Interference effects of novel word-object learning on visual perception

Shane Lindsay (s.lindsay@hull.ac.uk)
Robert Nightingale (r.nightingale@hull.ac.uk)

Psychology, University of Hull, HU6 7RX, United Kingdom

Abstract

Previous studies investigating effects of language comprehension on spatial processing have used existing words with pre-existing spatial associations. Here participants learnt novel words and novel objects with spatial associations. Following training, participants had to judge whether a visual object matched a word. Objects could match in identity or in spatial location. In Experiment 1, participants learnt just novel words and objects; Experiment 2 compared performance with existing objects with pre-existing spatial associations. We found mismatching (but task irrelevant) spatial information interfered with judgements of object identity, but only for novel words. In Experiment 3, we altered correspondence between visual targets and semantics using a target discrimination task, where the target had no relationship to the verbal cue. We found the opposite results to the previous two studies, as responses to spatially matching targets were slower than spatially mismatching targets. We discuss implications for embodied and non-embodied accounts of these findings.

Keywords: word learning; embodied cognition; spatial memory; interference; semantics

Introduction

Barsalou (1999) describes how understanding language may involve the activation of perceptual simulations of sentences or the concepts that individual words stand for. In support of this claim, there have been a range of studies indicating language can facilitate visual processing (e.g., Zwaan & Stanfield, 2012), particularly in cases where there is a match between representations activated by language and perception. In some cases, however, language appears to interfere with visual processing. For example, reading a sentence describing an event directed towards upper visual space (e.g. "*The mule climbed*") can lead to a reduction of perceptual discrimination performance where there is a spatial mismatch between the activated semantics and a target (e.g., discriminating an X or O in the upper visual field; Bergen, Lindsay, Matlock & Narayanan, 2007).

These results, referred to here as spatial match interference effects, have been found with sentences (Bergen, Lindsay, Matlock & Narayanan, 2007) and with single words (Estes, Verges & Barsalou, 2008), and have been argued to result from language semantics leading to an automatic shift in spatial attention. This shift is followed by a perceptual simulation interfering with perception. This has been described as a Perky effect (Perky, 1910), and argued to arise from embodied visuo-spatial simulations interfering with visuo-spatial perception due to shared resources or processing (Bergen, 2007).

A recent study proposed an alternative account to the embodied simulation theory to explain interference on spatially congruent trials. Estes, Verges and Adelman (2015) suggested that perceptual matching provides an alternative explanation for the interference and enhancement effects of language on visual processing, and were able to find both interference and facilitation effects within a single task.

As part of their explanation for how both effects could emerge, Estes et al. describe a coding system to account for the perceptual matching combinations. To illustrate the coding scheme, imagine a scenario shown in Figure 1, where you have learnt the word for a novel object, a *"Brend"*, and that object always appeared in an upwards location, seen in the top left panel. You are then given that verbal label again and an object that matches or mismatches the label, and the object's location matches or mismatches the trained location

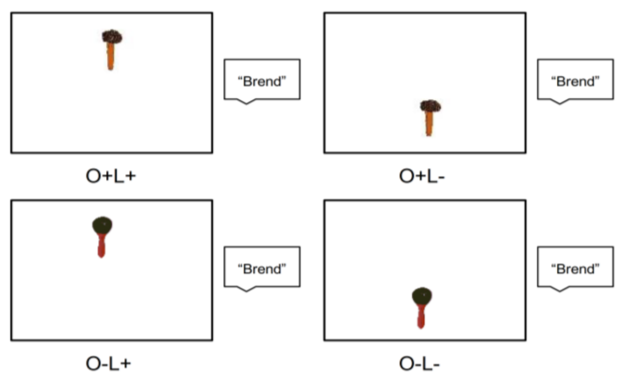

Figure 1: An illustration of a coding scheme for object and location matching and a preview of design of the main task in Experiment 1.

In Figure 1 above, O+ indicates an object match between the novel word *"Brend"* and its paired novel object, and O- a mismatch. L+ indicates a location match of a target object with the typical associated location for a named object, and L- a mismatch of the location with where it is typically found.

