Raising the roof: Situating verbs in symbolic and embodied language processing

John Hollander[1]* and Andrew Olney[1]

[1]University of Memphis Department of Psychology and Institute for Intelligent Systems

*Corresponding author: jmhllndr@memphis.edu

Keywords: Embodied cognition; Verbs; Semantic judgment; Language processing; Distributional semantic models

1. Abstract

Recent investigations on how people derive meaning from language have focused on task dependent shifts between two cognitive systems. The symbolic (amodal) system represents meaning as the statistical relationships between words. The embodied (modal) system represents meaning through neurocognitive simulation of perceptual or sensorimotor systems associated with a word's referent. A primary finding of literature in this field is that the embodied system is only dominant when a task necessitates it, but in certain paradigms this has only been demonstrated using nouns and adjectives. The purpose of this paper is to study whether similar effects hold with verbs. Experiment 1 evaluated a novel task in which participants rated a selection of verbs on their implied vertical movement. Ratings correlated well with distributional semantic models, establishing convergent validity, though some variance was unexplained by language statistics alone. Experiment 2 replicated previous noun-based location-cue congruency experimental paradigms with verbs and showed that the ratings obtained in Experiment 1 predicted reaction times more strongly than language statistics. Experiment 3 modified the location-cue paradigm by adding movement to create an animated, temporally decoupled, movement-verb judgment task designed to examine the relative influence of symbolic and embodied processing for verbs. Results were generally consistent with linguistic shortcut hypotheses of symbolic-embodied integrated language processing; location-cue congruence elicited processing facilitation in some conditions, and perceptual information accounted for reaction times and accuracy better than language statistics alone. These studies demonstrate novel ways in which embodied and linguistic information can be examined while using verbs as stimuli.

2. Introduction

The processes by which people extract meaning from language are essential subjects of psycholinguistics and cognitive science. Advancements in the cognitive study of semantic processing have unfolded along two trajectories: an account based on the principles of embodied cognition, and an account based on distributional language statistics. The theoretical frameworks that have developed around these accounts often do not parsimoniously accommodate one another. In fact, some consider them to be in theoretical opposition (see De Vega et al., 2008; Shapiro, 2014 for multiple perspectives). However, both distributional and embodied accounts are generally considered to be essential parts of cognitive processing (Barsalou et al., 2008; Louwerse & Jeuniaux, 2008; Vigliocco et al., 2009) such that empirical support for one is not necessarily evidence against the other. Recent developments in cognitive science suggest that these two seemingly incongruous accounts may be parallel, complimentary, or intertwined, and have sought to reconcile the theoretical and methodological gaps between them (Andrews et al., 2014; Barsalou, 2010, 2016; Bruni et al., 2014; Louwerse, 2011, 2018; Willems & Francken, 2012; Zwaan, 2014). Integrated theories recognize that contextual demands during language comprehension may necessitate the usage of embodied or symbolic systems in concert: both types of systems may be necessary to fully account for human language processing, but the degrees to which either is dominant may be contextually mediated. The challenge that remains is determining when, how, and why linguistic and embodied information interact during language processing.

However, integrated theories must synthesize from two bodies of literature and evidence that tend to conflict theoretically and methodologically, which constrains behavioral investigations in some ways, sometimes resulting in homogenized stimulus sets and experimental designs. In particular, previous experimentation in integrated language processing theory has neglected to include verbs, both in isolation and as components of predicates in propositional accounts. While experiments in embodied and symbolic literature have accounted for verbs independently, many of the designs involved are often methodologically incompatible or insufficient to test integrated theories. As a result, researchers seeking to experimentally reconcile these two theories have done so primarily with nouns, resulting in an incomplete account of language processing. Accordingly, it is important to examine the evidence in support of embodied and symbolic theories with respect to verbs so that the current state of integrative theories can be more comprehensively evaluated.

2.1. Embodied theory

The fundamental assertion of an embodied theory of language processing is that meaning is constituted by simulations of the perceptual states that have been present in previous experience with the referents of words. This strength of this premise varies between threads of embodied language processing research (Meteyard et al., 2012), which incorporate evidence developed along a few trajectories of neuropsychological and cognitive experimental paradigms.

