Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Hearing “moon” and looking up: Word-related spatial associations facilitate saccades to congruent locations

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN

1069-7977

Authors

Dunn, Ben
Kamide, Yuki
Scheepers, Christoph

Publication Date

2014

Peer reviewed

Hearing “moon” and looking up: Word-related spatial associations facilitate saccades to congruent locations

Ben M. Dunn (b.dunn.1@research.gla.ac.uk)
Institute of Neuroscience and Psychology, University of Glasgow
58 Hillhead Street, Glasgow
G12 8QB UK

Yuki Kamide (y.kamide@dundee.ac.uk)
School of Psychology, University of Dundee
DD1 4HN UK

Christoph Scheepers (christoph.scheepers@glasgow.ac.uk)
Institute of Neuroscience and Psychology, University of Glasgow
58 Hillhead Street, Glasgow
G12 8QB UK

Abstract

In the experiment reported here, 30 participants made a lexical decision on 120 spoken words and 120 spoken non-words. The words had either an upward (e.g. ‘moon’) or downward (e.g. ‘sewer’) spatial association, or they were neutral in this respect (e.g. ‘letter’). Participants made their lexical decisions by fixating a target located either above or below the centre of the screen, counterbalanced across participants. Saccade launch latencies to targets in a congruent spatial location (e.g., hearing ‘moon’ and looking up to confirm that the stimulus is a word) were significantly faster than those to targets in an incongruent location (e.g., hearing ‘moon’ and looking down to confirm that it is a word). Crucially, saccade launch latencies to incongruent target locations did not differ from those launched after hearing neutral words. Our results extend earlier findings (Dudschig et al., 2013) by showing that language-related spatial associations *facilitate* eye movements towards congruent locations rather than inhibiting eye movements towards incongruent locations.

Keywords: Language processing; Perceptual-spatial representation; Mental simulation; Embodied cognition.

Introduction

Dante Alighieri wrote, “Heaven wheels above you, displaying to you her eternal glories, and still your eyes are on the ground” (Sisson, 1993, p. 261). If we think of heaven as above us, should we not look up towards it? Certain concepts activate perceptual-spatial locations; more specifically, words such as ‘push’ or ‘pull’ seem to be mentally represented on a horizontal axis whereas ‘float’ and ‘sink’ are more likely to activate vertical mental representations (Richardson, Spivey, Barsalou, & McRae, 2003). To use an example, the concept associated with the word *moon* is likely to entail an upward spatial representation due to our experience that when talking about the moon, we are often looking up towards it, or about to look up towards it, respectively. Similarly, the word *sewer* is more likely to invoke a downward mental representation as sewers usually tend to be on the ground below our line of sight.

Recent research has shown that the perceptual-spatial representation of words can have either facilitatory or inhibitory effects on processing. For instance, Richardson et al. (2003) showed that verbs with horizontal or vertical associations can *inhibit* participants’ ability to detect a visual target in a location compatible with the spatial axis implied by the verb. Specifically, upon hearing *The ship sinks in the ocean*, participants took longer to detect a small abstract target (a black circle or square) when it appeared on the top or bottom of the screen compared with the left or right. Similar results were shown by Estes, Verges and Barsalou (2008) who visually presented a context word (e.g. *cowboy*), then a noun associated with either an upper or lower location (e.g. *hat* or *boot*) on the centre of the screen, and after a 50ms delay, an *X* or *O* located at the top or bottom of the screen. Participants’ task was to identify the letter by pressing a key. The results showed shorter reaction times (RTs) when the letter was located in an incongruent compared with a congruent spatial location (e.g., seeing *cowboy* then *hat* and identifying a target at the bottom of the screen, compared with identifying a target at the top of the screen). Attempts to explain such inhibitory effects are often based on the amount of perceptual-featural overlap (in this case, the spatial location) between the direction word cue and the location of the target symbol or letter to be identified. In tasks such as those described above, the cue words direct attention to a congruent spatial location (i.e. *cowboy hat* would direct attention upwards), hence to identify a target in an ‘up’ location requires inhibition of the spatial traces activated by the cue, resulting in longer RTs. Conversely, in a similar scenario a target in a ‘down’ location requires no inhibition of the spatial traces activated by *cowboy hat* and therefore results in shorter RTs.