The account of Estes et al. would suggest that hearing "*Brend*" should trigger a perceptual search for the referent. This search would initiate in its typical location. In cases where a mismatching object appears in the matching associated location for "*Brend*" (i.e., O-L+ in the bottom left) there is a processing cost that explains the spatial matching interference. This cost is suggested to be due to the inconsistent codes triggering additional processing to resolve the discrepancy, based on other findings indicating that inconsistent object and location codes can interfere with responses (Lu & Proctor, 1995)

In previous studies that have shown slower reaction times to perceptual targets that mismatched typical locations (e.g., Bergen et al., 2007; Estes et al., 2008) there has typically been a mismatch between a target (e.g., a X or O in a perceptual discrimination task) and language semantics (e.g., a referent to the word *"bird"*). Estes et al. (2015) used a task that manipulated object match, with cue words either matching objects (e.g., the word "*bird*" and a picture of a bird) or mismatching objects (e.g., a picture of a wrench). They were able to find support for the predictions of the perceptual matching account, with reaction times faster for matching objects and matching locations (O+L+; spatial match facilitation) but slower responses to mismatching objects but matching locations (O-L+; spatial match interference). This provides an alternative explanation to embodied simulation accounts of spatial interference effects by explaining it as the code-based matching of object and locations. Further evidence against visual simulations causing spatial match interference comes from additional findings that visual strength did not predict the size of effects (Estes et al., 2015), though spatial association did. This graded effect of spatial association was used to argue against polarity correspondence principle explanation of the findings (see also Santiago & Lakens, 2015).

In the studies described here we investigated these object and spatial matching effects using novel words and novel objects. Effects of novel word and novel object pairs offer a unique contrast to the results obtained by studies with existing words. For novel word and novel object pairs, the object will not be part of any existing semantic systems. When new words are learned for unfamiliar objects it is hypothesized that associations between them and their typical locations will be encoded within an episodic memory system. But it is unknown whether activation of these specific episodic-based visuo-spatial representations, triggered through language comprehension, will lead to the same kind of effects as words from an existing semantic system, established through a life history of experiencing objects in the world.

In Experiment 1, we trained participants with novel words and novel objects that systematically appeared in either top or bottom locations (see Figure 1, top left, for an example of a "top" object). Spatial location was always irrelevant to the training and test tasks. At test, we presented the cue word followed by an object where the task was to judge if the object matched the cue word. Objects and locations could either match or mismatch the recently learned novel objects and implicit spatial associations.

If perceptual simulation at cued locations leads to perceptual interference, then performance should be reduced both for object match and mismatches (O+L+ and O-L+). However, if perceptual simulation facilitates perception when there is a close visual match between what is simulated and observed (as occurs with our novel word-object paradigm), then we may see facilitation for both object match and mismatches. In contrast, the perceptual matching account gives a more nuanced prediction that spatial match interference should only occur with mismatching objects, as in this case the object and spatial codes are not congruent (O-L+), and comparatively, facilitation should be seen with matching codes (O+L+ & O-L-).

Experiment 1: Novel Words and Object Recognition

Method

Participants Twenty-eight participants were recruited from first and second year Undergraduate psychology modules at the University of Hull and received course credits. All participants had normal or corrected to normal vision.

Stimuli Sixteen novel words and images were selected from the NOUN database (Horst & Hout, 2016). Recordings of the words were from a male British English speaker. Mean utterance length was 644 ms (*SD* = 144 ms).

Thirty-two foil images were also selected from the NOUN database. Following pilot work, to increase memory demands foils were chosen to match targets in shape and orientation; see the bottom two O- trials in Figure 1 for an example foil. Colors were adjusted to increase homogeneity and reduce visual distinctiveness.

