COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 43 (2019) e12742

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12742

Affordance Compatibility Effect for Word Learning in Virtual Reality

Chelsea L. Gordon, Timothy M. Shea, David C. Noelle, Ramesh Balasubramaniam

Cognitive and Information Sciences, University of California, Merced, CA, USA

Received 16 April 2018; received in revised form 16 April 2019; accepted 23 April 2019

Abstract

Rich sensorimotor interaction facilitates language learning and is presumed to ground conceptual representations. Yet empirical support for early stages of embodied word learning is currently lacking. Finding evidence that sensorimotor interaction shapes learned linguistic representations would provide crucial support for embodied language theories. We developed a gamified word learning experiment in virtual reality in which participants learned the names of six novel objects by grasping and manipulating objects with either their left or right hand. Participants then completed a word–color match task in which they were tested on the same six words and objects. Participants were faster to respond to stimuli in the match task when the response hand was compatible with the hand used to interact with the named object, an effect we refer to as affordance compatibility. In two follow up experiments, we found that merely observing virtual hands interact with the objects was sufficient to acquire a smaller affordance compatibility effect, and we found that the compatibility effect was driven primarily by responses with a compatible hand and not by responses in a compatible spatial location. Our results support theoretical views of language which ground word representations in sensorimotor experiences, and they suggest promising future routes to explore the sensorimotor foundations of higher cognition through immersive virtual experiments.

Keywords: Embodied language; Virtual reality; Action observation

1. Introduction

1.1. Action/sensory language is grounded in sensorimotor processes

Embodied language theories propose that linguistic representations are grounded in sensorimotor experiences: that words evoke sights, sounds, and movements in the mind, and

---

Correspondence should be sent to Ramesh Balasubramaniam, Cognitive and Information Sciences, University of California, Merced, 5200 Lake Road, Merced, CA 95343, USA. E-mail: ramesh@ucmerced.edu

those features constitute the representations of the words (Barsalou, 1999, 2008; Gallese & Lakoff, 2005; Glenberg, 1997; Glenberg & Kaschak, 2002). In many cases, theories of language embodiment have focused on how sensorimotor experiences are reactivated and recombined in mental simulations (Barsalou, 2009; Zwaan & Madden, 2005). Simulations are thought to constrain ongoing sensorimotor processes, perturbing the actions and perceptions of language listeners. Furthermore, interactions between the sensorimotor foundations of language and real time sensorimotor processes are bidirectional; thus, concurrent motor and perceptual states bias the comprehension and production of language. Evidence supporting a general view of language as situated and embodied extends from embodied spatial language (Spivey, Tyler, Richardson, & Young, 2000) to emotional language (Glenberg, Havas, Becker, & Rinck, 2005; Havas, Glenberg, & Rinck, 2007) and abstract language (Casasanto & Boroditsky, 2008; Matlock, Holmes, Srinivasan, & Ramscar, 2011).

In a seminal finding, Glenberg and Kaschak (2002) demonstrated how language processing can influence seemingly unrelated aspects of action. In an effect referred to as the action-sentence compatibility effect, participants heard sentences with implied motion away from (“She closed the drawer”) or toward a protagonist, and judged whether the sentences were sensible by pulling a lever toward or away from themselves. When the direction of the response was the same as the implied motion of the sentence, participants made sense of the sentences more quickly. The speed at which a response could be prepared and executed was influenced by the semantics of the sentence. This is consistent with a view that sentence meaning is understood by how, in terms of body and environment, the actions in the sentence are accomplished.

In addition to the substantial evidence indicating *that* language is grounded in sensorimotor experiences, numerous studies show *how* language generates patterns of activity in the nervous system which correspond to sensorimotor experiences. For instance, olfactory areas are activated by words associated with smells (González et al., 2006), and sound-related words like “ringing” activate auditory regions more than non-sound-related words (Kiefer, Sim, Herrnberger, Grothe, & Hoenig, 2008). Many studies have shown that action words and action sentences activate somatotopic regions of the motor cortex, such that “kick” recruits leg area of vPMC and “lick” recruits the face area (Buccino et al., 2005; Hauk, Johnsrude, & Pulvermüller, 2004; Tettamanti et al., 2005), and this differential activation has been found to occur as early as 200 ms after word onset (Hauk & Pulvermüller, 2004), suggesting involvement of these areas in early semantic processing. Not only does language perception evoke sensorimotor activity, non-invasive brain stimulation of action regions can facilitate or inhibit linguistic processes. Pulvermüller, Hauk, Nikulin, and Ilmoniemi (2005) found that priming the respective effector representation area of primary motor cortex (M1) using single-pulse transcranial magnetic stimulation (TMS) decreased reaction times for responding to words describing actions performed by the stimulated effector. Vukovic, Feurra, Shpektor, Myachykov, and Shtyrov (2017) found that online repetitive TMS to the motor cortex slowed reaction times to action words, while leaving reactions to abstract words unaffected. These experiments bolster the view that sensorimotor systems in the brain are not only activated by language comprehension but play active roles in understanding.