In stark contrast, other studies reported facilitatory effects of perceptual-spatial representation on processing. For example, Stanfield and Zwaan (2001) presented participants with a sentence implying a horizontal or vertical orientation of an object (e.g., *He hammered the nail into the* {*wall* or *floor*}) followed by a congruent or incongruent depiction of the object (a nail laying flat or standing up). When asked if the visual target had been mentioned in the sentence, participants responded faster when the cued orientation matched that of the target. Moreover, Zwaan, Stanfield and Yaxley (2002) showed that participants were quicker to identify a picture of an eagle when its shape (wings outstretched, compared with wings folded) was compatible with the cue sentence (*The ranger saw the eagle* {*in the sky* or *in its nest*}). Finally, Zwaan and Yaxley (2003) found facilitatory effects of perceptual-spatial representations using a semantic relatedness task. Specifically, when two words were presented on the computer screen in line with the real-world arrangement of the objects they referred to (e.g. the word *branch* displayed above the word *root*), participants were faster to provide a relatedness judgement than when the order was reversed (*root* displayed above *branch*). The same was found with abstract concepts such as *master* and *slave* (Schubert, 2005).

The conflicting results (facilitation on the one hand, inhibition on the other) may be explained by assuming that activating the spatial location of a concept may interfere with responses to abstract targets that do not possess any 'inherent' spatial locations (i.e. an *X* or a small circle has no a-priori association with either 'up' or 'down'). Conversely, when the spatial location is implied and a target with visual features overlapping with the implied location is presented, responses are facilitated (e.g. the words 'eagle in the sky' followed by a picture of an eagle with outstretched wings). However in order to reconcile the conflicting results, further empirical and theoretical analysis is required.

Embodied theories of language processing postulate that concepts are understood via a process of perceptual simulation (Barsalou, 1999, 2008; Glenberg & Kaschak, 2002; Zwaan, 2004), whereby the mental representation of a linguistically cued concept activates experiential traces. The experiential traces combine all aspects of an individual's knowledge surrounding that concept; hence reactivation by something related to the concept (e.g. perceptual-spatial location) affects the outcome behaviour. In the research quoted above, abstract target detection seems to require *inhibition* of the initially activated experiential traces, therefore slowing down response times; whereas targets that reactivate experiential traces due to their visual appearance (e.g. shape, visual form) facilitate target detection and hence speed up response times.

Recently, saccadic eye movements have been used to determine the effects of perceptual-spatial representations during language processing. Using a lexical decision task based on eye movements, Dudschig et al. (2013) showed that participants were quicker to launch a saccade towards a 'yes' target located at the top of the screen after reading a centrally presented word like *sun* (associated with an upward spatial location) than after reading a word like *shoe* (associated with a downward spatial location); the reverse was true when the 'yes' target was at the bottom of the screen. This is an interesting finding particularly because the saccadic lexical decision task appears to tap directly into associations between word meanings on the one hand and motor responses by the visual system on the other.

However, note that Dudschig et al. (2013) chose visual presentation of their linguistic stimuli, which might have affected the resulting vertical eye movements to some extent due to variations in saccade starting position as a result of reading. More importantly, their experiment did not include a baseline condition. Note that without the latter, it is actually not possible to determine whether words like *sun* or *shoe* facilitate saccades to congruent target locations, inhibit saccades to incongruent target locations, or both. Contrary to Dudschig et al.'s (2013) own conclusions, it may actually be the case that perceptual-spatial traces inhibit saccades to incongruent spatial locations (making it more difficult to launch a downward saccade upon reading *sun*) rather than facilitating saccades to congruent spatial locations (making it easier to launch an upward saccade upon reading *sun*). Given that previously reported research has shown conflicting results in terms of whether word-related perceptual-spatial traces facilitate or interfere with the processing of a spatial target, we aimed to clarify whether 'direction words' such as *moon* and *sewer* facilitate or inhibit saccades towards congruent versus incongruent target locations. To achieve this, we introduced a baseline condition which consisted of words that do not evoke any particularly strong spatial association in the vertical dimension (e.g., *letter*).