All participants learned the same pairings of novel objects and words, but during each training and recall task the order of presentation for each pair was varied randomly. In order to create systematic spatial associations for the objects, objects were presented in the higher or lower section of the screen, and in the same location throughout tasks. For the mismatching object location trials, the Y co-ordinates were reversed. With a screen resolution of 1024 x 768 pixels, an initial base location was created at the mid-point of the X axis (512 pixels), and either one quarter or three quarters of the screen on the Y axis (384 and 576 pixels). To produce some variability in locations, these base coordinates were modified with Gaussian noise on the X axis (M = 0 pixels, SD = 50 pixels) and Y axis (M = 0 pixels, SD = 40 pixels).

Procedure Participants were taught to associate each of the 16 novel words with a novel object using an exposure task followed by a verbal cued recall task, repeated as a block, with each participant completing a minimum of three blocks. Training was completed if accuracy of recall was greater than 80% by the end of the third block, at which point they then completed the matching task. Otherwise, additional blocks of exposure/recall were run until the 80% accuracy criterion had been met. The order of presentation of word and object pairings were randomized per participant per task. Participants were not told that spatial location was important or to pay attention to the location.

Images were displayed on a 17" computer monitor, with each image subtending an angle of approximately 2 degrees. Words were played over headphones. In the exposure task, a fixation cross appeared for 1000 ms. Then the screen became blank and there was a variable delay, ranging from 221 ms to 779 ms (*M* = 605 ms, *SD* 142 ms), followed by presentation of the word over speakers. This delay and the word combined together equaled 1000 ms. Following word offset, the novel object was displayed in its canonical upper or lower screen location for 5000 ms.

For the cued recall task, a fixation cross appeared on screen for 1000 ms, then the novel object appeared in its canonical location for 5000 ms. Participants were instructed to verbally produce the associated word, with accuracy immediately recorded by the experimenter. After 5000 ms, feedback was provided (regardless of successful recall) by playing the correct word over headphones.

A schematic for the matching task is shown in Figure 1. In the matching task there were four blocks of 16 trials. Matching combinations of word and image stimuli appeared in four possible combinations; two combinations consisted of an object match between object and word (O+L+ and O+L-) and two of a mismatch (O-L+ and O-L-). Objects appeared in the top or bottom of the screen, which either could spatially match the trained location training (L+) or mismatch the trained location (L-; i.e., appear in the opposite vertical location). There were two counterbalanced lists, inverting the order of presentation of the matching conditions. Before starting this task participants had 12 practice trials.

Timings were the same as for the exposure task, but there was a fixed delay of 250 ms from word offset and presentation of the object on screen. Participants had to respond if this object matched the word by pressing the M key for a match and the Z key for a mismatching object. They were instructed to respond as fast and accurately as possible. RT's were calculated from onset of the object of the screen.

We also included a final location recall task to assess participants' knowledge of the trained locations. All 16 novel trained objects were presented one at a time in the center of the screen, and participants had to use the mouse to "drag and drop" the objects to where they thought the object appeared in the training phase.

Results

Two participants failed to complete the task due to experimental software error. Accuracy of matching words to objects was close to ceiling, for both when word and object matched and location of the object and the meaning of the word were the same (O+L+; *M* = 97%, *SD* = 4.4%), and for when the novel word and object matched but the location of the word was in the opposite part of the screen to the spatial meaning of the word (O+L-; *M* = 97%, *SD* = 5.6%). In analyses of errors we found no significant effects of condition or interactions.

To assess object and spatial compatibility effects on reaction times, a repeated measures analysis of variance was used with incorrect responses removed. Mean RTs are shown in Figure 2. There was no significant main effect of object match, $F(1, 25) = 2.99$, $p = .10$, though RTs for when the cue word and object matched (O+; *M* = 855 ms, *SD* = 255 ms) were higher than non-matching cases (O-; *M* = 823 ms, *SD* = 185 ms).

However, there was a significant main effect of location match. RT's were slower when the spatial meaning of the novel word's location from training mismatched the screen location of the object on screen (L-; *M* = 864 ms, *SD* = 202 ms) compared to location matching trials (L+; *M* = 815 ms, *SD* = 238 ms), $F(1, 25) = 10.94$, $p = .003$. This overall pattern was opposite to a spatial match interference effect.