1.2. How do novel words become grounded in sensorimotor systems?

Despite the abundant evidence that language is grounded in the experiences of the body, a key area of embodied language theories lacking empirical support is the process by which sensorimotor experiences come to underlie the representation of novel words. The studies described above observe neural activation during perception of well-known words, which reflects the long-term semantic networks of these concepts and top-down knowledge of affordances. Research on the acquisition of embodied language effects is sparse (see Öttl, Dudschig, & Kaup, 2017; Richter, Zwaan, & Hoever, 2009), and it has yet to be demonstrated that natural interactions with novel objects can give rise to the kinds of effects discussed above. Empirical support for this initial phase of embodied language learning provides a crucial test of embodied language theories. If novel words do not show early effects of the sensorimotor context in which they were learned, this would undermine the view that the sensorimotor processes are truly constitutive of the word meanings, rather than more passive associations. However, if specific sensorimotor experiences that take place during word learning influence how those words subsequently affect behavior, it would provide powerful support for grounded word learning.

Circumstantial evidence for sensorimotor interaction playing an important role in the acquisition of new words can be found in studies of infant word learning. Yu, Smith, and Pereira (2008) found in a novel word learning study with 18-month-old children that the proportion of time that an object remained in an infant's visual field, as well as the amount of time holding a named object when its name was spoken, was predictive of successful word learning. This suggests that sensorimotor properties are important features of word learning, but it leaves unclear whether the kinds of interactions experienced by a learner influence the semantic representations of learned words.

One reason that evidence for the acquisition of embodied language effects is sparse is due to a tendency for word learning studies to rely on standard computer tasks where participants learn the novel words for flat images of objects on a screen (Kirkham, Slemmer, & Johnson, 2002; Smith & Yu, 2008; Trueswell, Medina, Hafri, & Gleitman, 2013;Yu & Smith, 2011). This kind of learning is very unlike real-world word learning, where children learn about objects by picking them up and interacting with them. Research with real objects finds that real world objects are remembered more accurately than their photographic counterparts (Snow, Skiba, Coleman, & Berryhill, 2014). Neuroimaging studies also show that the neural mechanisms involved in processing 3D objects may be distinct from mechanisms involved in processing 2D versions of those same objects (Snow et al., 2011). In addition, when action animations that align with the meaning of learned verbs are presented with the verbs, learning is greater than when those animations do not align with the verbs (Hald, van den Hurk, & Bekkering, 2015). This suggests that having concurrent representations of the actions implied by words improves learning of the words. This corresponds to how words are learned in the world, where a spoken word often co-occurs with the object or action it refers to, or a gesture indicating the action. This body of work suggests that realistic objects and movements will be more likely to result in embodied language effects.

We conducted a series of three experiments to investigate sensorimotor grounding of novel words acquired through sensorimotor interaction with objects in a virtual environment. In the first experiment, we investigated whether participants would be faster to respond to novel words that were learned through sensorimotor interaction when the action required for the response used the same hand and movement as the affordance learned for the word. We refer to this relationship as an affordance compatibility effect. One previous study has shown evidence of spatial congruency effects for novel words (Öttl et al., 2017). The authors had participants learn the names of novel objects in the environment in front of them that were either located in the upper or lower visual field. In a test phase, recollection of the objects was facilitated when participants made an up or down movement congruent with the original location of the object. This study reveals a spatial component of the learned representations; however, it does not involve the kind of realistic sensorimotor interaction with objects thought to underpin natural word learning. In contrast, participants in our experiments learned novel object names in a virtual environment with naturalistic affordances and were then tested in a word-color match test. We found that participants acquired an affordance compatibility effect where they were faster to respond to matches in the test phase with the hand used to interact with the named object from the training phase. In a follow-up experiment, participants learned the same words by observing virtual hands interacting with the objects. This was done to investigate whether the affordance compatibility effect is dependent upon direct object manipulation. Finally, in a third experiment, we explored the extent to which spatial affordance compatibility could be separated from effector-specific affordance compatibility.

2. Experiment 1: Direct manipulation induces affordance compatibility

In a first experiment, we explored whether sensorimotor experience during novel word learning would influence later processing of the learned words. We developed a gamified virtual reality experiment using Unreal Engine 4. Using virtual reality enabled us to attach specific manual affordances to virtual objects with rich visual properties. We predicted that learners would associate the affordances of objects with the words for those objects and that these associations would influence behavioral responses even when explicit retrieval of the affordances was not necessary. We tested this by comparing responses in a word-color matching task. If responses which were compatible with the affordance of a word were faster than incompatible responses, this would be evidence of a learned affordance compatibility effect. In addition, we incorporated a variety of visual and auditory consequences of actions (e.g., potions pouring a stream of liquid) to motivate learning and encourage participants to engage with the virtual environment.

