensorimotor simulations in language comprehension in perceptual tasks (Bergen et al., 2007; Estes et al., 2008, 2015; Ostarek & Vigliocco, 2017; Richardson et al., 2003; Stanfield & Zwaan, 2001; Zwaan et al., 2002; Zwaan & Yaxley, 2003a). For instance, in Zwaan and Yaxley (2003a), participants identified whether spatial words such as *root* and *branch* are semantically related. Results showed faster and more accurate reactions when words were presented in their canonical locations (e.g., *root* was located under *branch)* compared to a non-canonical arrangement (e.g., *branch* under *root*) (see also Šetić & Domijan, 2007). In a follow-up study, Zwaan and Yaxley (2003a) replicated the processing advantage with spatially congruent presentation but only when words were presented to the left visual field that corresponds to the right hemisphere, which is responsible for storing and processing visuospatial relations. However, the interaction between match effect and visual field was not replicated when the response side (i.e., right-handed vs. left-handed responses) was counterbalanced between participants (Berndt et al., 2019).

Emotionally charged words with positive or negative valence such as *hero* or *liar* similarly evoke sensorimotor simulations based on the metaphorical relationship between valence and vertical space (i.e., GOOD IS UP" and "BAD IS DOWN") (Lakoff & Johnson, 1980). These sensorimotor simulations are reflected in faster reaction times (RTs) in semantic judgment tasks when there is an alignment between the word's metaphorical location and its actual position on the screen even though neither location is relevant for a successful semantic judgement. For example, Meyer and Robinson (2004) demonstrated that positive words are evaluated faster (i.e., "rate *hero* on a scale from 1 – negative to 5 – positive") if they are presented at the top of the screen and negative words (e.g., *liar*) are evaluated faster if they are presented at the bottom (see also Chasteen et al., 2010; Giessner & Schubert, 2007; Leitan et al., 2015; Meyer & Robinson, 2004; Schubert, 2005; Taylor et al., 2015; Zanolie et al., 2012). Words with spatial associations can also guide visual attention to an associated location (i.e., a *conceptual cueing effect*). For instance, words such as *bird* or *hero* can orient attention upward, whereas words such as *root* or *liar* can orient attention downward (Amer et al., 2017; Dudschig et al., 2013; Estes et al., 2015; Gozli et al., 2016). Such attentional bias has been found to interfere with visual perception at an associated location and thus; counterintuitively, hinders identification of visual targets. To illustrate, *bird* elicits slower identification of a visual target presented at the top of a display than at the bottom of the display. Several studies have demonstrated this phenomenon, known as a *spatial interference effect* or a *mismatch advantage* (Bergen et al., 2007; Estes et al., 2008; Gozli et al., 2013; Richardson et al., 2003; Verges & Duffy, 2009).

Although a spatial interference effect has been reliably reported (see Estes & Barsalou, 2018 for a comprehensive meta-analysis), there are different theories posited to explain the phenomenon. Below we discuss both the grounded-embodied point of view and the memory retrieval point of view. From a grounded-embodied point of view, a spatial interference effect has been attributed to perceptual competition between the spatial representation of a word triggered by sensorimotor simulation and an

actual visual cue. The result of this competition is slowed perception in the overlapping location (see

Quadflieg et al., 2011). Alternative accounts are mostly based on *spatial indexing* or *spatial coding*

(Pylyshyn, 1989) and assume that word locations on the computer screen are encoded in an amodal

fashion (i.e., without the perceptual simulation) along with the words themselves (see Spivey et al.,

2004 for a review). In support of spatial coding account against modal simulation in semantic

judgement, Estes et al. (2015) found that words with strong (e.g., *sky*) and weak (e.g., *flourish*) visual

strength and imageability elicit a similarly strong interference effect. In memory retrieval tasks, spatial

indexing is assumed to account for a plethora of studies showing a *compatibility effect* (see Wynn et al.,

2019 for a review). That is, a match between an initial location where information is presented and a

later location where participants look to retrieve information improves recognition. For example,

participants in Scholz et al. (2016) listened to four factual statements presented with a loudspeaker

image in one of the four spatial areas on a two by two grid. Another statement probed participants'

recognition memory and during retrieval, with gaze behaviour manipulated by a visual cue appearing

either in a congruent location or a diagonal location with the original location of the loudspeaker image

associated with the probed information. Results showed that participants who were visually cued to the

congruent location were more successful in retrieving the information from memory despite the spatial

information being completely irrelevant for successful completion of the task (see also Johansson &

Johansson, 2014). These findings suggest that coding locations of information leads to a compatibility

effect and thus, a type of *match advantage* in memory, in contrast to spatial interference due to

conceptual cueing as described above. Hence, the alternative account from the memory point of view

predicts that target objects are coded not only for their location but also for their match with the

location of the cue word: if the physical location matches, then memory is improved and if it does not,

then memory is deteriorated.

The objective of the present study is to investigate the effect of spatial congruency of verbal stimuli on short-term verbal recognition memory as well as to compare evidence for either modal (i.e., sensorimotor simulation in the present study) or amodal (i.e., spatial indexing in the present study) theories of language representation and spatial interference. In particular, we aim to test whether implicit spatial information is perceptually (thus, modally) simulated and/or whether word locations are amodally indexed during short-term verbal recognition memory. Participants in Experiment 1 saw four words with implicit spatial meaning or metaphorical spatial associations presented simultaneously on a two-by-two grid at the encoding phase and they were auditorily probed with one of the four encoded words. During retrieval, a visual cue appeared on the grid either in the same location (i.e., congruent retrieval) or in a diagonal location (i.e., incongruent retrieval) as to the location of the probe word during encoding. Participants in Experiment 2 did not see a cue at retrieval but studied spatially neutral words in addition to upward and downward words.

Our predictions for Experiment 1 were straightforward: If sensorimotor simulation is indeed an intrinsic part of language representation with consequences in verbal recognition memory, then a word with a spatial association would be expected to activate the visuospatial experience of its referent and in turn, orient attention to the associated location (i.e., a modal explanation). In light of previous evidence showing spatial interference in semantic tasks, we predict that retrieval performance will be poorer in the case of a match between the location of the probe word and a visual cue during retrieval (i.e., congruent retrieval) compared to an incongruent retrieval condition, where the simulated location is empty at the time of retrieval. However, if spatial indexing is reflective of the retrieval process (i.e., an amodal explanation), then memory performance will be better when there is a matching visual cue at the retrieval screen as per the evidence demonstrating compatibility effect in memory. Adjudicating modal and amodal explanations is timely and critical considering the conflicting evidence and non-replications (e.g., Berndt et al., 2019) discussed above.

Experiment 1

Method

Participants

Experiment 1 was carried out with 48 students at the University of Birmingham (five males; M_{age} = 20.13, SD = 3.27, range: 18 – 41). One participant's accuracy (46%) was identified as an outlier based on the interquartile range and thus replaced. All participants were monolingual native speakers of British English (speaking/learning only English from birth and currently using English as their primary language) as determined with the Language History Questionnaire (version 2.0; Li et al., 2013). Participants reported normal or corrected-to-normal vision, no speech or hearing difficulties and no history of any neurological disorder. They received either £8 or course credit for participation. All participants were fully informed about the details of the experimental procedure and gave written consent. Post-experiment debriefing revealed that all participants were naïve to the purpose of the experiment. Ethical review and approval for data collection were conducted and granted by the University of Birmingham Ethics Committee.

Stimuli

The experiment consisted of 192 trials in total. Trials were evenly divided into two groups (n = 96) as experimental trials (positive probe) and fillers. Probe words in the experimental trials were among the four study words in the encoding phase, whereas a different, not seen, word was probed in fillers. Thus, 384 (96 x 4) words were used in positive probe trials and 480 (96 x 5) words were used in fillers. In total, there were 864 (384 + 480) unique words across 192 trials. All words were nouns (see supplemental online materials for the list of words used in the experiments). 384 words in the experimental trials have implicit spatial meaning or metaphorical spatial associations. The spatial associations of experimental words were verified with a separate spatial association norming study (see

below). Experimental words were also a subset of the extension of Paivio et al. (1968) norms (Clark &

Paivio, 2004).

Words in the experimental trials were arranged into sets with four words. Words in filler trials

were sets of five words, including four for the presentation and one unseen word for the following

probe. Words within both experimental and filler sets were matched on imageability, concreteness and

word length (all *SD*s < 2.00). Words were further controlled so that no word started with the same letter

or was semantically related with any other word in the trial set. Monosyllabic, disyllabic and trisyllabic

words were evenly distributed making 1/3 of each type (e.g., [3, 3, 3, 3], [1, 2, 1, 2] or [3, 2, 3, 2, 3] etc.).

Welch's t-tests revealed no significant difference between the probe and distractor words in

imageability, concreteness, word length or number of syllables (all *p*s > .05). Thus, any word among the

four words in each trial set was as likely to be remembered as any other word. We further formed 192

unique mathematical equations (e.g., $[2 \times 3] - [2 + 3] = 1$) to present as memory interference between

encoding and retrieval phases (see Conway & Engle, 1996 for a similar design). Half of the equations

were correct. Incorrect equations were further divided into two equal groups: The results were either

plus or minus one from the correct result.

Spatial Association Norming

Before running Experiment 1, norming was conducted with a different group of 30 monolingual

native speakers of British English (five males; M_{age} = 19, *SD* = 0.69, range: 18 – 20) who rated the spatial

associations of 1482 English nouns horizontal and vertical (x, y) coordinate space in a lab-based

experiment (Brysbaert et al., 2014 for a similar design). Handedness was tested with the Edinburgh

Handedness Inventory (Oldfield, 1971). Results showed that there were 16 right-handed, one left-

handed and 13 ambidextrous participants. The mean laterality quotient was 65 on a scale of -100

(absolute left-handedness) to +100 (absolute right-handedness) showing that right-handedness was

dominant among participants in particular including participants who can use their both hands, which is

a frequently seen case considering the prevalence of right-handedness (Corballis, 2003). The coordinate

space was a blank, central square of 480 x 480 in pixels (11° x 11° of visual angle) on the computer

screen. Horizontal and vertical axes of the coordinate were 960 pixels long (+/- 480 pixels in both

directions from the origin) giving 921,600 different clickable points. The material set was randomly

distributed over 5 lists of 290 words. The order of words in the lists was fully randomised. Participants

were presented a centrally placed single word (Courier, font size = 52) in capital letters. They were

instructed to read the word carefully and visualise its meaning before moving onto the rating phase.