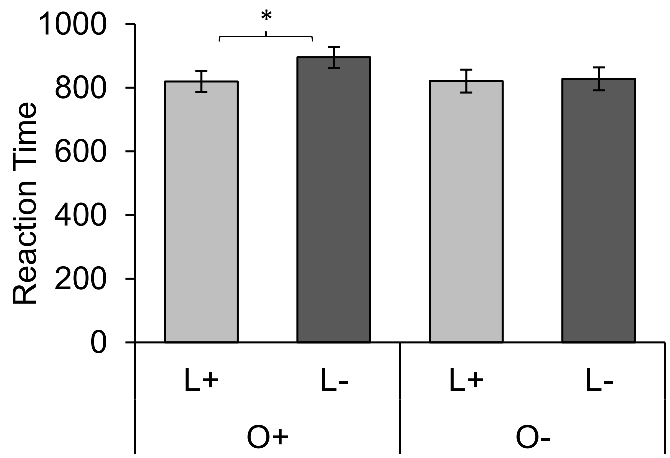

Figure 2: Reaction times to an object match and mismatch (O+ or O-) and location match or mismatch (L+ or L-). Error bars indicate within-participant 95% confidence intervals for the L+ vs. L- comparisons. (cf. Cousineau, 2005).

This main effect was qualified by a significant interaction between object and location match, $F(1, 25) = 14.26$, $p = .001$. When the cue word and object matched (O+), participants were significantly faster to assess a match when the location of the object matched the spatial meaning of the word (O+L+; *M* = 813 ms, *SD* = 225 ms) compared to when object location and the spatial meaning cue word did not match (O+L-; *M* = 899 ms, *SD* = 290 ms). $t(25) = 4.27$, $p < .001$.

In contrast, for object mismatch trails (O-), there was no significant difference between the location match (O-L+; *M* = 817 ms, *SD* = 184 ms) or mismatched (O-L-; *M* = 830 ms, *SD* = 194 ms), $t(25) = .84$, $p = .41$.

Results for the location task showed that participants had developed a strong memory of the trained location. Mean pixel offset from trained to located object was -2 pixels, and in only 2% of trials the object was placed in the opposite half of the screen to where it was trained.

Discussion

In Experiment 1 we did not find evidence consistent with the idea that perceptual simulation interfered with judgements of word-object identity. However, we also did not find evidence for the prediction of Estes et al.'s (2015) perceptual matching account that when objects mismatched but spatial codes matched (O-L+) then spatial match inference would occur. Instead, we found evidence for better performance when spatial location matched the cue word's referent's typical location but only when the object also matched.

One possible explanation of this finding is attentional priming, whereby the cue word primed the expected object location. This could facilitate matching judgements and slow down processing when the object location mismatched the cue, due to the need for an attentional shift to a new location to verify the object match. However, simple spatial priming does not explain no spatial effect for mismatching objects.

Experiment 1 was successful in its attempt to see whether novel words and novel spatial meanings could automatically influence visuo-spatial processing, though we did not find the spatial match interference seen in Estes et al. (2015). One question is whether discrepancies in our findings from theirs arise from using novel words instead of existing words.

Therefore, in Experiment 2, we wanted to directly compare these novel words and their newly learnt spatial meanings to words with pre-established spatial associations. Accordingly, Experiment 2 repeated our paradigm but included existing words and objects in the test task. As accuracy was close to ceiling in Experiment 1, we also made changes with the aim to increase the difficulty of the task to assess if object or spatial matches could influence errors as well as reaction times.

Experiment 2: Novel and Existing Words and Object Recognition

Method

Participants Twenty-six participants took part in Experiment 2 in exchange for a payment of £5. Most were psychology students at the University of Hull. None had taken part in Experiment 1.