2.1. Methods

Twenty-seven participants (23 women; 25 right handed) completed a two-part experiment using an HTC Vive virtual reality system (Fig. 1). We initially planned 30

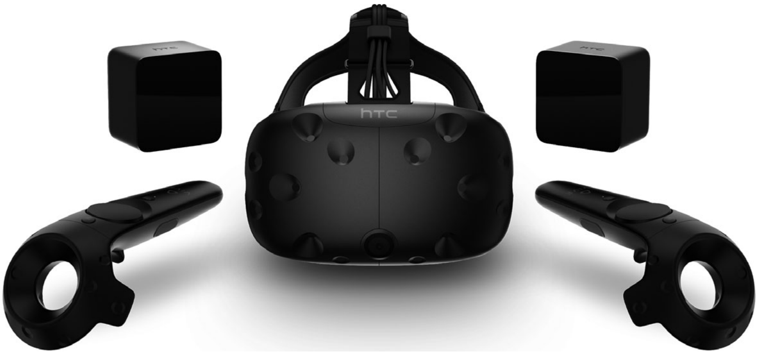

Fig. 1. The HTC Vive virtual reality system consisting of a motion-tracked head-mounted display with 2,160 × 1,200 resolution, two handheld motion controllers, and two wall-mounted infrared sensors. During the experiment, participants cannot see their actual body or surroundings, but instead see a game-like environment and virtual hands (see Fig. 2). Image source: HTC Vive Press Kit.

participants, but due to technical issues and low performance during training, only 27 completed the experiment. Participants were adult undergraduate students (age 19–23 years) recruited from the University of California, Merced behavioral subjects research pool. All participants had normal or corrected vision and normal hearing and spoke fluent English (19 bilingual). Participants provided informed consent prior to beginning the experiment. During the experiment, participants interacted via hand-held controllers with a virtual environment. The controllers were visually represented to the participants as virtual hands which tracked the position and orientation of the participant's actual hands. All participants completed a pre- and post-exposure comfort survey (Data S1) adapted from the Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, & Lilienthal, 1993) and were debriefed on the nature of the experiment upon completion.

The training phase of the experiment was a gamified novel word learning task. Participants learned the names of six novel objects. Names were selected from the NOUN Database (Horst & Hout, 2016) and randomly assigned to the six objects. The objects were modelled to resemble potion bottles with visually distinct shapes and colors. The objects were arranged on either side of a large cauldron which occupied the center of the virtual space. Objects on the right side had handles on the right side and could only be grasped with the right hand, and symmetrically for the left side. Each novel object afforded either a left- or a right-handed grasp.

Participants were instructed to "Pour in these ingredients…" followed by one of the novel words. Following the prompt, participants picked up one of the objects by the handle (pulling the trigger on the controller to grasp) and tilted it over the virtual cauldron to pour the ingredient (Fig. 2). If the word matched the ingredient poured, a swirling particle effect indicated success. If the ingredient did not match, the cauldron exploded, and the potions were reset to the sides of the cauldron in random positions but without changing the side on which a given object appeared. Pouring trials were grouped into recipes of 2–6 non-repeating ingredients. If the participant correctly poured all of the ingredients

Fig. 2. Screenshots from the perspective of a participant in Experiment 1. (Top) Six novel objects arranged on either side of a cauldron. During each training trial, the participant heard one of six object names, then grasped one object by the handle (silver or gold rings) and poured it into the cauldron. (Bottom Left) If the correct object was poured into the cauldron, swirling "magical" particles would indicate success. A series of successful pours would cause an object (e.g., floating globe) to appear. (Bottom Center) An incorrect pour resulted in the cauldron exploding. (Bottom Right) During the test phase, the participant heard one of the novel words and saw a patch of color. The participant was instructed to pull the left (or right) trigger on the motion controller if the patch of color matched the color of the named object.

in a recipe, a short musical tune was played, a virtual object (e.g., a floating globe) appeared somewhere in the environment, and the potions were randomly reset to their sides. The training process was repeated until participants completed 20 recipes. Most participants completed the training in 10–25 minutes. Two participants failed to complete the training phase in one hour and were excluded from further analysis.

After training, participants performed a match-mismatch reaction time task based on the Action-Sentence Compatibility Effect (Glenberg & Kaschak, 2002). Participants heard one of the words from the previous phase and were presented with a patch of color matching one of the objects. The patch of color was shown 100 ms after the start of the audio. Participants were instructed to respond as quickly as possible by pulling one controller trigger if the color matched the named ingredient or pulling the other if it did not match (counterbalanced between subjects). Trigger pulls performed during the match response were the same movements used to grasp the objects in the learning phase. If the word referred to a potion poured with the same hand as the trigger response, the trial was coded as *compatible*, otherwise it was *incompatible*. There were 200 randomized trials, half of which were compatible and half of which were incompatible trials. There were also an equal number of *match* ("yes" response) and *mismatch* ("no" response) trials. If no response was made within 1.5 s, the trial ended and was recorded as a non-response. Trials were completed in 5 blocks of 40 trials separated by 10 second breaks. Word-object mappings were randomized in the training phase after every 10 subjects to control effects of word or color during the test phase. We recorded and analyzed which response was made and reaction times for all trials.

2.2. *Results*

A total of 5,400 test trials were completed with a no-response rate of 7.4% and an incorrect response rate of 3.9%. We excluded mismatch trials from further analyses of affordance compatibility. Mismatch trials require a participant to retrieve both the object that the word references and the object that the color references. Either, none, or both of these objects may correspond to a compatible affordance. Match trials only require recalling the object referred to by both. We also eliminated incorrect and no-response trials. Response times that were 2.5 standard deviations away from the mean for each subject were discarded as outliers (<1% of the data). There was a small decrease in reaction times for most subjects during the first 10–20 trials indicating an effect of practice. However, we determined that practice effects did not interfere with further analysis, so we did not exclude initial trials.