There was no time limit for the study phase and it ended by hitting the space key. Participants were

presented a rating screen during the rating phase. They were expected to identify the word's location by

left clicking any point within the square. When a participant clicked a location, the experimental

software automatically recorded the click location both horizontally and vertically in the coordinate

space.

Control words (n = 32) were included to test inter-rater reliability. These words spanned the full

range of locations and were randomly distributed into each list. Analyses indicated excellent inter-rater

reliability based on intraclass correlation coefficients (ICC) for horizontal (ICC = 0.95, CI [95%] = 0.92 –

0.97) and vertical positions (ICC = 0.98, CI [95%] = 0.97 – 0.99) (see Cicchetti, 1994; Hallgren, 2012 for

cut-offs). The mean standard error in ratings was 3.02 for the horizontal plane and 5.37 for the vertical

plane, indicating a high degree of consistency between participants in location ratings. We also

computed rater correlation coefficients between each participant's average rating for a given word and

the mean rating for the word in question within each presentation list. The mean coefficient of the

correlation sets was 0.76 (SD = 0.07, range: 0.59 – 0.84) for the horizontal position and 0.65 (SD = 0.12,

range: 0.36 – 0.82) for the vertical position. Words that overlapped words with a data set from Estes et

al. (2015) (n = 95) showed a strong, positive correlation in vertical locations, r (92) = .92, p < .001.

Variances (horizontal position = 13,128.40; vertical position = 41,536.82) clearly showed that

ratings on the vertical axis were more spread out than the ratings on the horizontal axis. In other words,

participants tended to use the whole range of the vertical scale but rated the words more conservatively

on the horizontal scale. A heat map showing the distribution of the words across the coordinate system

indicated that the majority of the words were rated on a diagonal axis from bottom left to top right (see

Figure 1). This was also evident from the strong, positive correlation between horizontal and vertical

positions; r (1437) = .44, p < .001.

[Figure 1 near here]

384 words in the experimental trials were selected from the normed and labelled words to be

used as the experimental words in Experiment 1. An equal number of words (n = 24) in the experimental

set were associated with top, top right, bottom and bottom left locations in space. Horizontal positions

were not considered as a variable due to the dominance of vertical space over horizontal space (Franklin

& Tversky, 1990; Marmolejo-Ramos et al., 2013). Words were grouped as upward (top and top right)

and downward words (bottom and bottom left) because the majority of the words were normed in a

diagonal axis from bottom left to top right and there were very few words located in top left and bottom

right positions. All participants saw an equal number of upward and downward words in total (n = 48).

That said, the horizontal position was taken into consideration when placing the words in the grid: While

words associated with the upward and rightward position (e.g., *friend*) were placed in the top right

quadrant, words associated with upward position only (e.g., *heaven*) were placed in the top left

quadrant. Similarly, words associated with the downward and leftward position (e.g., *jail*) were placed in

the bottom left quadrant and words associated with the downward position only (e.g., *mule*) were

presented in the bottom right quadrant. Descriptive and spatial statistics of words used as a probe are

presented in Table 1. In addition to horizontal and vertical positions, we also calculated two spatial measures: spatiality and spatial variability. Spatiality shows the distance of a word from the centre. Hence, it is the extent to which a word is spatial as opposed to being spatially neutral. The spatiality value is the sum of the horizontal and vertical locations ratings (negative values were transformed into positive), normalised into a range from 0 to 1. Spatial variability was computed to investigate the words that participants rated on vastly different locations of the coordinate system. The spatial variability value is the average of the standard deviations between ratings among participants within each list, normalised into a range from 0 to 1.

[Table 1 near here]

Apparatus

Stimuli were presented on a TFT LCD 24-inch widescreen monitor operating at 144 Hz with a resolution of 1920 x 1080 pixels (508 mm x 285.75 mm). The experiment was programmed in and run on OpenSesame (version 3.1.7; Mathôt et al., 2012). Probe words that were presented auditorily in the retrieval phase were produced by a native female speaker of British English in a sound-attenuated room and recorded using Audacity (version 2.1.10; https://www.audacityteam.org). Participants responded (yes/no they had seen the word) by pressing one of two keys on a standard keyboard.

Data were analysed and visualised in R programming language and environment (R Core Team, 2020). Mixed-effects models were constructed with *lme4* package (Bates et al., 2015). Significance values of the coefficients in models were computed based on the t-distribution using the Satterthwaite approximation with *lmerTest* package (Kuznetsova et al., 2016). Bayes factors were computed with *BayesFactor* package (Morey & Rouder, 2015).

Procedure

The task was a yes/no verbal recognition memory test. As spelt out in detail below, each trial was composed of five consecutive phases: (1) fixation, (2) encoding, (3) interference, (4) retrieval and (5) feedback (see Figure 2). The task was to decide as quickly and accurately as possible whether an auditorily presented word had appeared previously or not. As soon as participants made a yes/no judgement (by hitting one of the response keys) the trial ended, and a new trial began. (1) Fixation: A fixation cross appeared at the centre of the screen for 500 ms. (2) Encoding: Participants were presented four words on a two by two grid for 1800 ms (see Richardson & Spivey, 2000 for a similar, "Hollywood Squares" design). The duration of the encoding phase was based on previous studies in which a compatibility effect was observed with a similar experimental design (Kumcu & Thompson, 2016, 2020). Encoding location was a between-subjects variable to minimise the learning effects between conditions. Half of the participants (n = 26, randomly chosen) saw the probe word in a spatially congruent location and the other half in an incongruent location. (3) Interference: Participants were presented mathematical equations and asked to identify whether the equation was correct or not within 10,000 ms (or they timed out). (4) Retrieval: The probe word was auditorily presented as participants looked at the screen with a blank grid. Participants were tasked to make a yes/no judgement to determine whether they had seen the probe word among the four words shown in the encoding phase within 3,000 ms (or they timed out). During the retrieval phase, a square appeared either in the same (congruent retrieval) or in the diagonal quadrant (incongruent retrieval) as the original location of the probe word in the encoding phase. Auditory words and visual probes were presented at the same time. Participants were instructed to look at the square as they gave their responses. Retrieval manipulation was a within-subjects variable. All participants were asked to retrieve the probe word in congruent and incongruent retrieval conditions. Participants were asked to respond as quickly and as accurately as possible. Therefore, they were instructed to keep their index fingers positioned on the response keys at all times. The key press was randomised such that half of the

participants pressed the 'J' key for yes and the 'F' key for no, while the other half pressed 'F' for yes and 'J' for no. (5) Feedback: Participants were shown their accuracy (correct or incorrect) and response time (RT) for 500 ms after each trial. Total accuracy and average RT were shown at the end of the experiment. For a detailed schema of the experiment, see Figure 2.

The order of trials and equations were fully randomised independently of each other. The experiment was divided into four equal blocks with 48 trials in each block and a short break between the blocks. A typical session lasted approximately 45 minutes. Overall accuracy in interference equations and the recognition memory test for words (including fillers) were 79% and 81% respectively, suggesting that participants attended to the task with high concentration.

Following the experiment, a computerised version of the Corsi block-tapping task (Corsi, 1972) was used to measure visuospatial short-term memory and a forward digit span test was used to measure short-term memory (using Psychology Experiment Building Language [PEBL], version 0.13, test battery version 0.7, http://pebl.org) (Mueller & Piper, 2014). Participants were instructed to reproduce lit-up blocks in the same order in Corsi block-tapping task. Corsi block span was reported as the dependent measure, which was calculated by taking the minimum list length, adding the total number correct, and dividing by the number of lists at each length. Participants were asked to repeat the visual/auditory presentation of number strings by typing in the forward digit span test. The reported dependent measure was digit span, which is the sum of the longest string of digits repeated without error over two trials. Participants were also asked to report their level of daytime sleepiness on the Epworth Sleepiness Scale.

[Figure 2 near here]

Results

Data Preparation

Homogeneity

Variability between the participants was considered as a random effect in the mixed-effects models and thus, was controlled in the analysis. Along with that, participants who were tested on congruent and incongruent encoding conditions were compared on individual variability measures (i.e., Corsi Block-tapping test, forward digit span, Epworth Sleepiness Scale) to verify the homogeneity between the groups. There was not a significant difference between participants tested on congruent and incongruent encoding conditions in any of the measures (see Table 2).

[Table 2 near here]

Measures

Accuracy and RT were used as measures of memory performance. Accuracy was the proportion of experimental trials to which the participants correctly responded yes. RT was the time in milliseconds between the onset of auditory presentation of the probe word and correct keyboard response. Responses identified as outliers that were 2 *SD* away (678.5 ms) from the mean (1079.41 ms) were removed (*n* = 179 out of 3,721 observations, 4.81%) from the RT dataset. The average accuracy in the memory retrieval task (excluding the fillers) was 80.75% (*SD* = 0.39) and the average RT (excluding fillers and incorrect responses) was 1029.02 ms (*SD* = 253.17). Overall, 4,608 observations for accuracy (excluding fillers) and 3,542 observations for RTs (excluding fillers and incorrect responses) collected from 48 participants were analysed.