Stimuli Items had the same design as Experiment 1, but with the addition of existing words. Sixteen existing concrete words and 16 matching images were chosen from words which had been previously normed to either have a strong up or strong down spatial association (e.g., *sun, tower, plane*; *shoes, grass, mouse*; Bergen et al., 2007). An additional 16 concrete words were selected based on ratings of being neutral in vertical association. Corresponding object images of the 32 existing words were selected from the Bank of Standardized Stimuli (BOSS; Brodeur, Dionne-Dostie, Montreui, & Lepage, 2010), along with 32 foil images, and resized to 120 x 120 pixels. The names for these 32 existing objects were recorded by the same male native English speaker from Experiment 1.

Procedure Training was the same as Experiment 1, but with criterion of accuracy in the verbal recall reduced to 75%**.** Only the novel word-object pairs were trained. We introduced a distraction task between training end and the matching task, where participants had to watch a five minute clip from a BBC wildlife documentary over headphones.

In the matching task, as with Experiment 1 there were two counterbalanced lists and 4 blocks. However, each block consisted of 32 trials containing both novel objects or existing objects and words. This task was the same for both novel and novel words of keyboard presses to indicate if the object matched or mismatched the verbal cue word.

Results and Discussion

Although we made some modest changes from Experiment 1 aiming to increase difficulty to reduce ceiling effects, accuracy was again close to ceiling in Experiment 2 for object matching trials, (M = 96%, SD = 6.6%) as was the accuracy rejecting object mismatching trials (M = 98%, SD = 3.1%), with no significant effects of errors.

A 2 x 2 x 2 ANOVA was performed with object match (O+ vs. O-), location match (L+ vs. L-) and word type (existing vs. novel word), and mean results are shown in Figure 3.

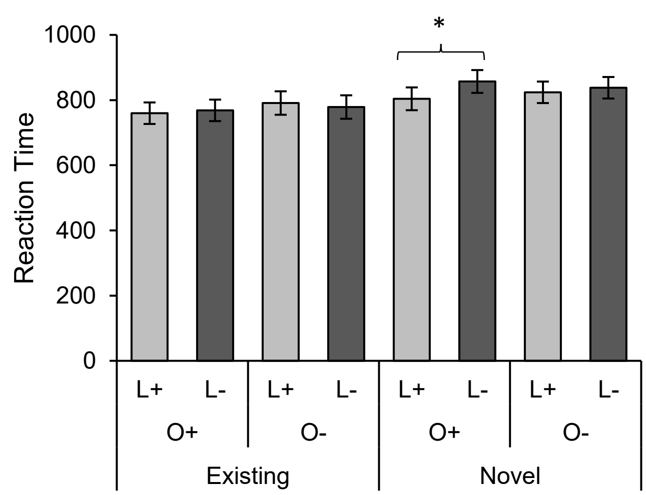

Figure 3: RTs for Existing and Novel words for object match and location matches. Error bars indicate within-participant 95% confidence intervals for the L+ vs. L- comparisons.

There was a significant main effect of word type $F(1, 25) = 14.79$, $p = .001$. Overall, existing words (M = 797 ms, SD = 162 ms) were faster to be correctly matched with their corresponding objects than novel words (M = 808 ms, SD = 180 ms).

Like Experiment 1, there was no significant main effect of object match, $F(1, 25) = .79$, $p = .39$, and no interaction of object match and word type, $F(1, 25) = .99$, $p = .33$.

There was, however, a significant main effect of location match $F(1, 25) = 4.97$, $p = .035$, with mean RT faster (M = 795 ms, SD = 172 ms) when the spatial meaning of the word and the object location matched (L+) compared to a mismatch between location and spatial meaning (L-; M = 810 ms, SD = 167 ms). This indicated an overall benefit on object matching performance when the location of the object matched the meaning of the cue word. This effect was qualified by a marginal interaction between location match and word type, $F(1, 25) = 3.96$, $p = .058$.

Further explorations of this effect were tested in additional ANOVAs separating word type. For existing words, there was no significant main effect of location match, $F(1, 25) = .029$, $p = .866$, no significant interaction effect between location match and object match, $F(1, 25) = .577$, $p = .455$, and no significant pairwise comparisons.