Current Study

Similar to Dudschig et al. (2013), the present study employed an eye movement activated lexical decision task. Participants had to indicate whether a spoken word candidate was an actual word of the English language or not, by looking at either a 'yes' or a 'no' target presented above or below the centre of the screen (the latter was marked by a cross). The relative positioning of the 'yes' target (either above or below the centre of the screen) was counterbalanced across participants. Apart from using a different set of word and non-word stimuli and a different language (English rather than German), our experiment differed from the study by Dudschig et al. (2013) in two major respects. First, our stimuli were presented in the auditory modality (i.e., participants listened to spoken word candidates via headphones) as this was deemed to produce less interference with the visual response required by the task. Second, in addition to words with independently attested upward (e.g. *moon*) and downward (e.g. *sewer*) spatial association, we also included words without any particular association in the vertical dimension (e.g. *letter*). The latter formed our baseline condition for comparison. In the present design, *facilitatory* effects of perceptual-spatial word associations should manifest themselves in faster saccade onset latencies (relative to the baseline) whenever a given word's spatial association is congruent with the required visual response (i.e., hearing *moon* and looking up to say 'yes' or hearing *sewer* and looking down to say 'yes', respectively). Conversely, *inhibitory* effects would become apparent in slower saccade onset latencies (again, relative to the baseline) whenever spatial associations are incongruent with the required response (hearing *moon* and looking down to say 'yes' or hearing *sewer* and looking up to say 'yes', respectively).

Method

Participants Thirty individuals (22 Female; M=24.2 years) from the University of Glasgow participated in the study, each receiving £4 or course credits. All participants had either normal or corrected-to-normal vision and were native English speakers.

Materials One hundred and twenty words were chosen as linguistic stimuli. There were 40 'up' (e.g. *moon*), 40 'down' (e.g. *sewer*) and 40 'neutral' (e.g. *letter*) words, as determined by a pre-test (see below). The 'up' and 'down' stimuli each consisted of 20 verbs, 12 nouns and 8 adjectives whilst the 'neutral' condition had 20 verbs, 11 nouns and 9 adjectives. Across conditions, the words were matched on lexical frequency, number of syllables, and number of phonemes as determined by the MRC Psycholinguistic Database (Coltheart, 1981). To control for concreteness, we asked 38 participants to rate each word on a scale of 1 (very abstract) to 7 (very concrete). Mean concreteness ratings did not differ across conditions (all *p*s >.7 by between-item *t*-tests). There were also no cross-condition differences in lexical decision times for written instances of the words (norms from Balota, et al., 2007).

All words, as well as 120 non-word fillers (see below), were recorded as separate sound files using a computer generated male British-English voice ('Brian', implemented in IVONA Reader software)[1]. Word stress was controlled so that each word had a steady tone with no rising or falling intonation. Spoken durations for 'up' words (M = 708 ms, SD = 129 ms), 'down' words (M = 721 ms, SD = 160 ms), and 'neutral' words (M = 710 ms, SD = 156 ms) did not differ reliably from one another (*p*s > .6). The 120 non-word fillers were pronounceable psuedowords constructed from novel composites of existing English phonemes (e.g. *asteng*). Each sound file had the volume normalised to -6dB (peak level) using Sound Studio (Felt Tip Software).