Mixed-effects modelling

Data were analysed using linear and binomial logit mixed-effects modelling. There are several advantages of mixed-effects models with crossed random effects over traditional by-participant (*F1*) by-

item (*F2*) ANOVAs (Baayen et al., 2008). The use of mixed-effects models is particularly safer, more

powerful and more informative in psycholinguistics (see Hutchinson et al., 2013) because language

materials such as words used as test items in these studies show substantial variation even if they are

controlled on several properties. Further, it is misleading to generalise the findings beyond the language

materials specifically used as a sample in a given study (Clark, 1973). In the current study, both

participants and items were treated as random effects to explain by-participant and by-items variation

and to neutralise potentially relevant covariates such as learning, fatigue or order effects. Thus, we

captured the fact that some participants were more accurate and faster than others and some items

were easier to remember than others, which is not possible with a traditional ANOVA.

Visual inspections of residual plots did not reveal any obvious deviations from homoscedasticity

or linearity. Linear models were fit for continuous target variables (RT). Binomial models were fit for

categorical target variables (accuracy) and with *bobyqa* optimiser to prevent nonconvergence.

We started fitting models by building the random effects structure and followed a maximal

approach. That is, random effects were included as both random intercepts and correlated random

slopes (random variations) as long as they converged and were justified by the data (Barr et al., 2013).

Random intercepts and slopes were included even if they did not improve the model fit to control for

possible dependence due to repeated measures or order effects. Random slope models allowed the

explanatory variable to have a different effect for each group. Thus, we also captured that the fact that

the effect of encoding or retrieval condition on memory performance may vary by the lexical or

semantic properties of words such as word length.

Random effects structure was simplified step by step as per the magnitude of the contribution

of a random effect to the explanation of the variation in the data. That is, the random effect with the

weakest contribution was dropped first and if necessary, the structure was further reduced accordingly

(Winter, 2013). Results are reported in terms of the unstandardised regression coefficient, *B*.

Factor analysis

Diagnostic tests indicated collinearity between word variables (i.e., imageability, concreteness, word length, number of syllables, horizontal position and vertical position) (MeanVIF = 2.86, RangeVIF = 2.06 – 3.46; kappa = 2400.33). To address collinearity, we performed exploratory factor analysis using a principal component analysis extraction method with an orthogonal (varimax) rotation method. Kaiser's criterion and the scree test criterion showed the presence of three factors in our data. Three factors were interpreted as follows based on the loadings (see Table 3): (1) Imagery: Imageability and concreteness, (2) Length: word length and syllable length and (3) Word Position: Horizontal and vertical position. The three-factor solution explained 89.38% of the variance in the data. VIF of the factors were below one and thus, below our threshold of two. Regression scores calculated for three factors were employed in the subsequent linear mixed-effects multiple regression models.

[Table 3 near here]

Memory Performance

Mixed-effects models with accuracy and RT as target variables were fit. Fixed effects were encoding condition with two levels (i.e., spatially congruent encoding and spatially incongruent encoding) and retrieval condition with two levels (i.e., spatially congruent retrieval and spatially incongruent retrieval) and their interaction (see supplemental online materials for analysis codes).

All factors (i.e., imagery, word length and word position) were added as random slopes into participants and items for the general accuracy model. Summary of the general accuracy model with encoding and retrieval conditions as fixed effects indicated that overall, participants retrieved the probe word more accurately when they were cued to the congruent position (i.e., where the probe word appeared before) (*M* = 0.81, *SD* = 0.39) during retrieval; *B* = 0.25, *z* = 2.14, *p* = .03 compared to the

incongruent position (M = 0.80, SD = 0.40). For the general RT model, imagery and word length factors

were added as random slopes into participants and imagery factor was added as random slopes into

items as the maximal model did not converge. Summary of the general RT model with encoding and

retrieval conditions as fixed effects indicated that overall, words encoded in congruent locations (e.g.,

bird in the top left quadrant) were retrieved slower (M = 1075.26 ms, SD = 252.59) than words encoded

in incongruent locations (e.g., *bird* in the bottom right quadrant) (M = 985.63 ms, SD = 245.99); B =

103.49, t = 2.95, p = .005 regardless of the cue position at retrieval. Importantly, there was an

interaction effect between retrieval condition and imagery factor, showing that highly imageable words

slowed down retrieval when there was a visual cue in the same quadrant as to the initial location of the

probe word, B = 17.32, t = 2.39, p = .02.

Likelihood tests showed that encoding condition; x^2(1) = 0.49, p = .48 or retrieval condition; x^2(1)

= 1.23, p = .27 did not improve the null model fit for accuracy. Encoding condition improved the null

model fit for RT; x^2(1) = 6.30, p = .01, BF = 2.71 but not retrieval condition; x^2(1) = 0.75, p = .39.

No interaction was found between imagery factor and encoding for accuracy or RT and also,

imagery factor and retrieval for accuracy (the smallest p = .08). There was no correlation between the

performance in the mathematical equation and the performance in the recognition task (the smallest p

= .08).

The general accuracy model revealed an interaction effect between encoding and retrieval and

indicated worse accuracy with congruent encoding and congruent retrieval; B = -0.30, t = -1.93, p = .05.

A condition variable with four levels involving all four encoding and retrieval combinations was

formulated to investigate the interaction of encoding and retrieval conditions on memory performance

more precisely (see Table 4). Descriptive measures of memory performance across conditions were

presented in Table 4 in the order of accuracy from the most accurate to the least accurate.

[Table 4 near here]

Smaller models were fit only with congruently and incongruently encoded words. All factors

(i.e., imagery, word length and word position) were added as random slopes into participants and items

for the accuracy model. Retrieval condition within incongruent encoding condition improved the null

model fit for accuracy; x^2(1) = 4.59, *p* = .03, BF = 0.39. Summary of the accuracy model demonstrated

that words were retrieved more accurately when they were encoded in an incongruent position and

(e.g., *bird* in the bottom right quadrant) and a visual cue appeared in the same position (e.g., top right

quadrant) during retrieval; *B* = 0.23, *z* = 2.11, *p* = .03. Retrieval condition within congruent; x^2(1) = 0.13, *p*

= .72 or incongruent encoding; x^2(1) = 2.42, *p* = .12 did not improve model fit for null RT model.

[Figure 3 near here]

Discussion

Results of Experiment 1 demonstrated that spatial manipulation at encoding and retrieval

influenced memory performance. Incongruent encoding, with a congruent retrieval cue, was the most

accurate combination. We argue that this is because reading words with literal or metaphorical spatial

associations simulated the suggested locations. The performance was better when the simulated

location did not overlap with the visual cue in an incongruent encoding (e.g., bird in the bottom right

quadrant) followed by a congruent retrieval (e.g., visual cue in the bottom right quadrant. Participants

retrieved the words encoded in incongruent locations faster compared to congruent locations. That

said, we also found an interaction effect between imageability of the words and cue position on retrieval

times: Highly imageable words (e.g., *mountain*, *volcano* or *university* etc.) were retrieved more slowly

than less imageable words (e.g., *outcome*, *magnitude* or *truth* etc.) when there was a visual cue in the

same quadrant as the location of the probe word regardless of the spatial arrangement at encoding.

Such a finding suggests that highly imageable words with spatial associations elicit stronger simulations

of the associated locations and, in turn, lead to a stronger interference effect when matched with visual

information in the same location. In light of our findings and following the converging evidence

supporting perceptual competition (e.g., Verges & Duffy, 2009), we argue that sensorimotor simulation

dictated memory performance in Experiment 1. We attribute the spatial interference effect to

perceptual competition between the sensorimotor representation of words and visual cues. If an

amodal account, such as spatial indexing, were reflective of the retrieval process, memory performance

would have been better whenever there was a match between the location of the probe word and the

visual cue (i.e., congruent retrieval) as shown in previous studies (Johansson & Johansson, 2014; Scholz

et al., 2016). However, a match between the simulated location and the visual probe resulted in worse

performance.

Experiment 2

Experiment 1 examined spatial manipulation both at encoding and retrieval stages with the

encoding manipulation as a between-participants variable. Hence, we conducted a second experiment

in an attempt to replicate the findings with a larger sample and a simpler, more powerful design. To be

more precise, we tested the effect of spatial congruency on memory performance with a repeated

measures within-participants design. In other words, all participants saw the words in congruent and

incongruent positions at encoding. The retrieval manipulation was removed altogether from Experiment

2 to isolate the effect. Further, a control encoding condition with spatially neutral words normed to

central locations was introduced. In Experiment 1, words encoded in incongruent positions were

remembered better as compared to words encoded in congruent position. However, it was not clear

whether the mismatch advantage stemmed from an actual advantage for mismatched information or a

disadvantage for the matched information as there was no control condition with neutral words. The

neutral condition in Experiment 2 was added to disambiguate direction of the effect more precisely. If

spatial information was a part of verbal memory, then memory performance would be different across

encoding conditions, which would support results from Experiment 1. As there was no cue in this

experiment, we predicted that the performance would be better in the congruent encoding condition

compared to incongruent and/or neutral conditions.

Method

Participants

Experiment 2 was carried out with 55 monolingual native speakers of English recruited via

Prolific (https://www.prolific.co), an online participant recruitment system. The participant pool was

filtered to recruit individuals with normal or corrected-to-normal vision, no speech or hearing difficulties

and no history of any neurological disorder. They received £10 for participation. All participants were

fully informed about the details of the experimental procedure and gave consent. Ethical review and

approval for data collection were conducted and granted by the University of Birmingham Ethics

Committee. One participant's accuracy (48%) was identified as an outlier based on interquartile range

and thus removed from the dataset. The analysis reported below was conducted with 54 participants

(28 males; M_{age} = 21.13, SD = 2.28, range: 18 – 25).

Stimuli

As in Experiment 1, Experiment 2 consisted of 192 trials in total. The same materials as in

Experiment 1 were used except for spatially neutral words (n = 32) (e.g., *board* etc.) associated with the

central location in the norming study. Neutral words were shown in the neutral encoding condition.

Based on results from Experiment 1 and in Kumcu and Thompson (2020), the lexicosemantic variables

that were used to control the words were reduced: Only imageability and word length in letters were

selected from the imagery and length factors of Experiment 1. There was a high correlation between

spatiality and emotionality in the norming study; r (1437) = .37, p < .001 showing a difference between

spatial and neutral words in emotionality. Words that were located in the centre were less emotional.