Incorrect response rates were nearly identical for compatible and incompatible trials while the no-response rate was slightly greater for incompatible trials (Fig. 3). Both kinds of errors were relatively infrequent, which may be due to a lack of pressure to respond quickly. Compatible match trials were 29 ms (90% CI = (−5, 69) ms) faster on average for the right hand and 16 ms (90% CI = (−14, 44) ms) faster for the left hand.

We performed a linear mixed effects analysis on reaction times using R and the nlme package (Pinheiro et al., 2017), following recommendations from Zuur, Ieno, Walker, Saveliev, and Smith (2009). Fixed effects included in the model were response (left or right), affordance (left or right), and the interaction between these variables. We added random intercepts for subjects, as maximum likelihood tests performed using REML (restricted maximum likelihood estimation) indicated this was the best fit for the random term of the model. We ran a likelihood-ratio chi-square test comparing the full model to a null model without the affordance by response interaction to determine whether the full model performed significantly better. We obtained p-values for individual model predictors by running the full model with REML. We then used a Monte Carlo simulation method to compute the probability of finding a significant interaction (ß = 0.88) and 90% confidence intervals for the model coefficients. We calculated the Monte Carlo estimates by sampling subjects and trials with replacement from our original data.

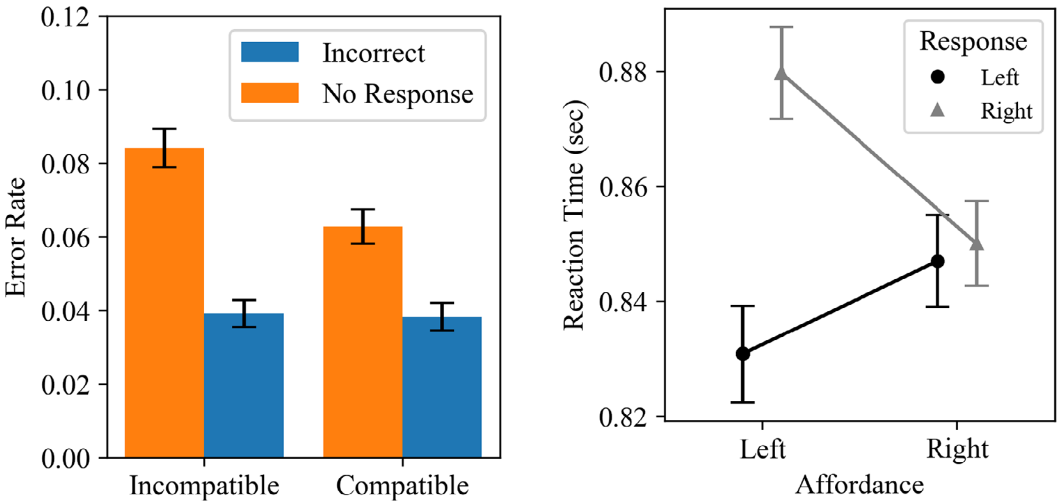

Fig. 3. (Left) The incorrect and no response rates in Experiment 1, for compatible versus incompatible trials. Error bars denote *SEM*. The no-response rate was roughly 2% greater for incompatible than compatible trials, corresponding to an average of 4 time-outs per subject. (Right) The mean reaction times for left and right responses by the affordance of the named object for match trials, excluding incorrect and no-response trials. Error bars denote *SEM*. The affordance compatibility effect is shown as an interaction between affordance and response.

The full model significantly outperformed the null model ($p < 0.0001$, L ratio = 17.0). We did not find a significant main effect of response (coef = 54 ms, 90% CI [−12, 127], $t = 1.36$, $p = 0.18$) or affordance (coef = 19 ms, 90% CI [−10, 44], $t = 1.75$, $p = 0.08$). As predicted, we observed an interaction between response and affordance (Fig. 3), such that participants responding to matches with their left hand had quicker responses to words associated with the left affordance and participants responding to matches with their right hand had quicker responses to words associated with the right affordance (coef = −59 ms, 90% CI [−102, −11], $t = 4.13$, $p < 0.0001$, Cohen's $d = 0.29$). This interaction indicates that the novel words acquired an affordance compatibility effect: that actions which were compatible with the sensorimotor interactions practiced while learning a word were facilitated over incompatible actions.

3. Experiment 2: Action-observation induces affordance compatibility

We next asked whether the affordance compatibility effect for learned words could be induced without direct manipulation of objects. We conducted a second experiment in which participants performed the training task verbally while observing virtual hands manipulating the objects. Observation of actions recruits a network of brain regions significantly overlapping with the areas active during execution of the same action (Buccino et al., 2001; Fadiga, Craighero, & Olivier, 2005; Grezes & Decety, 2001; Hari et al., 1998). Furthermore, observation of motor learning is found to facilitate motor learning in the observer upon later learning of the same task (Mattar & Gribble, 2005), suggesting that the observer was simulating the motor experience of the actor as they watched the action unfold. Thus, we expect the neural processes occurring in an action observation version of our task to substantially overlap with those in Experiment 1, resulting in similar formation of the associations between motor affordances and object labels. This would reduce the need for an individual to have exhaustive experience with a referent object or action in order to acquire fully grounded representations, since many grounded features can be acquired through social learning or observation of others.