However, as with Experiment 1, for novel words there was a main effect of location match. When the location of the object matched the spatial meaning of the cue word (L+) reaction times were significantly faster ($M = 814$ ms, $SD = 172$ ms) than when the spatial meaning of the cue word did not match (L-; $M = 847$ ms, $SD = 166$ ms), $F(1, 25) = 8.34$, $p = .008$. Unlike Experiment 1, there no significant interaction between location matching and object matching, $F(1, 25) = 2.484$, $p = .128$. However, pairwise comparisons showed the same pattern as Experiment 1 of significantly faster RTs for spatial matches on object matching trials, $t(25) = 3.09$, $p < .01$, $d = .61$, but no significant effect of location on object mismatching trials ($p = .42$).

The findings with novel words replicates our key results from Experiment 1, but we found null effects of spatial match for the existing words. In both Experiments 1 and 2, we did not find spatial match interference effects using this paradigm. While we used an object matching task in Experiment 1 and 2, involving a direct link between the cue and target objects, other studies have used a more low-level target discrimination task with semantically unrelated targets (e.g., Estes et al., 2008, with discrimination between a X or O).

In Experiment 3, we used the same training regime for novel words but used a target discrimination task at test. Here there was no relationship between the object learned and the target, but there was an implicit spatial relationship based on the match between the training location and target location (i.e., coding of O-L+ or O-L0). We wanted to see if there were faster responses to matching locations as might be expected from spatial priming, or if a match interference effect occurred leading to slower response times, consistent with previous studies using existing words and sentences.

Experiment 3: Novel Words and Target Discrimination

Method

Participants Thirty participants were recruited from the undergraduate psychology program at the University of Hull on. All participants had normal or corrected to normal vision

Stimuli and Design The same 16 words and objects were used from Experiment 1, along the with same training paradigm. Targets were the symbols ㄹ and 5, 11 x 11 pixels in size, and were the exact mirror image of each other.

Procedure Shortly after the training phase was completed participants completed the target discrimination task. A fixation cross appeared in the middle of the screen for 1000 ms between trials. Words were heard over headphones while participants viewed a white screen. Four hundred ms after word offset a target appeared on the screen. Targets appeared in the same coordinates as the associated object for that cue word, or in the reverse Y coordinates. Participants were instructed to identify the target as quickly and accurately as possible by pressing the 2 or 5 key on the number row of the keyboard. Responses timed out after 1500 ms. There were 32 trials, with each novel word heard twice, with a target appearing once in its trained location and once in the opposite location.

Results

Accuracy for L+ trials was 92.4% and 91.7% for L- trials. A paired samples t-test on correct trials found there was a significant difference between matches of training location and target location on reaction times, $t(29) = 7.473$, $p < .001$, $d = 1.36$. As can be seen Figure 4, responses were slower when the target matched the trained location, for both words associated with objects in the top or bottom of the screen.

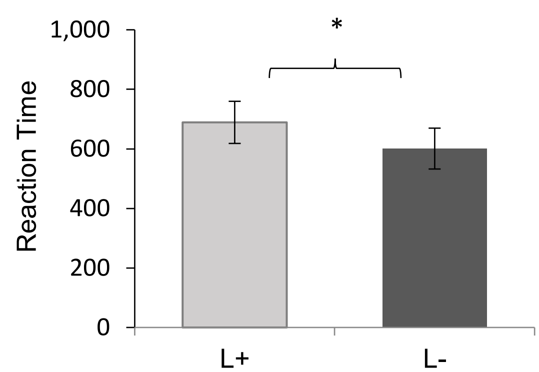

Figure 4: Target discrimination RTs for training location and target location matches and mismatches. Error bars indicate within-participant 95% confidence intervals for the L+ vs. L- comparisons.