Emotionality was thus added to control the words to make sure that any difference in memory between

spatial and neutral words stems from the spatial congruency manipulation rather than emotionality. As

expected, spatial and neutral words in Experiment 2 were different in emotionality; $t(94) = 4.87$, $p < .001$

but there was no such difference between up and down words; $t(62) = 1.50$, $p = 0.14$. One-way ANOVAs

showed that there was no difference in imageability; $p = 0.49$ or word length; $p = 0.3$ between the up,

down or neutral words. There was no difference between spatial (up and down words combined) and

neutral words in imageability; $p = 0.76$ or word length; $p = 1$ either. There were also reliable differences

between the up, down and neutral words in horizontal; $F(2,93) = 71.09$, $p < .001$ and vertical positions;

$F(2,93) = 661.09$, $p < .001$ as per the spatial norming (see Table 5).

Procedure

The same procedure as in Experiment 1 was used. The exceptions were the neutral encoding

condition and the exclusion of spatial cues shown at retrieval. Participants were now asked to retrieve

the probe word as quickly and as accurately as possible without any looking instructions. The main

manipulation of Experiment 2 was the spatial match between the spatial association of the probe word

and its position on the grid during encoding. Neutral probe words were counterbalanced across

participants to be shown in each one of the four quadrants of the grid. The spatial match condition,

location of the distractors, location of the neutral probe words, response hand-sides (old/new for the

memory task and correct/incorrect for the equation task) were counterbalanced across participants. As

a result, there were eight versions of the experiments. There was no difference in accuracy or RTs

between the test versions (all $ps > .05$). The experiment was conducted online via Gorilla

(https://gorilla.sc), an online experiment development and hosting service.

[Table 5 near here]

Results

Measures

Outliers at the higher and lower extremes (n = 94 out of 4,069 observations, 2.31%) that were identified based on the interquartile range were removed from the RT dataset. The average accuracy in the memory retrieval task (excluding fillers) was 80% (SD = 0.4) and the average RT (excluding fillers and incorrect responses) was 1310.40 ms (SD = 459.97). Overall, 5,063 observations for accuracy (excluding fillers) and 3,975 observations for RTs (excluding fillers and incorrect responses) collected from 54 participants were analysed.

Memory Performance

Word type with two levels (spatial vs. neutral) significantly improved both the null accuracy model; x^2(1) = 5.27, p = .02, BF = 0.40 and the null RT model; x^2(1) = 5.62, p = .02, BF = 0.63. Participants retrieved the spatial words less accurately (M = 0.80, SD = 0.4); B = 0.19, z = 2.20, p = .03 but faster (M = 1301.07 ms, SD = 457.31) than the spatially neutral words (M = 0.82, SD = 0.38 and M = 1328.91 ms, SD = 464.82); B = 26.26, t = 2.37, p = .02 regardless of the spatial match.

Spatial match condition with three levels (congruent, incongruent and neutral) improved the null accuracy model; x^2(2) = 6.53, p = .04, BF = 0.06 (see Figure 1 and Table 2). Participants retrieved the spatial words in incongruent positions (e.g., *victory* in downward position) (M = 0.79, SD = 0.41) less accurately than the spatially neutral words associated with the central position (M = 0.82, SD = 0.38); B = 0.24, z = 2.57, p = .01. There was no such difference in accuracy between the retrieval of spatial words in congruent position (e.g., *victory* in upward position) (M = 0.80, SD = 0.4) and neutral words; B = 0.14, z = 1.46, p = .14. A smaller model fitted with spatial words only indicated that there was no difference in accuracy between the retrieval of spatial words in congruent and incongruent positions either; x^2(2) = 1.29, p = .26. Spatial match condition improved the null RT model significantly; x^2(2) = 6.12, p = .05, BF =

0.06. Participants retrieved the words in congruent positions (e.g., *victory* in upward position) (*M* =

1290.57 ms, *SD* = 461.14) faster than the spatially neutral words associated with the central position (*M*

= 1328.91 ms, *SD* = 464.82); *B* = 30.77, *t* = 2.41, *p* = .02 (see Table 6 and Figure 4). There was no such

difference between the retrieval of spatial words in incongruent positions (e.g., *victory* in downward

position) (*M* = 1311.75 ms, *SD* = 453.31) and neutral words; *B* = 21.49, *t* = 1.68, *p* = .09. A smaller model

fitted with spatial words only indicated that there was no difference in accuracy between the retrieval of

spatial words in congruent and incongruent positions either; $x^2(1)$ = 0.36, *p* = .55.

[Table 6 near here]

[Figure 4 near here]

Spatial word type with two levels (up words vs. down words) did not improve the null accuracy

model fitted with only spatial words; $x^2(1)$ = 0.07, *p* = .8. Participants retrieved the up (*M* = 0.79, *SD* =

0.41) and down words (*M* = 0.80, *SD* = 0.4) equally accurately. That said, spatial word type improved the

null RT model; $x^2(1)$ = 5.85, *p* = .02, BF = 1.19. Overall, participants retrieved the up words (*M* = 1315.57

ms, *SD* = 470.81) slower than down words (*M* = 1286.69 ms, *SD* = 443.24); *B* = 31.24, *t* = 2.43, *p* = .02.

Smaller models with only up and down words indicated that the words associated with downward

location were retrieved slower when they were presented in incongruent positions (*M* = 1322.17 ms, *SD*

= 452.10) compared to congruent positions (*M* = 1251.68 ms, *SD* = 431.82); *B* = 50.48, *t* = 2.86, *p* = .004.

There was no such difference between the retrieval of up words in congruent (*M* = 1329.63 ms, *SD* =

486.02) and incongruent positions (*M* = 1301.19 ms, *SD* = 454.64); $x^2(1)$ = 2.77, *p* = .1.

Results showed that imageability (but not word length; *p* = .42 or emotionality; *p* = .32)

improved the accuracy in general; $x^2(1)$ = 4.33, *p* = .04, BF = 0.42. Overall, participants retrieved the high

imageable words more accurately than the low imageable words; *B* = 0.08, *z* = 2.13, *p* = .03. Interaction

models showed that imageability improved the accuracy for neutral words to the most extent; $B = 0.10$,

$z = 2.60$, $p = .009$ and also spatially congruent words with a smaller magnitude; $B = 0.08$, $z = 2.07$, $p = .04$.

However, imageability did not improve the accuracy of spatial words presented in incongruent

positions; $B = 0.06$, $z = 1.55$, $p = .12$. Results showed that word length (but not imageability; $p = .52$ or

emotionality; $p = .96$) affected the RTs in general; $x^2(1) = 4.32$, $p = .04$, BF $= 0.36$. Overall, participants

retrieved the longer words more slowly than the shorter words; $B = 7.62$, $t = 2.09$, $p = .04$. The effect of

word length was evident in the retrieval of spatially neutral words with the largest magnitude; $B = 10.73$,

$t = 2.82$, $p = .005$ and also in the retrieval of spatial words in incongruent positions; $B = 7.26$, $t = 1.94$, $p =$

.05. Word length did not affect the retrieval of spatial words in congruent positions; $B = 6.07$, $t = 1.63$, p

$= .1$.

Discussion

Results from Experiment 2 demonstrate that spatial manipulation at encoding and spatial

meaning of words manipulate memory performance. Participants retrieved spatial words less accurately

but faster than the spatially neutral words even though lexicosemantic differences between spatial and

neutral words were strictly controlled both at the stimuli development and analysis stages. Congruency

of spatial information affected retrieval performance as well. Participants retrieved spatial words in

incongruent positions less accurately than neutral words. Likewise, they remembered spatial words in

congruent positions faster than neutral words. Results from Experiment 2 support and extend results

from Experiment 1, although the direction of the effect is different as discussed below.

General Discussion

The present study investigated the effect of spatial associations on short-term verbal

recognition memory and disambiguated between modal and amodal accounts of spatial interference

effects in two experiments. In particular, Experiment 1 explored the role of sensorimotor simulation vs.

spatial indexing in verbal recognition short-term memory. Previous studies have yielded mixed results as

to how words with intrinsic or metaphorical spatial associations are processed: While there is evidence

that visually-loaded words guide attention to the associated location which in turn slows down the

identification of visual targets in the same location (e.g., Verges & Duffy, 2009), there is also evidence

that spatial compatibility between a word and a visual cue improves recognition memory (Wynn et al.,

2019). Against this background, we investigated modal (sensorimotor simulation) and amodal (spatial

indexing) theories of language representation and spatial interference effects in a short-term verbal

recognition memory task in Experiment 1. Words encoded in a spatially incongruent position from their

spatial association (e.g., *bird* in the bottom right quadrant or *mule* in the top left quadrant) were

retrieved more accurately when there was a visual cue in the same quadrant during retrieval (e.g.,

bottom right and top left, respectively) compared to a spatially incongruent retrieval cue. Words

encoded in a spatially incongruent position (e.g., *bird* in the bottom right quadrant or *mule* in the top

left quadrant) were retrieved faster than words encoded in a spatially congruent position (e.g., *bird* in

the top left quadrant or *mule* in the bottom right quadrant) regardless of the cue position during

retrieval. As a follow-up study, Experiment 2 investigated the effect of encoding manipulation on

memory performance with a control condition, a larger sample and more powerful within-participants

design. Here, participants retrieved the spatially-loaded words faster but less accurately than spatially

neutral words even when lexical and semantic differences between spatial and neutral words were

controlled. As to the congruency between the location of the probe word on the screen and the location

implied by the word, the results indicated that participants retrieved incongruent words less accurately

and congruent words faster than the control words, suggesting a match advantage.