Spatial Association Norming An internet-based rating study was conducted to verify the intended spatial associations per condition. Participants rated 402 English candidate words for vertical association on a Likert scale ranging from −5 to +5 (see below). The words were split into 15 lists (each seen by at least 21 subjects), with 25-30 items per list. Underneath each printed candidate word, there was an 11-point bipolar scale on which participants had to provide their spatial association ratings. The leftmost point on the scale (scored as −5) was labelled "down" (for downward association), the rightmost point (scored as +5) was labelled "up" (for upward association), and the midpoint (scored as 0) was labelled "neutral" (for no vertical association). Participants also marked a word as 'known' if they were familiar with the word or 'unsure' if they were not. Eleven cases (0.1%) with 'unsure' ratings were removed from analysis. The mean rating for the final selection of 'up' words was +3.65 (N = 40; min. +3.00; max. +4.36); the 'neutral' words scored an average of +0.03 on the scale (N = 40; min. −0.27; max. +0.43); finally, the 'down' words had an average rating of −3.48 (N = 40; min. −4.48; max. −2.82).

Apparatus The stimuli were presented on a 21-inch CRT monitor of a DELL Optiplex GX 720 desktop computer with a display refresh rate of 85 Hz. Chin and forehead rests, positioned at a distance of 70 cm from the screen, were used to minimise head movements. Participants' eye movements were continuously monitored using a desk-mounted SR Research EyeLink 1000 eye-tracker, sampling at 1000 Hz. Although viewing was binocular, only the dominant eye was tracked, as established by a variation of the Miles test (Miles, 1930; Roth, Lora, & Heilman, 2002). Stimulus presentation and data collection were controlled using Experiment Builder software (SR Research).

Procedure Each participant was presented with 240 auditory stimuli (120 words and 120 non-words) in an individually determined random order. As shown in Figure 1, each trial began with the presentation of a central fixation cross for drift correction while the participant kept looking at the cross. 150 ms after drift correction, the sound file was played via headphones and at the same time, a green and a red square appeared on screen. Each square measured 10 × 10 screen pixels and appeared 8^o above and below the central fixation cross, respectively. The participant's task was to decide, as quickly and accurately as possible, whether what they just heard was an actual English word or not by looking at either the green square (if they thought they heard a word) or the red square (if they thought they heard a non-word). The location of the red and green square was counterbalanced across participants; 15 participants had the red square at the top and the green square at the bottom (and vice versa for the remaining 15 participants) for all 240 trials. This between-subject manipulation lowered the chances of participants figuring out the purpose of the study. Each trial terminated when a fixation was detected in one of the target areas (dashed rectangles in Figure 1), or after a timeout of 3000 ms, respectively. The target areas for the trial-terminating gaze trigger were defined as the inside edge of the coloured square to the top or bottom edge of the screen (200 pixels, 5.5^o) and were 800 pixels (22^o) wide. Before the first trial and after every 40 subsequent trials, the eye-tracker was recalibrated and validated using a 9-point fixation procedure. An experimental session lasted approximately 40 minutes.