3.1. Methods

Experiment 2 was conducted using the method from Experiment 1 with several modifications. In the training phase, following the prompt, instead of grasping and pouring one of the objects with the controller, participants verbally indicated which object they wished to pour. Each position from left to right was marked by a floating number (1 through 6). Participants indicated their choice by reading the number above the object. Participants were not given controllers. When the participant made a selection, the experimenter entered the choice on a keyboard and a virtual hand followed a pre-recorded trajectory to reach out, pick up, and pour the potion. Because the no-response rate was relatively high in Experiment 1, we increased the trial duration in the test phase from 1.5 s to 2 s to avoid truncating the reaction time distribution. We decreased the number of test trials from 200 to 160 due to concerns of fatigue, although our post-exposure comfort survey ultimately determined this was not an issue. Twenty-seven participants (18 women; 26 right handed; 18 bilingual; age 18–21 years) took part in this experiment.

3.2. Results

A total of 4,200 test trials were completed with a no-response rate of 5.5% and an incorrect response rate of 5.2%. The mean reaction time for correct match trials was 981 ± 292 ms. As in Experiment 1, outliers were discarded. Likely due to the increased time to respond, no-response rates were lower in Experiment 2, and overall reaction times were greater and more variable. No differences were observed for no-response or incorrect response trials as a function of compatibility.

Statistical modeling was conducted as in Experiment 1. The full model outperformed the model without the interaction ($p = 0.036$, L-ratio = 4.39). We did not find a significant main effect for either response (coef = 31 ms, 90% CI [−78, 141], $t = -1.11$, $p = 0.62$) or affordance (coef = −17 ms, 90% CI [−56, 19], $t = 0.51$, $p = 0.27$). There was again a significant interaction between response and affordance (coef = 50 ms, 90% CI [−111, 21], $t = -2.09$, $p = 0.036$, Cohen's $d = 0.17$; Fig. 4), consistent with our prediction that novel words can become associated with their affordances through action observation without action execution. After completing the study we calculated a power estimate using Monte Carlo simulation as in Experiment 1 (ß = 0.538), suggesting that

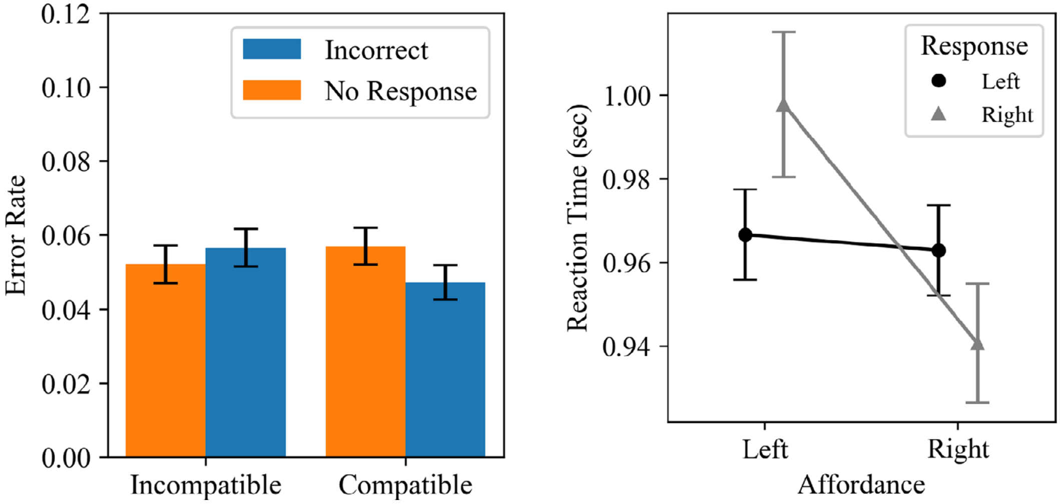

Fig. 4. (Left) The error rates for all participants in Experiment 2. There were no clear differences in error rates between conditions. With the increased trial duration, the overall rate of no-response trials decreased roughly 2% compared with Experiment 1, while the rate of incorrect responses increased 1%. (Right) The mean reaction times for left and right responses by the affordance of the named object in match trials. The interaction between affordance and response demonstrates an affordance compatibility effect, though largely driven by participants responding with their right hands.

this study design was slightly underpowered and a larger replication would help to determine the reliability of this finding.

The affordance compatibility effect in this experiment was driven only by right hand responses. Given that most of the participants were right-handed (96%), it is possible they were more likely to mentally simulate the reach and grasp movement performed with the right artificial hand. Furthermore, the effect size in this experiment was smaller than that of experiment 1. This may indicate that action observation alone gives rise to a weaker association between words and affordances.

4. Experiment 3: Space and hand interact in affordance compatibility

In Experiments 1 and 2, the affordance of each novel object was represented redundantly through the position of the object and the orientation of the handle, as well as through corrective instructions to participants if they attempted to use the incorrect hand. This is consistent with many natural interactions with objects, in which both spatial and visual features indicate affordances. However, because the relative location of an object and the hand used to interact with it was always consistent, it was not possible to distinguish between the contribution of the specific hand and the side of space in which the interaction occurred. The affordance compatibility effects we observed in Experiments 1 and 2 could have been caused by either factor. Therefore, in a third experiment, we separated the spatial and hand compatibility dimensions by swapping the positions of some objects during training. We suspected that the affordance compatibility effect in Experiments 1 and 2 was primarily driven by handedness and we would observe a significant interaction between affordance hand and response hand, but we did not have any a priori hypotheses regarding spatial compatibility.