The results from both experiments overall suggest that participants tapped into the spatial

associations of words during their retrieval such that spatial meanings and physical locations of words

on the computer screen impacted how well they were remembered. In this regard, results seem to

support the claims of embodied semantics that perceptual relations are encoded in the language

(Louwerse, 2008) with consequences for verbal memory. Crucially, these consequences occurred even though spatial information was neither relevant nor necessary for successful retrieval. In other words, participants were only instructed to retrieve the words as quickly and as accurately as possible in both experiments. They were not asked to remember where they saw the words, visualise them or judge their spatial iconicity. This occurred even though there was a difference between the conditions manipulated based on the location of the words and their referents. If spatial associations were not simulated in the current study, there would have been no systematic differences between the conditions. Further evidence in this direction is the effect of word imageability on perceptual simulation: In Experiment 1, highly imageable words were retrieved more slowly when there was a visual cue in the same quadrant as the location of the probe word (i.e., congruent retrieval) regardless of the spatial arrangement at encoding. The fact that imageability, which is a semantic variable, magnified the interference effect is important. This suggests that participants did not merely encode the word locations, as the spatial indexing theories assume but accessed the spatial meanings of words.

Metaphorical spatial referents of abstract words such as *liar* had similar effects on how they were remembered. As the large body of literature on the embodiment of abstractness has indicated (e.g., Barsalou, 2003), abstract words are also grounded in the sensorimotor system through conceptual metaphors (e.g., "GOOD IS UP" and "BAD IS DOWN") (Lakoff & Johnson, 1980). We argue that participants in the current study accessed the spatial part of their meaning which then cued them to certain locations (Experiment 1) or facilitated/hindered their retrieval according to their position on the screen (Experiment 2). This occurred even though abstract words intuitively have weaker spatial associations compared to highly imageable, concrete words. Estes et al. (2015) cast doubt on the perceptual competition account on the grounds that, in their study, words with strong and weak visual strength and imageability elicited equivalent spatial interference. Our results are at odds with these results. The results provide evidence that comprehenders perceptually simulate the intrinsic spatial

associations of words. Thus, our data support perceptual, grounded-embodied theories of language

representation (Barsalou, 1999, 2008; Paivio, 1986; Pulvermüller, 1999). That said, we do not argue that

spatial indexing and a compatibility effect should be discarded altogether given the abundant evidence

(Wynn et al., 2019). The argument here is that when analysed in the same experiment (Experiment 1),

the perceptual simulation evoked by presented words is so salient that any compatibility effect due to

spatial indexing is overshadowed. Thus, amodal explanations, such as spatial indexing, are not sufficient

to explain findings in studies showing spatial iconicity.

The effect of the sensorimotor simulation in the experiments was systematically in opposite

directions: In Experiment 1, words encoded in incongruent locations were retrieved more accurately

when there was a spatially matching cue. In Experiment 2, however, words encoded in incongruent

locations were retrieved less accurately compared to the control condition. Likewise, words encoded in

congruent locations were retrieved slower in Experiment 1 but they were retrieved faster in Experiment

2 compared to the control condition. These results are in line with the studies investigating the spatial

words in other semantic tasks and reporting match/mismatch advantage under different conditions. As

briefly discussed in the Introduction, tasks in which participants are asked to judge semantic relations

between spatial words result in a match advantage (e.g., Zwaan & Yaxley, 2003a), that is, spatial

congruency typically results in improved accuracy and/or response speed. On the other hand, a

mismatch advantage has been found in tasks in which participants acted on a visual target after reading

a spatial word (e.g., Bergen et al., 2007). We argue that this is also what we observed in both

experiments.

To be more precise, in Experiment 1 there was a mismatch advantage due to the spatial cueing

effect triggered by reading spatial words and competition between the simulated locations and the

visual cue. To illustrate, reading *bird* gave rise to a simulation of an upward location and seeing a visual

cue in the same position deteriorated retrieval accuracy due to a perceptual competition between the

spatial representation of the simulated location and the cognitive cost of perceiving actual visual

information in the same location. In this regard, results confirm previous studies reporting spatial

interference effect or mismatch advantage due to conceptual cueing effect (Bergen et al., 2007; Estes et

al., 2008; Gozli et al., 2013; Richardson et al., 2003; Verges & Duffy, 2009) and extend the mismatch

advantage to short-term verbal recognition memory, a relatively understudied task in the literature.

However, participants were not faced with such a perceptual competition because there was no visual

cue in Experiment 2 at retrieval. In turn, memory performance was better whenever the physical and

the simulated locations matched at encoding.

An issue to consider is why spatial manipulations in the current study affected accuracy and RTs

differently under different conditions. To be more precise, words encoded in congruent locations

slowed down their retrieval in Experiment 1 and words encoded in incongruent locations (with a

congruent cue) were retrieved more accurately. Likewise, results in Experiment 2 indicate that there

was a disadvantage for the mismatched words in accuracy and an advantage for the matched words in

response speed. Further, spatial words were remembered less accurately but faster than neutral words.

Taken together, it seems as the spatial incongruency affected accuracy rates while spatial congruency

affected retrieval times. Some other studies investigating sensorimotor simulation with spatial probes

noted similar discrepancies between accuracy rates and RTs. For example, top targets elicited faster but

less accurate responses than bottom targets in Estes et al. (2015). They attributed the difference to a

speed-accuracy tradeoff. However, Pearson correlation of the response times and accuracy scores in the

present study was weak in Experiment 1 (-0.24) and negligible in Experiment 2 (-0.06), ruling out a

speed-accuracy tradeoff. In any case, our results suggest that sensorimotor simulation affects accuracy

and response speed differently in memory, and further research disambiguating the effects of

sensorimotor simulation on memory is necessary.

Our results are crucial both for memory and language. Spatial and other sensory attributes of

words are often disregarded when developing verbal material for memory experiments. The current

study demonstrates that these variables are as important as more traditional variables such as

imageability and frequency and might affect how well individuals remember. As to language, we extend

the growing literature of embodiment of language processing. Given the experimental design and the

variety of verbal stimuli in the present study, our results have further implications in favour of modal

explanations over amodal and symbolic accounts. For example, Louwerse (2008) claimed that spatial

iconicity results can be explained by the frequency of linguistic relations between spatial pairs. In

consequence, iconic word pairs (*branch* over *root*) are processed faster than non-iconic counterparts

(*root* over *branch*) because language comprehenders are exposed to iconic linguistic symbols more

often (e.g., *branch* more frequently precedes *root* in everyday language use). Similarly, Zwaan and

Yaxley (2003a) speculated that words such as *attic* and *basement* could have links to the concepts of top

and bottom respectively given that attics are typically found in the top parts of the houses. Accordingly,

they argued that when the position of the word *attic* is amodally encoded on the computer screen, it is

tagged with the concept of top. Thus, a conflict between the tag assigned to the word and the contents

of its meaning yields a slower response. Our data speak against these amodal, symbolic accounts of

spatial iconicity. Participants in our study were not asked to judge semantic relations of pairs but to

retrieve a single word among three others, which were controlled based on several predictors of verbal

memory. Given that conformity of a single word location with its meaning influenced how well the word

was retrieved from memory, it is unlikely that linguistic relations played a role in our results. Even if

amodal theories could explain the findings for concrete words with literal spatial associations such as

attic or *basement*, they fail to explain why retrieval of words with metaphorical spatial meanings such as

hope or *chaos* was modulated according to their spatial arrangement. The fact that abstract words

triggered the simulation of metaphorical locations with consequences in memory in the present study

sheds further light on the grounding of abstract words, which is a long-standing challenge for grounded-embodied cognition (Dove, 2016). Along with that, future studies in non-Western languages with a right-to-left reading direction are particularly important to generalise the results.

To conclude, our results provide experimental support for converging evidence in favour of a simulation-based account of language representation against amodal, symbolic processes. Thus, in addition to the growing evidence, we argue that experiential and/or metaphorical spatial associations of words are perceptually reactivated in a task that is seemingly unrelated to the spatial information. Moreover, the mediating role of word imageability on the spatial interference effect in our study endorses a perceptual competition account to explain the spatial interference effect. The unique contribution of the present study is extending these results to short-term recognition memory involving language with further implications in support of a simulation account of language based on grounded-embodied cognition.

Disclosure statement

No potential conflict of interest was reported by the authors.

Open practices statement

Data associated with this article can be retrieved at https://osf.io/6wcen.

References

Amer, T., Gozli, D. G., & Pratt, J. (2017). Biasing spatial attention with semantic information: an event

coding approach. *Psychological Research*, 1–19. https://doi.org/10.1007/s00426-017-0867-5

Amorim, M., & Pinheiro, A. P. (2019). Is the sunny side up and the dark side down? Effects of stimulus

type and valence on a spatial detection task. *Cognition and Emotion*, *33*(2), 346–360.

https://doi.org/10.1080/02699931.2018.1452718

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random

effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory

hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

https://doi.org/10.1016/j.jml.2012.11.001

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–609;

discussion 610-660. https://doi.org/10.1017/S0140525X99252144

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the

Royal Society B: Biological Sciences*, *358*(1435), 1177–1187.

https://doi.org/10.1098/rstb.2003.1319

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

https://doi.org/10.1146/annurev.psych.59.103006.093639

Bates, D. M., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4.

Journal of Statistical Software, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bergen, B. K. (2016). Embodiment, simulation and meaning. In N. Riemer (Ed.), *The Routledge handbook

of semantics* (pp. 142–156). Routledge. https://doi.org/10.4324/9781315685533.ch8

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual

imagery in sentence comprehension. *Cognitive Science*, *31*(5), 733–764.

https://doi.org/10.1080/03640210701530748

Berndt, E., Dudschig, C., Miller, J., & Kaup, B. (2019). A replication attempt of hemispheric differences in

semantic-relatedness judgments (Zwaan & Yaxley, 2003). *Acta Psychologica*, *198*(December 2018).

https://doi.org/10.1016/j.actpsy.2019.102871

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally

known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911.

https://doi.org/10.3758/s13428-013-0403-5

Chasteen, A. L., Burdzy, D. C., & Pratt, J. (2010). Thinking of God moves attention. *Neuropsychologia*,

48(2), 627–630. https://doi.org/10.1016/j.neuropsychologia.2009.09.029

Chomsky, N. (1957). *Syntactic structures*. Mouton de Gruyter.