---

[1] See http://www.ivona.com/en/reader

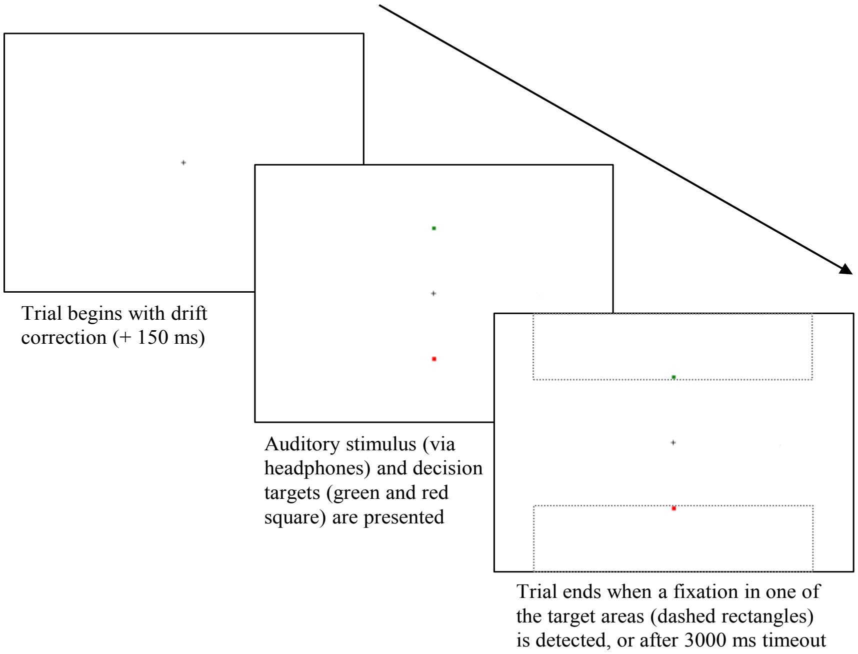

Figure 1: Schematic of a single trial
The green (for ‘yes’) and red (for ‘no’) squares represent the targets for lexical decision; their vertical positioning was counterbalanced across participants

Data Analysis Lexical decision accuracy was greater than 95% in each condition. Only the critical word trials were considered for analysis. The dependent variable of interest was the saccade launch latency for correct ‘word’ decisions (saccade towards green square), measured from the onset of the auditory stimulus presentation until the eye started moving away from the central fixation cross (as determined by saccadic acceleration and velocity thresholds). Trials with multiple saccades, containing eye-blinks, saccades not landing within 100 pixels of the green target, saccades launched in the incorrect direction (i.e. towards the red square), or trials that terminated after 3000 ms timeout, were removed; this affected ca. 10% of the critical trials. Saccade launch latency outliers of more than 2.5 SDs away from the mean of a given condition per participant were also removed from further analysis (affecting less than 3% of the data).

Statistical analyses were performed in SPSS 20 using Generalized Estimating Equations (GEE; e.g. Hardin & Hilbe, 2003). GEE allows for more accurate modelling of skewed data compared with ANOVA. Since saccade launch latencies (like RT distributions in general) tend to be positively skewed, we used a *gamma* distribution and *log* link function in the GEE model specifications. Two types of analyses were performed, with the requirement that both should yield a significant result (at $p < .05$) for an effect to be considered statistically meaningful. In the *by-subject* analysis, word direction (‘up’, ‘neutral’, ‘down’) was entered as within-subjects factor and saccade direction (‘upwards’, ‘downwards’) as between-subjects factor. In the *by-item* analysis, word direction was between- and saccade direction within-items. All analyses assumed an exchangeable covariance matrix for repeated measurements.

Results

The descriptive data are shown in Figure 2, and Table 1 summarizes the inferential results from GEE modelling (GSχ^2 refers to the *Generalized-Score Chi-Square* statistic).

Table 1. Results from a log gamma GEE analyses modelling saccade launch latencies as a function of word direction (W) and saccade direction (S).

Effect		**By Subjects**		**By Items**	
	df	**GS χ^2**	***P***	**GS χ^2**	***P***
Word Direction (W)	2	10.38	< .01	2.43	0.30
Saccade Direction (S)	1	1.87	0.17	50.75	<.001
W × S Interaction	2	13.11	< .01	29.85	<.001

The main effect of word direction (ca. 28 ms higher saccade launch latencies in the ‘neutral’ condition compared to the other word direction conditions) was significant within-subjects but not between-items (presumably due to reduced power in the latter case). Likewise, the main effect of saccade direction (ca. 50 ms higher saccade launch latencies for downward than for upward saccades overall) was significant within-items but not between-subjects (again, suggesting reduced power for the ‘between’ factor). Note that previous research (including Dudschig et al.,