4.1. Methods

The methods were similar to those of Experiments 1 and 2, aside from the location of each of the novel objects during training. One left-handed object always appeared on the left side and could only be picked up with the left hand. Another always appeared on the right side and could only be picked up by the left hand, requiring participants to reach across their body. A third object was always picked up by the left hand, but randomly alternated between the left and right sides. Right-hand objects had a corresponding flipped arrangement. During the test phase, the hand used to respond was either compatible or incompatible with the hand used to grasp the named object (hand compatibility) and was either on the same side of space, the opposite side of space, or mixed (space compatibility). To ensure the participants knew which hand to use for each object, the object handles were adjusted to face prominently in the direction of the correct hand, and the participants were instructed to pick the object up with the hand matching the handle direction. To collect sufficient data for the factorial design (2 hand × 3 space), 43 participants (34 women; 39 right handed; 26 bilingual; age 18–41 years) completed the training phase and 200 test trials each.

4.2. Results

Participants in Experiment 3 completed 8,160 test trials with an overall no-response rate of 3.8% and incorrect response rate of 7.3% (no significant differences by trial types). The mean reaction time for correct match trials was 927 ± 291 ms. Test trials in which the response hand was compatible with the hand used to grasp the named object were 22 ms faster than incompatible trials. Spatially incompatible trials were 9 ms faster than mixed trials and 13 ms faster than spatially compatible trials (Fig. 5).

As in Experiments 1 and 2, we applied an iterative model testing procedure to choose the best linear mixed effects model determined by model fit and complexity but including the additional interactions with spatial affordance. The model fit procedure and related data processing can be found in the code included in Data S1. The random structure of the model was the same as that of Experiments 1 and 2. Fixed effects included in the optimal model were response hand, spatial affordance (left, right, both), hand affordance (left, right), the interaction between space and hand, and the interaction between hand and response. Our model testing indicated that including the interaction between space and response did not significantly improve model fit ($p = 0.13$, L-ratio = 4.06); therefore, this interaction was omitted from the optimal model. The final model significantly outperformed a model without the hand by the response interaction ($p < 0.0001$, L-ratio = 11.3,

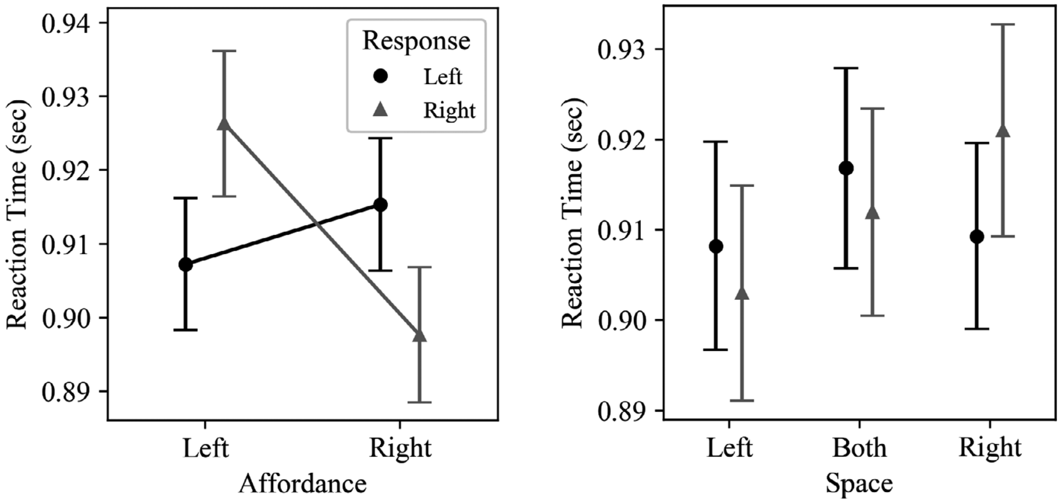

Fig. 5. (Left) Mean reaction times (correct match trials) for left and right responses by the hand used to grasp the named object. The hand compatibility effect is similar to the overall affordance compatibility effect in Experiments 1 and 2. (Right) We did not find any significant interaction between response hand and spatial affordance. Spatial compatibility did significantly interact with hand compatibility.

ß = 0.78) and without the hand by space interaction ($p < 0.0001$, L-ratio = 14.9; Table 1). There was no significant main effect for response or hand affordance. We did observe a main effect of left spatial affordance. As in Experiments 1 and 2, we observed an affordance compatibility effect for the affordance hand (Cohen's $d = -0.19$). We also observed an interaction between hand affordance and space affordance, where having the hand and space share the sidedness feature (both right or both left) speeds responding to the corresponding word (Cohen's $d = 0.24$). These results are consistent with our prediction that hand compatibility is likely the more significant factor in the effects observed in Experiments 1 and 2. Although we did not make predictions regarding the effects of space on reaction time, our results indicate that space might modulate the hand-specific effects. We discuss potential explanations for this below.

5. Discussion

We conducted three novel word learning experiments using virtual reality to investigate the ways in which object affordances become associated with words. Participants learned the names of virtual potions through interaction in several training conditions. Then, in a word–color matching task, we observed faster responses for words which referred to objects grasped with the same hand used for the response than those grasped with the opposite hand. We refer to this as an affordance compatibility effect. This is an important demonstration of naturalistic sensorimotor interaction during word learning giving rise to embodied language effects.