Cicchetti, D. V. (1994). Guidlines, criteria, and rules of thumb for evaluating normed and standardized

assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290.

https://doi.org/10.1037/1040-3590.6.4.284

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological

research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359.

https://doi.org/10.1016/S0022-5371(73)80014-3

Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior

Research Methods, Instruments, and Computers*, *36*(3), 371–383.

https://doi.org/10.3758/BF03195584

Conway, A. R., & Engle, R. W. (1996). Individual differences in working memory capacity: more evidence

for a general capacity theory. *Memory*, *4*(6), 577–590. https://doi.org/10.1080/741940997

Corballis, M. C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness.

Behavioral and Brain Sciences, *26*(2), 199–208. https://doi.org/10.1017/S0140525X03000062

Corsi, P. (1972). *Human memory and the medial temporal lobe region of the brain* [Unpublished doctoral

dissertation]. McGill University.

Dove, G. (2016). Three symbol ungrounding problems: Abstract concepts and the future of embodied

cognition. *Psychonomic Bulletin & Review*, *23*(4), 1109–1121. https://doi.org/10.3758/s13423-015-

0825-4

Dudschig, C., Souman, J., Lachmair, M., de la Vega, I., & Kaup, B. (2013). Reading "sun" and looking up:

The influence of language on saccadic eye movements in the vertical dimension. *PLoS ONE*, *8*(2),

e56872. https://doi.org/10.1371/journal.pone.0056872

Estes, Z., & Barsalou, L. W. (2018). A comprehensive meta-analysis of spatial interference from linguistic

cues: Beyond Petrova et al. (2018). *Psychological Science*, *29*(9), 1558–1564.

https://doi.org/10.1177/0956797618794131

Estes, Z., Verges, M., & Adelman, J. S. (2015). Words, objects, and locations: Perceptual matching

explains spatial interference and facilitation. *Journal of Memory and Language*, *84*, 167–189.

https://doi.org/10.1016/j.jml.2015.06.002

Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head up, foot down: Object words orient attention to the

objects' typical location. *Psychological Science*, *19*(2), 93–97. https://doi.org/10.1111/j.1467-

9280.2008.02051.[x]

Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental

Psychology: General*, *119*(1), 63–76. https://doi.org/10.1037/0096-3445.119.1.63

Giessner, S. R., & Schubert, T. W. (2007). High in the hierarchy: How vertical location and judgments of

leaders' power are interrelated. *Organizational Behavior and Human Decision Processes*, *104*(1),

30–44. https://doi.org/10.1016/j.obhdp.2006.10.001

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency.

Journal of Experimental Psychology, *61*(1), 23–29. https://doi.org/10.1037/h0040561

Gozli, D. G., Chasteen, A. L., & Pratt, J. (2013). The cost and benefit of implicit spatial cues for visual

attention. *Journal of Experimental Psychology: General*, *142*(4), 1028–1046.

https://doi.org/10.1037/a0030362

Gozli, D. G., Pratt, J., Martin, K. Z., & Chasteen, A. L. (2016). Implied spatial meaning and visuospatial

bias: Conceptual processing influences processing of visual targets and distractors. *PLoS ONE*,

11(3), e0150928. https://doi.org/10.1371/journal.pone.0150928

Hallgren, K. A. (2012). Computing Inter-rater reliability for observational data: An overview and tutorial.

Tutorials in Quantitative Methods for Psychology, *8*(1), 23–34.

https://doi.org/10.20982/tqmp.08.1.p023

Hunt, R. R., & Worthen, J. B. (Eds.). (2006). *Distinctiveness and memory*. Oxford University Press.

Hutchinson, S., Wei, L., & Louwerse, M. M. (2013). Avoiding the language-as-a-fixed-effect fallacy: How

to estimate outcomes of linear mixed models. In P. Bello, M. Guarini, M. McShane, & B. Scassellati

(Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2393–2398).

Cognitive Science Society.

Johansson, R., & Johansson, M. (2014). Look here, eye movements play a functional role in memory

retrieval. *Psychological Science*, *25*(1), 236–242. https://doi.org/10.1177/0956797613498260

Johns, M. W. (1991). A new method for measuring daytime sleepiness: The Epworth sleepiness scale.

Sleep, *14*(6), 540–545. https://doi.org/10.1016/j.sleep.2007.08.004

Kessels, R. P. C., Zandvoort, M. J. E. Van, Postma, A., Kappelle, L. J., & Haan, E. H. F. De. (2000). the Corsi

block-tapping task: Standardization and normative data. *Applied Neuropsychology*, *7*(4), 252–258.

https://doi.org/10.1207/S15324826AN0704_8

Kumcu, A., & Thompson, R. L. (2016). Spatial interference and individual differences in looking at

nothing for verbal memory. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.),

Proceedings of the 38th Annual Conference of the Cognitive Science Society (pp. 2387–2392).

Cognitive Science Society.

Kumcu, A., & Thompson, R. L. (2020). Less imageable words lead to more looks to blank locations during

memory retrieval. *Psychological Research*, *84*, 667–684. https://doi.org/10.1007/s00426-018-1084-

6

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2016). lmerTest: Tests in linear mixed effects models. R

package version 3.0.0. Retrieved from https://cran.r-project.org/package=lmerTest.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Leitan, N. D., Williams, B., & Murray, G. (2015). Look up for healing: Embodiment of the heal concept in

looking upward. *PLoS ONE*, *10*(7), e0132427. https://doi.org/10.1371/journal.pone.0132427

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.).

Oxford University Press.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2013). Language history questionnaire (LHQ 2.0): A new dynamic

web-based research tool. *Bilingualism: Language and Cognition*, *17*(3), 673–680.

https://doi.org/10.1017/S1366728913000606

Louwerse, M. M. (2008). Embodied relations are encoded in language. *Psychonomic Bulletin & Review*,

15(4), 838–844. https://doi.org/10.3758/PBR.15.4.838

Marmolejo-Ramos, F., Elosúa, M. R., Yamada, Y., Hamm, N. F., & Noguchi, K. (2013). Appraisal of space

words and allocation of emotion words in bodily space. *PLoS ONE*, *8*(12), 18–20.

https://doi.org/10.1371/journal.pone.0081688

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment

builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324.

https://doi.org/10.3758/s13428-011-0168-7

Meyer, B. P., & Robinson, M. D. (2004). Why the sunny side Is up: Associations between affect and

vertical position. *Psychological Science*, *15*(4), 243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x

Montoro, P. R., Contreras, M. J., Elosúa, M. R., & Marmolejo-Ramos, F. (2015). Cross-modal metaphorical mapping of spoken emotion words onto vertical space. *Frontiers in Psychology*, *6*(August), 1–13. https://doi.org/10.3389/fpsyg.2015.01205

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.10-1. Retrieved from http://cran.r-project.org/package=BayesFactor

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods*, *222*, 250–259. https://doi.org/10.1016/j.jneumeth.2013.10.024

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Ostarek, M., & Vigliocco, G. (2017). Reading sky and seeing a cloud: On the relevance of events for perceptual simulation. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(4), 579–590. https://doi.org/10.1037/xlm0000318

Paivio, A. (1986). *Mental representation: A dual coding approach*. Oxford University Press.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1).

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*(2), 253–279; discussion 280-336. https://doi.org/10.1017/S0140525X9900182X

Pylyshyn, Z. W. (1986). *Meaning and cognitive structure: Issues in the computational theory of mind*. Ablex.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*(1), 65–97. https://doi.org/10.1016/0010-0277(89)90014-0

Quadflieg, S., Etzel, J. A., Gazzola, V., Keysers, C., Schubert, T. W., Waiter, G. D., & Macrae, C. N. (2011). Puddles, parties, and professors: linking word categorization to neural patterns of visuospatial coding. *Journal of Cognitive Neuroscience*, *23*(10), 2636–2649. https://doi.org/10.1162/jocn.2011.21628

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*(3), 269–295. https://doi.org/10.1016/S0010-0277(00)00084-6

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*(5), 767–780. https://doi.org/10.1016/S0364-0213(03)00064-8

Scholz, A., Mehlhorn, K., & Krems, J. F. (2016). Listen up, eye movements play a role in verbal memory retrieval. *Psychological Research*, *80*(1), 149–158. https://doi.org/10.1007/s00426-014-0639-4

Schubert, T. W. (2005). Your highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology*, *89*(1), 1–21. https://doi.org/10.1037/0022-3514.89.1.1

Šetić, M., & Domijan, D. (2007). The influence of vertical spatial orientation on property verification. *Language and Cognitive Processes*, *22*(2), 297–312. https://doi.org/10.1080/01690960600732430

Spivey, M. J., Richardson, D. C., & Fitneva, S. A. (2004). Thinking outside the brain: Spatial indices to visual and linguistic Information. In J. Henderson & F. Ferreira (Eds.), *Interfacing language, vision, and action: Eye movements and the visual world* (pp. 161–190). Psychology Press. https://doi.org/10.4324/9780203488430

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156. https://doi.org/10.1111/1467-

9280.00326

Taylor, J. E. T., Lam, T. K., Chasteen, A. L., & Pratt, J. (2015). Bow your head in shame, or, hold your head

up with pride: Semantic processing of self-esteem concepts orients attention vertically. *PLoS ONE*,

10(9), e0137704. https://doi.org/10.1371/journal.pone.0137704

van Ede, F., de Lange, F. P., & Maris, E. (2012). Attentional cues affect accuracy and reaction time via

different cognitive and neural processes. *Journal of Neuroscience*, *32*(30), 10408–10412.

https://doi.org/10.1523/JNEUROSCI.1337-12.2012

Verges, M., & Duffy, S. (2009). Spatial representations elicit dual-coding effects in mental imagery.