Discussion

Experiment 2 extended the findings of Experiment 1 by comparing responses to novel words-object pairs with recently learnt visuo-spatial associations with objects with pre-existing spatial associations, formed from a long history of world interactions. We replicated the findings of Experiment 1, revealing that visuo-spatial associations activated by novel words could interfere with object matching, even when the spatial association was irrelevant to task performance. These findings show that automatic activation of spatial word semantics still occurs with recently learnt words. However, we did not find automatic spatial interference or facilitation effects for existing words.

Using novel words in Experiment 3, we found the opposite results, with interference on spatially matching trials. The findings of Experiment 3 serve as additional empirical support for the pattern of spatial match interference using a target discrimination task following presentation of spatially associated language, alongside previous results in the literature using visual target discrimination tasks (e.g., Bergen et al., 2007; Estes et al., 2008).

Unlike Estes et al. (2015), we did not find spatial match interference effects with mismatching objects in Experiment 1 and 2. As with their study, we combined both matching and mismatching objects in a single task. However, there were some differences in the tasks other than the use of novel words. Their task was to discriminate between real objects (e.g. *bird* or *wrench*) and nonsense line-drawings, and the preceding linguistic cue was not relevant to the task, though presumably participants were aware that in some cases the object matched the cue and some cases it did not. Here, the task was to directly report if the visual object matched the cue. It is unclear if and why this difference may have reduced the likelihood of finding spatial match interference, especially since one might expect that explicitly asking for object matching may boost object matching effects. In order to better understand the discrepancy, a more direct replication of their study could use a similar task to Estes et al. with novel and existing words.

Another source of uncertainty for explaining the match task results of Experiments 1 and 2 is why we did not find spatial match interference with existing words or faster responses for congruent spatial locations for the existing words, while we did find them for the novel words. One possible explanation is that a consequence of training novel objects occurring at specific locations is that these recently acquired episodic visuo-spatial representations should be both strong and specific, given sufficient training. In contrast, hearing the word "*bird*" outside of a constraining context could trigger very heterogeneous and individual representations, with comparatively less spatial association strength and/or specificity in visual representations. However, the heterogenous spatial semantics of existing words does appear to be sufficiently strong to cause effects with existing words in other work (e.g., Estes et al., 2015).

One methodological issue is that in Experiment 1 and 2 the mismatching objects were all unseen foils, as we did not want to dilute the spatial associations built in the training tasks during the test task. This meant that participants could potentially respond based on familiarity and not necessarily on language-based match. Experiment 3 did not have this issue and therefore shows a clearer influence of novel language semantics, but to unconfound this, future studies using the matching task could use the familiar novel objects as foils as well as targets.

In summary, by using novel words and novel objects we were able partially replicate and extend other work showing the influence of language on perceptual processing. Neither theoretical accounts of perceptual simulation or perceptual matching were able to fully explain the set of findings here. Future work aims to better understand potential mechanisms by using eye tracking to assess the contribution of reflexive attention to facilitation and interference effects, along with more direct tests of the perceptual interference explanation by using a more graded manipulations of object match within a single experiment.

Acknowledgments

Research in paper was submitted as part of Robert Nightingale's Masters in Research and Tegan Rymer's BSc Psychology thesis at the University of Hull.

References

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology, 59*(1), 617-645.

Bergen, B. (2007). Experimental methods for simulation semantics. Methods in Cognitive Linguistics, 277-301.

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.

Brodeur MB, Dionne-Dostie E, Montreuil T, Lepage M (2010) The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLoS ONE 5*(5): e10773.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 4–45.

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.

Estes Z., Verges M. & Barsalou, L. (2008). Head up, Foot down: Object Words Orient Attention to the Objects Typical Location. *Psychological Science, 19(*2), 93-97

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods, 48*(4), 1393-1409.

Lu, C. H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic bulletin & review*, *2*(2), 174-207.

Perky, C. W. (1910). An experimental study of imagination. *The American Journal of Psychology, 21*(3), 422-452.

Santiago, J., & Lakens, D. (2015). Can conceptual congruency effects between number, time, and space be accounted for by polarity correspondence?. *Acta Psychologica, 156,* 179-191.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168-171.