Table 1
Linear mixed-effects model fit by REML

	Value (s)	*SE*	DF	*t*-value	*p*-value	90% CI
Intercept	1.050	0.038	3457	28.0	0.0000	[0.997, 1.139]
Right response	−0.023	0.051	41	0.45	0.65	[−0.088, 0.025]
Right hand	−0.021	0.017	3457	−1.25	0.21	[−0.079, 0.099]
Left space	−0.040	0.014	3457	−2.80	0.005	[−0.101, 0.019]
Right space	−0.012	0.014	3457	−0.81	0.42	[−0.057, 0.032]
Right response × Right hand	−0.056	0.017	3457	−3.36	0.0008	[−0.109, −0.003]
Right hand × Left space	0.075	0.020	3457	3.71	0.0002	[0.015, 0.185]
Right hand × Right space	0.019	0.020	3457	0.94	0.35	[−0.040, 0.095]

In our second experiment, we confirmed that an affordance compatibility effect can be induced, perhaps to a lesser extent, through observation of virtual effectors. This result is consistent with theories suggesting that action observation networks in the brain support imitation and social learning (Iacoboni et al., 1999), and suggests that words and concepts learned while watching others may activate sensorimotor experience-dependent networks in the same way as words learned by doing. Prior work has shown that motor learning can occur during passive observation of a motor task (Mattar & Gribble, 2005), but these findings further demonstrate that action observation influences language learning. It is important to note again, however, that this observation network does seem to activate this embodied representation to a weaker degree than direct manipulation. While the affordance compatibility effect was observed for the right hand responses, it was absent for left hand responses. Research shows that right-handers may have a more difficult time learning words that represent left-handed actions (De Nooijer et al., 2013), which may play a role in our study. Additional work will be needed to explore the role and limits of observation as a means of acquiring embodied semantic knowledge.

In our third experiment, we sought to understand the relationship between effects of hand versus space on the affordance compatibility effect. We confirmed that the effector used to interact with an object acquired an affordance compatibility effect even when crossing the body to interact. We did not identify a direct relationship between the space in which affordances were learned and the resulting response dynamics. The interaction between space and hand suggests that a direct study of the relationship between spatial and motor affordances would be beneficial. Response times were faster when spatial and effector-specific affordances were consistent, regardless of the hand used to make the response. This suggests that a shared affordance feature in this context might facilitate response preparation. Given that spatial location frequently corresponds with effector-specific affordances, interacting with and learning about objects for which these features are inconsistent may engage distinct cognitive and neural mechanisms. The present study was not able to test these predictions directly, preventing a clear picture of spatial affordance effects from emerging. More work is needed to address these questions. This experimental paradigm introduces a platform which can be extended for further exploration of cognition grounded in naturalistic body movements.

Another explanation for these results that we considered is that the affordance compatibility effect is related to a Simon effect (Roest, Pecher, Naeije, & Zeelenberg, 2016; Simon, 1969). The Simon effect refers to a pattern of faster responses when the response and the stimuli share an overlapping spatial dimension. This widely replicated finding can be seen, for instance, by showing participants a picture of a mug, where the participant needs to push a button with the left hand to classify the object. In this example, the participant should be faster to respond if the handle of the mug faces left than if it faces right. In our experiments, an overlap between the response and the handle direction of the recalled object resembles a Simon task. This effect has also been shown when recalling stimuli that were previously overlapping with the response direction (Pellicano, Vu, Proctor, Nicoletti, & Umiltà, 2008; Vankov, 2011; Wühr & Ansorge, 2007). It is possible, therefore, that a Simon effect was a part of the mechanism speeding compatible hand responses, as recalling the direction of the handle would be enough to create this effect. We cannot rule out this interpretation, but future work can address this concern by including a response task where only one hand is used to respond, but the spatial location of response buttons differs. We would expect no difference between the response button locations if our affordance compatibility account were correct.

One important caveat of this work, and much of the existing research on embodied language, is that it is difficult to determine precisely how interactions between language and sensorimotor systems give rise to action compatibility effects. These effects are typically small perturbations of response latency or accuracy which could be caused because cognitive conflict is induced by incompatible action representations or because motor preparation is facilitated by linguistic activation of motor regions. In either case, response incompatibility is obviously regularly overcome during natural behavior, so it is difficult to know the importance of affordance compatibility for everyday cognition. These challenges are not unique to embodied language research: Visuospatial compatibility effects (e.g., S-R Compatibility, Michaels, 1988) similarly rely on slightly speeded responses to investigate motor representations. Nevertheless, a more direct test of the efficacy of sensorimotor activity in linguistic representations is needed if embodied language accounts are to replace, rather than complement, amodal symbolic representations.

A crucial piece in the understanding of embodied language will come from bridging short-term embodied learning effects like those demonstrated here with longer-term embodied language effects observed in fluent adults (Barsalou, 1999; Glenberg & Kaschak, 2002). Embodied language theories predict a progression from specific sensorimotor associations to more flexible, generalized sensorimotor simulations as words are expressed in a broader set of contexts (Barsalou, 2009; Zwaan, 2004). Many unanswered questions remain about how associations between affordances and words change over weeks and months of sensorimotor experience. The effort to answer these questions will benefit from the integration of naturalistic infant and child language learning research with virtual reality experiments offering greater control over sensorimotor interactions. As these questions are tackled, we may come to better understand how our bodies and environments give meaning to the words we use.