Cognitive Science, *33*(6), 1157–1172. https://doi.org/10.1111/j.1551-6709.2009.01038.x

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*.

http://arxiv.org/abs/1308.5499

Wynn, J. S., Shen, K., & Ryan, J. D. (2019). Eye movements actively reinstate spatiotemporal mnemonic

content. *Vision*, *3*(2), 21. https://doi.org/10.3390/vision3020021

Zanolie, K., Dantzig, S. van, Boot, I., Wijnen, J., Schubert, T. W., Giessner, S. R., & Pecher, D. (2012).

Mighty metaphors: Behavioral and ERP evidence that power shifts attention on a vertical

dimension. *Brain and Cognition*, *78*(1), 50–58. https://doi.org/10.1016/j.bandc.2011.10.006

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan

(Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking*

(pp. 224–245). https://doi.org/10.1017/CBO9780511499968.010

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory.

Psychological Bulletin, *123*(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162

Zwaan, R. A., Stanfield, R., & Yaxley, R. (2002). Language comprehenders mentally represent the shapes

of objects. *Psychological Science*, *13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430

Zwaan, R. A., & Yaxley, R. H. (2003a). Hemispheric differences in semantic-relatedness judgments.

Cognition, *87*, B79–B86. https://doi.org/10.1016/S0010-0277(02)00235-4

Zwaan, R. A., & Yaxley, R. H. (2003b). Spatial iconicity affects semantic relatedness judgments.

Psychonomic Bulletin & Review, *10*(4), 954–958. https://doi.org/10.3758/BF03196557

Table 1

Descriptive Measures (Mean and SD) of the Probe Words in Positive Trials of Experiment 1

Variable	Upward words	Downward words	Total
Number	48	48	96
Imageability	4.84 (1.41)	5.02 (1.28)	4.93 (1.34)
Concreteness	4.19 (1.91)	4.67 (2)	4.43 (1.96)
Length in letters	7.02 (1.83)	6.5 (1.75)	6.76 (1.80)
Syllable length	2.33 (1)	2.12 (0.91)	2.23 (0.96)
Horizontal position	127.8 (114.85)	-118.88 (143.5)	4.46 (179.13)
Vertical position	308.6 (73.62)	-301.2 (98.22)	3.7 (318.43)
Spatiality	0.58 (0.13)	0.61 (0.15)	0.59 (0.14)
Spatial variability	0.33 (0.11)	0.35 (0.1)	0.34 (0.11)

Table 2

Results of the Individual Variability Measures (Mean and SD) Among All Participants and Across Encoding

Groups

Measures	All	Congruent encoding	Incongruent encoding	*p*
Corsi block span	58.94 (24.25)	53.75 (25.01)	64.12 (22.81)	.140
Digit span	9.15 (2.41)	9.62 (2.58)	8.67 (2.16)	.170
Sleepiness	0.84 (0.55)	0.85 (0.51)	0.83 (0.6)	.897

Note. The average score for healthy adults in the Corsi block-tapping task is 55.7 (Kessels et al., 2000)

and 8 – 10 in forward digit span test (Lezak et al., 2012). The normal range of daytime sleepiness is 0 –

10 (Johns, 1991).

Table 3

Factor-Loadings of the Predictors

	Factor			
Predictors	Imagery	Length	Position	h2
Imageability	**0.93**	-0.21	-0.02	0.92
Concreteness	**0.95**	-0.15	-0.06	0.92
Length in letters	-0.19	**0.93**	0.11	0.91
Number of syllables	-0.17	**0.94**	0.05	0.91
Horizontal position	-0.02	0.09	**0.92**	0.85
Vertical position	-0.05	0.06	**0.92**	0.86
Factor statistics				
Eigenvalue	2.62	1.60	1.13	7.99
Variance (%)	30.55	30.28	28.54	89.93

Table 4

Descriptive Measures (Mean and SD) of Memory Performance in Experiment 1

Condition	Accuracy (%)	RT (ms)
Incongruent encoding + congruent retrieval	84 (0.37)	993.53 (252.77)
Incongruent encoding + incongruent retrieval	80 (0.40)	977.34 (238.54)
Congruent encoding + incongruent retrieval	80 (0.40)	1077.92 (254.31)
Congruent encoding + congruent retrieval	79 (0.41)	1072.54 (250.95)

Table 5

Descriptive Measures (Mean and SD) of the Probe Words in Positive Trials of Experiment 2

Variable	Spatial words	Upward words	Downward words	Neutral words
Number	64	32	32	32
Imageability	4.90 (1.27)	4.73 (1.36)	5.08 (1.16)	4.98 (1.06)
Word length	6.72 (1.54)	7.00 (1.68)	6.44 (1.34)	6.72 (1.25)
Emotionality	4.12 (1.2)	4.34 (1.18)	3.90 (1.2)	2.94 (0.93)
Horizontal position	-4.93 (170.40)	131.60 (95.85)	-141.46 (106.48)	24.64 (71.02)
Vertical position	-2.81 (315.07)	303.14 (70.11)	-308.76 (59.86)	1.89 (71.26)
Spatiality	0.58 (0.09)	0.56 (0.15)	0.60 (0.16)	0.15 (0.09)
Spatial variability	0.34 (0.11)	0.32 (0.1)	0.35 (0.11)	0.39 (0.14)

**Table 6**

Descriptive Measures (Mean and SD) of Memory Performance in Experiment 2

| Spatial match condition	Accuracy (%)	RT (ms)
Spatial words	80 (0.40)	1301.07 (457.31)
Congruent encoding	80 (0.40)	1290.57 (461.14)
Incongruent encoding	79 (0.41)	1311.75 (453.31)
Neutral words	82 (0.38)	1328.91 (464.82)

Figure 1

The Heat Map Shows the Distribution of Word Location Ratings over Positions on the Horizontal and

Vertical Axis
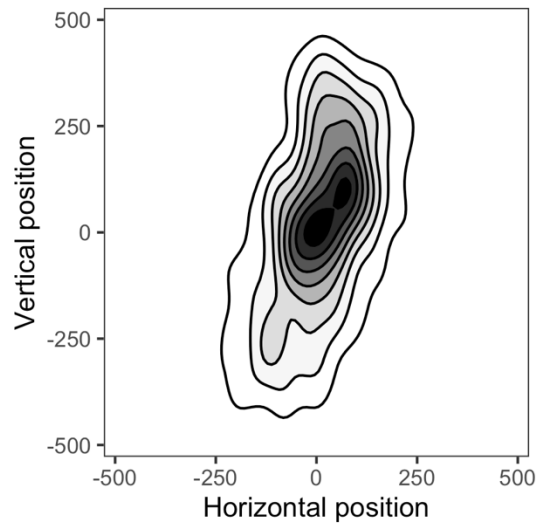
Note. Lighter colours indicate lower frequency and darker colours indicate higher frequency. Note that

the blank area outside the pattern does not refer to an absence of ratings, but rating frequency under a

certain threshold (see supplemental online materials for an interactive scatterplot showing all word

locations on the horizontal and vertical axes).

Figure 2

A Schematic Illustration of the Temporal Order of Events in an Example Trial Showing Two Different

Encoding and Two Different Retrieval Conditions in Experiment 1
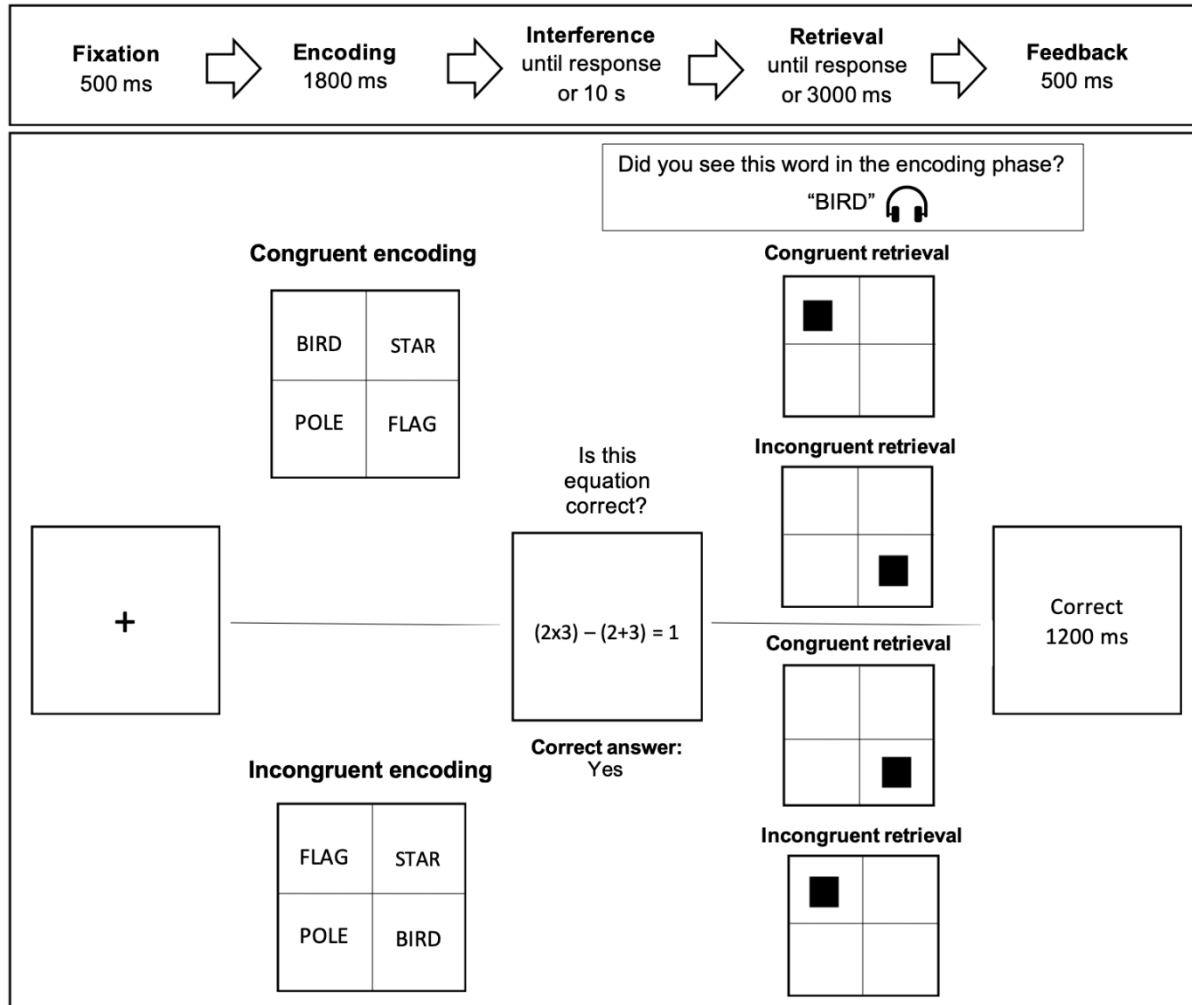
Note. A schematic illustration of the temporal order of events in an example trial showing two different encoding and two different retrieval conditions in Experiment 1. In this example, the relevant quadrant is the top left location, where the probe word (i.e., BIRD) appears. Feedback on this diagram (Correct 1200 ms) is given as an example.