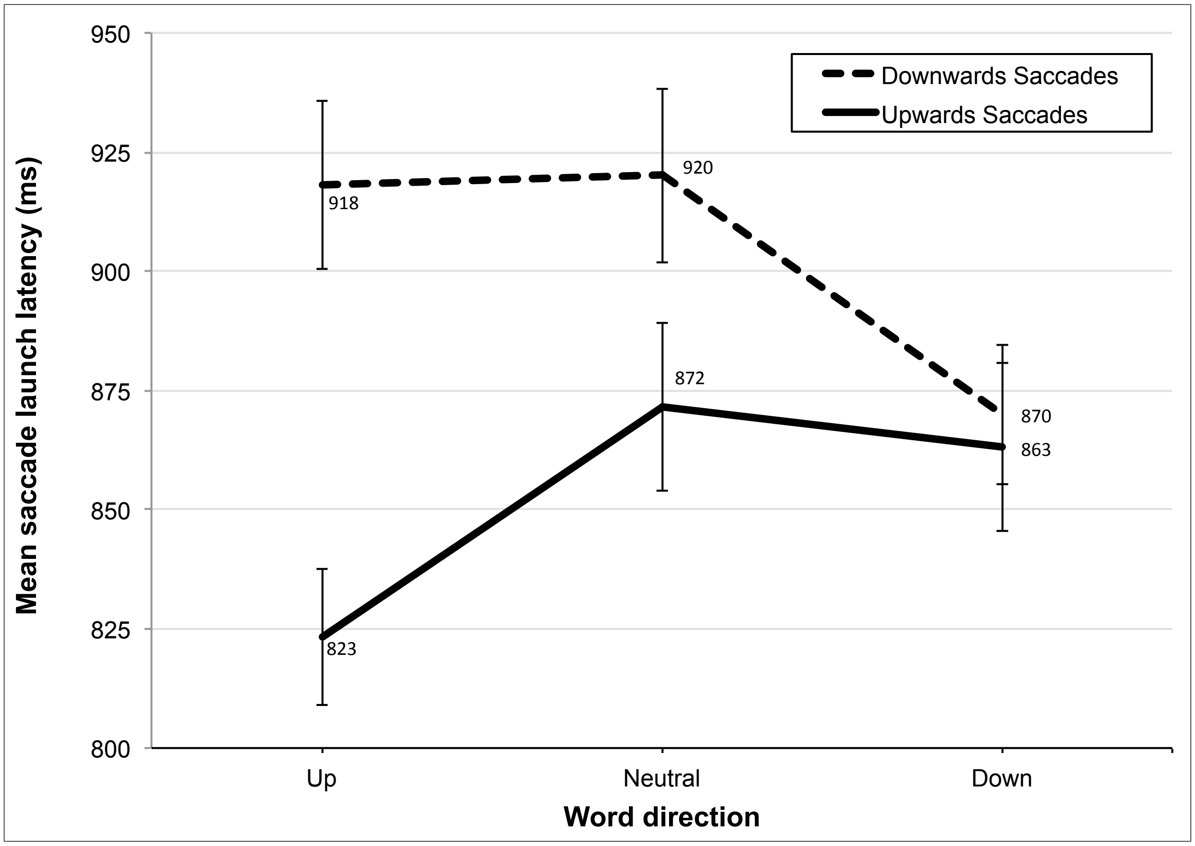

Figure 2: Mean saccade launch latencies per condition (in ms), relative to spoken word onset (the words ended ca. 715 ms after word onset on average). Error bars represent 95% confidence intervals for the means derived from the by-subject analysis.

2013) has shown a similar general disadvantage for downward saccades.

Most crucially, there was a significant word direction × saccade direction interaction in both the by-subject and the by-item analysis. 95% CIs for simple effects showed that **upward** saccades were launched more quickly upon hearing 'up' words like *moon* than upon hearing 'neutral' words like *letter* (by-subject contrast: 49 ± 22 ms; by-items: 54 ± 44 ms); the comparison between 'down' words like *sewer* and 'neutral' words like *letter* was not significant (by-subjects: 7 ± 17ms; by-items: 12 ± 46ms). Conversely, **downward** saccades were launched quicker after 'down' words like *sewer* compared to neutral words like *letter* (by-subjects: 50 ± 32 ms; by-items: 51 ± 44 ms), whereas the contrast between 'up' words like *moon* and 'neutral' words like *letter* was not significant (by-subjects: 2 ± 21 ms; by-items: 7 ± 47 ms).