Acknowledgments

The work was supported in part by an NSF NRT (DGE – 1633722) grant awarded to the corresponding author. We thank Wim Pouw and two anonymous reviewers for very useful comments on this paper. The paper also benefited from discussions with several members of the Sensorimotor Neuroscience Lab at UC Merced.

Declarations of interest

None.

References

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(4), 637–660.

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

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1281–1289.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., & Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, *13*(2), 400–404.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, *24*(3), 355–363.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*(2), 579–593.

De Nooijer, J. A., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). When left is not right: Handedness effects on learning object-manipulation words using pictures with left-or right-handed first-person perspectives. *Psychological Science*, *24*(12), 2515–2521.

Fadiga, L., Craighero, L., & Olivier, E. (2005). Human motor cortex excitability during the perception of others' action. *Current Opinion in Neurobiology*, *15*(2), 213–218.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3–4), 455–479.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*(1), 1–19.

Glenberg, A. M., Havas, D., Becker, R., & Rinck, M. (2005). Grounding language in bodily states: The case for emotion. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge, UK: Cambridge University Press.

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

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage*, *32*(2), 906–912.

Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, *12*(1), 1–19.

Hald, L. A., van den Hurk, M., & Bekkering, H. (2015). Learning verbs more effectively through meaning congruent action animations. *Learning and Instruction*, *39*, 107–122.

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Sciences*, *95*(25), 15061–15065.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307.

Hauk, O., & Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping*, *21*(3), 191–201.

Havas, D. A., Glenberg, A. M., & Rinck, M. (2007). Emotion simulation during language comprehension. *Psychonomic Bulletin & Review*, *14*(3), 436–441.

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.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*(5449), 2526–2528.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220.

Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience*, *28*(47), 12224–12230.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42.

Matlock, T., Holmes, K. J., Srinivasan, M., & Ramscar, M. (2011). Even abstract motion influences the understanding of time. *Metaphor and Symbol*, *26*(4), 260–271.

Mattar, A. A., & Gribble, P. L. (2005). Motor learning by observing. *Neuron*, *46*(1), 153–160.

Michaels, C. F. (1988). SR compatibility between response position and destination of apparent motion: Evidence of the detection of affordances. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 231.

Öttl, B., Dudschig, C., & Kaup, B. (2017). Forming associations between language and sensorimotor traces during novel word learning. *Language and Cognition*, *9*(1), 156–171.

Pellicano, A., Vu, K. P. L., Proctor, R. W., Nicoletti, R., & Umiltà, C. (2008). Effects of stimulus–stimulus short-term memory associations in a Simon-like task. *European Journal of Cognitive Psychology*, *20*(5), 893–912.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., & Maintainer, R. (2017). Package ‘nlme’. *Linear and Nonlinear Mixed Effects Models*, 3-1.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793–797.

Richter, T., Zwaan, R. A., & Hoever, I. (2009). Acquiring experiential traces in word-referent learning. *Memory & Cognition*, *37*, 1187–1196. https://doi.org/10.3758/MC.37.8.1187.

Roest, S. A., Pecher, D., Naeije, L., & Zeelenberg, R. (2016). Alignment effects in beer mugs: Automatic action activation or response competition? *Attention, Perception, & Psychophysics*, *78*(6), 1665–1680.

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, *81*(1), 174.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Snow, J. C., Pettypiece, C. E., McAdam, T. D., McLean, A. D., Stroman, P. W., Goodale, M. A., & Culham, J. C. (2011). Bringing the real world into the fMRI scanner: Repetition effects for pictures versus real objects. *Scientific Reports*, *1*, 130.

Snow, J. C., Skiba, R. M., Coleman, T. L., & Berryhill, M. E. (2014). Real-world objects are more memorable than photographs of objects. *Frontiers in Human Neuroscience*, *8*, 837.

Spivey, M. J., Tyler, M. J., Richardson, D. C., & Young, E. E. (2000). Eye movements during comprehension of spoken scene descriptions. (pp. 487–492).

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273–281.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156.

Vankov, I. (2011). The interplay of spatial attention and memory: Evidence for memory-driven spatial compatibility effects, but not for a facilitating role of attention in memory retrieval. *Attention, Perception, & Psychophysics*, *73*(7), 2124.

Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., & Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: An online rTMS study. *Neuropsychologia*, *96*, 222–229.

Wühr, P., & Ansorge, U. (2007). A Simon effect in memory retrieval: Evidence for the response-discrimination account. *Psychonomic bulletin & review*, *14*(5), 984–988.

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165–180.

Yu, C., Smith, L. B., & Pereira, A. F. (2008). *Grounding word learning in multimodal sensorimotor interaction*. Science Society.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects modelling for nested data. In Mixed Effects Models and Extensions in Ecology with R* (pp. 101–142). New York, NY: Springer.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of Learning and Motivation*, *44*, 35–62.

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In H. A. E. M. Pecher, & R. A. Zwaan (Eds.), *The grounding of cognition: The rol of perception and action in memory, language and thinking* (pp. 224–245). Cambridge, UK: Cambridge University Press.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Data S1. Supplemental Data.