Figure 3

Accuracy and RTs Between Four Combinations of Encoding and Retrieval Conditions in Experiment 1
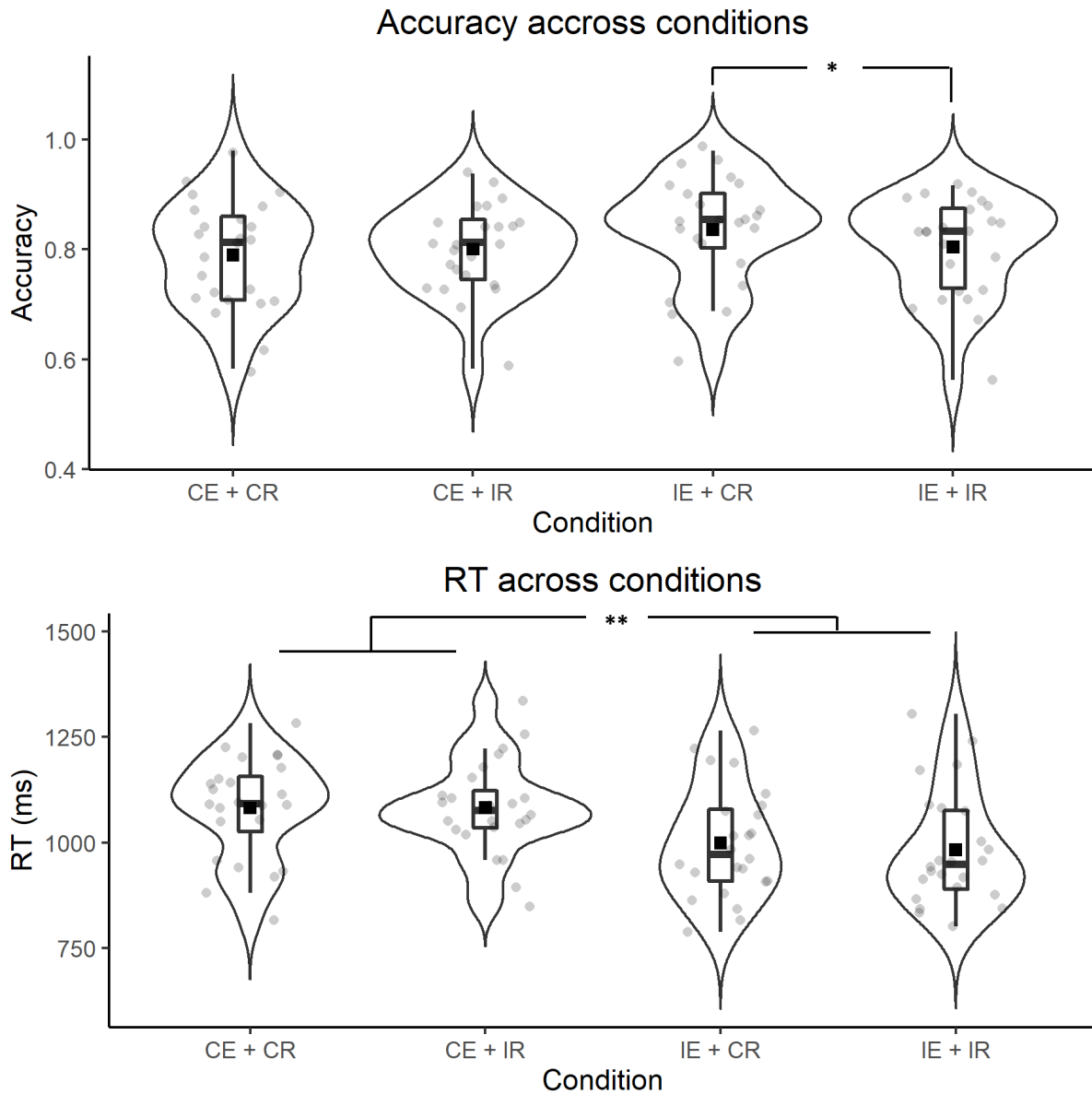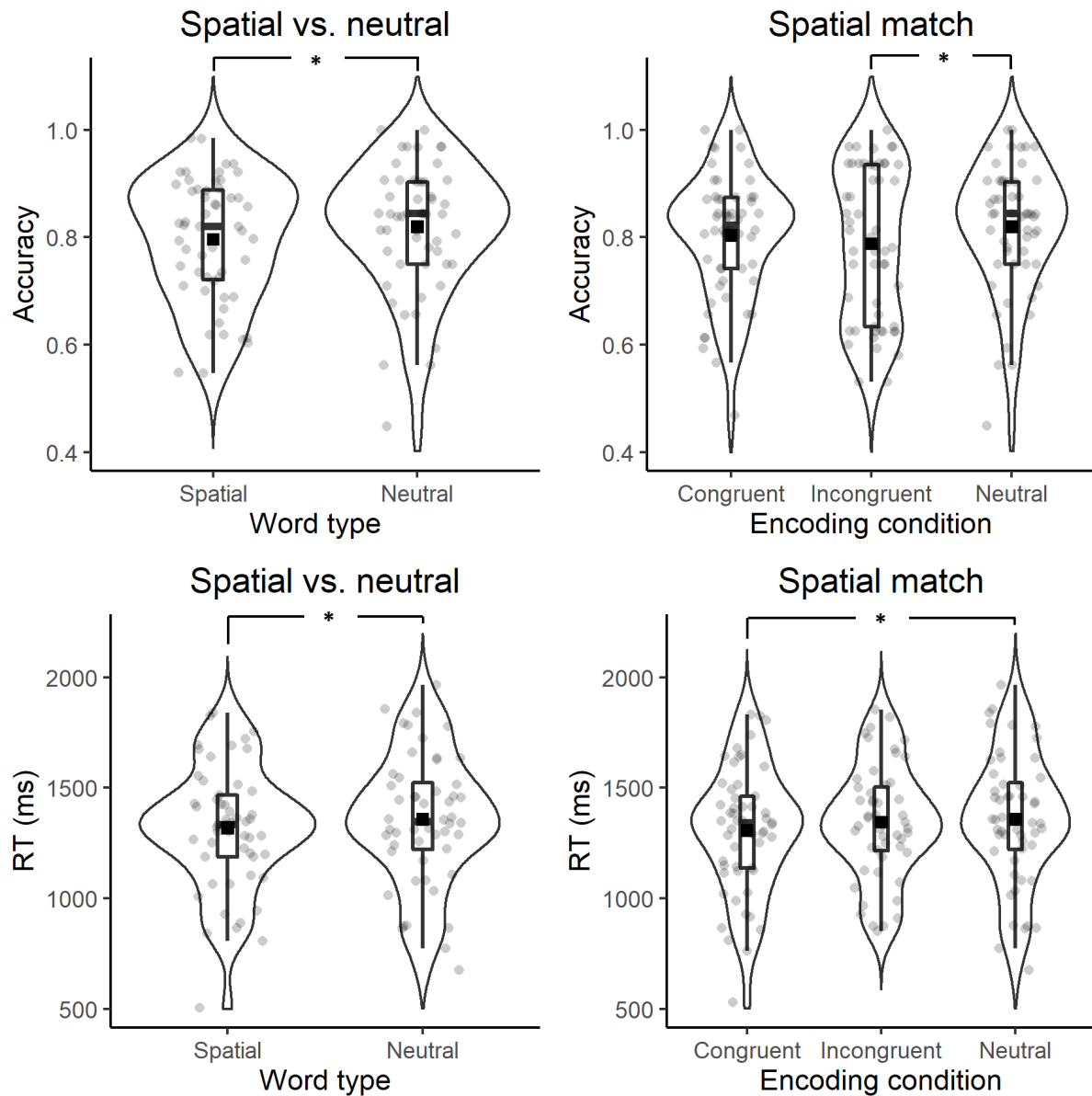
Note. Violin plots show the distribution of accuracy and RTs as per condition. The box plots within the

violins show the whole range (vertical line), the mean (black square), and the median (horizontal line).

Grey dots represent data points for individual participants. (CE + CR: Congruent Encoding and Congruent

Retrieval, CE + IR: Congruent Encoding and Incongruent Retrieval, IE + CR: Incongruent Encoding and

Congruent Retrieval, IE + IR: Incongruent Encoding and Incongruent Retrieval). * $p \leq$ .05 ** $p \leq$ .01.

Figure 4

Accuracy and RTs across Word Types and Encoding Conditions in Experiment 2

Note. Violin plots show the distribution of accuracy and RTs per condition. The box plots within the

violins show the whole range (vertical line), the mean (black square), and the median (horizontal line).

Grey dots represent data points for individual participants. * $p \leq .05$