Discussion

Using an eye movement activated lexical decision task, the present experiment investigated how perceptual-spatial representations associated with words affect launch latencies for saccades towards congruent or incongruent spatial locations. The results clearly showed that 'direction' words facilitate saccades towards congruent locations in the vertical dimension, but crucially, do not inhibit saccades towards incongruent locations. These results add to a growing body of literature suggesting that perceptual-spatial representations are automatically activated upon hearing relevant linguistic cues.

The present findings replicate the results from Dudschig et al. (2013), but using different stimuli, a different language (English rather than German), a different modality (spoken rather than written words) and most importantly, a baseline condition that allowed for distinguishing between facilitatory and inhibitory effects of language on eye movements in the vertical dimension. The Dudschig et al. (2013) findings were ambiguous as to whether 'up' words like *moon* and 'down' words like *sewer* facilitate saccades in the congruent direction, inhibit saccades in the incongruent direction, or both. The present results indicated, via comparison with the baseline, that saccades are *facilitated* when cued by words whose perceptual-spatial associations are congruent with the direction of the required saccadic response for lexical decision.

The question remains, why do compatible direction words facilitate saccades? Previously, facilitation was explained as the result of featural overlap between the cue and target –

hence a lack of featural overlap led to inhibitory effects (Estes et al., 2008). In the present investigation, there is at least one aspect of the target that might overlap with the cueing word, namely the vertical direction that leads to the target location. From an embodied cognition point of view, experiential traces associated with a concept become reactivated upon later presentation (Zwaan, 2004). Hearing the word *moon* would reactivate all experiential traces of the related concept (including perhaps, looking up to see the moon), and therefore, congruent visuomotor responses (saccading upwards) should be facilitated. By contrast, if a given word's vertical association is incongruent with the direction of the required saccadic response, then its influence on saccade launch latency is no different from that of a vertically 'neutral' word like *letter*. This suggests that experiential traces associated with words would not interfere with, or inhibit, incongruent saccadic responses.

What about the inhibitory effects shown by some of the studies discussed in the introduction (Estes et al., 2008; Kaschak, et al., 2005; Richardson et al., 2003)? Note that these studies did not record eye movements, making it difficult to compare with our results. As mentioned, further empirical work is necessary to determine the cause of the inhibitory effects, however the present study has gone someway to showing that the effects previously reported are not due to a lack of featural overlap between the directional cue and the abstract target.

In conclusion, our results have helped to establish the facilitatory role of word-related spatial associations on saccadic eye movements. Our findings confirm that words automatically activate associated perceptual-spatial representations. This supports the view held by embodied cognition theories that word-related concepts are grounded in perception and action (Barsalou, 1999, 2008; Glenberg & Kaschak, 2002; Zwaan, 2004). Furthermore, our results have shown that the perceptual-spatial overlap between direction words and abstract targets is facilitatory when preparing to saccade to a compatible location – this is despite the lack of other perceptual-featural properties (e.g. visual form) overlapping with the directional concepts. The results add to the growing debate surrounding the embodied view of language processing.

Acknowledgments

We acknowledge support from ESRC Grant RES-062-23-2842 awarded to YK and CS. The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

References

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445-459.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577-660.

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

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33*(4), 497-505.

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.

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.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558-565.

Hardin, J. & Hilbe, J. (2003). *Generalized Estimating Equations*. London: Chapman and Hall/CRC.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition, 94*(3), B79-B89.

Miles, W. R. (1930). Ocular dominance in human adults. *The Journal of General Psychology, 3*(3), 412-430.

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.

Roth, H. L., Lora, A. N., & Heilman, K. M. (2002). Effects of monocular viewing and eye dominance on spatial attention. *Brain, 125*(9), 2023-2035.

Schubert, T. W. (2005). Your highness: vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology, 89*(1), 1-21.

Sisson, C. (1993). *Dante Alighieri: The Divine Comedy*. Oxford, UK: Oxford University Press.

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.

Zwaan, R. A. (2004). The immersed experiencer: toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 44, pp. 35-62). New York: Academic Press.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science, 28*(4), 611-619.

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

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review, 10*(4), 954-958.