Neurocognitive researchers have observed motor cortex activation and embodied processing facilitation effects with action words and sentences. For example, word recognition studies have revealed that when people read the word *kick*, some of the areas of the brain responsible for movement of the feet and legs activate higher than baseline despite no leg movement taking place (Pulvermüller et al., 2005; Tettamanti et al., 2005). In another study, participants were instructed to hold their hands in certain configurations (either open palm or in a fist) while performing sentence sensibility judgments. When the sentences contained phrasing which matched participants' hand configurations (e.g., containing the word *applauded* matched with an open palm), participants were faster to complete their judgments than when the word was incongruous to their hand shape (Aravena et al., 2010). Further, increased neural activity (via ERP) was observed at nodes near the motor cortex during incongruous trials, suggesting that the hand shape manipulation could induce facilitative preactivation or interference of the accurate neuromotor simulation. Similar studies involving neuroimaging and physical posturing have found effects using abstract language (Guan et al., 2013) and idioms (Boulenger et al., 2009). This type of evidence demonstrates that language processing does not happen just in "language centers" of the brain, but in many regions, including the sensorimotor cortex.

Cognitive researchers have also devised ways to observe embodied effects without relying on the manipulation of gross motor movements. (Matlock, 2004) found that participants were slower to make judgments about sentences in which fictive motion (e.g., *the road runs through the valley*) was depicted in difficult terrain as compared to more easily traversable terrain. This effect was further replicated in an eye-tracking study, suggesting that, while participants were not instructed to make gross motor movements, and the sentences did not refer to any real motion, the language in the studies influenced processing to an impeded pace analogous to navigation of the implied topography (Richardson & Matlock, 2007). A review by Zwaan & Madden (2005) details other evidence of implicit perceptual simulation during language processing. One study tasked participants to read individual sentences and then name related pictures that exhibited either congruent or incongruent features as implied, but not explicitly stated, by each sentence. For example, after reading *The ranger saw the eagle in the sky*, the subsequent picture of an eagle would either have its wings outstretched (sentence congruent) or folded (incongruent). Participants were faster to name sentence congruent pictures (Zwaan et al., 2002). According to the tenets of embodied theory, the best explanation of this effect is that sentence comprehension involves interactive representations that contain perceptual information, such as shape and orientation, which is otherwise not explicitly conveyed.

The research described so far typically involves verbs, action phrases, or physically interactive experiments. However, nouns and adjectives have played a role in embodied experiments as well, and have come to dominate certain influential experimental paradigms, such as spatial iconicity or location congruency experiments. In an illustrative example of these experiments, two nouns are simultaneously presented on a computer screen in a vertical arrangement, meaning that one word appears in the upper part of the screen, and one appears lower. Sometimes, these words referred to object pairs that are canonically found in a fixed vertical order (e.g., an *attic* is higher than a *basement*). Zwaan & Yaxley (2003) originally devised this paradigm, and tasked participants to determine whether word pairs were semantically related. When pairs that possessed a canonical order were positioned in that order (e.g., *attic* above *basement*), participants were faster to respond than when the order was reversed (e.g., *basement* above *attic*). This effect was found to extend to more abstract relationships, including words indicating power differentials between people (e.g., *master* above *apprentice*) (Schubert, 2005). Additional experiments by both sets of researchers suggest that this effect was due only to the vertical positioning manipulation, as opposed to linguistic factors like typical word order. These researchers concluded that the facilitative effect of word-referent organization congruency could be explained by the application of a perceptual simulation and not by language statistics.

A similar line of experiments involves the presentation of single words at specific locations on the screen. Šetić & Domijan (2007) designed an experiment using a set of words referring to either animals or inanimate objects and were associated with either high or low positions in a typical visual setting (e.g., a *bird* is often in the upper visual field, while a *carpet* is often low). These words were presented individually in either high or low positions on a screen, and participants were asked to judge whether the word referred to a living or non-living thing. As an embodied account would predict, word-referent location congruency facilitated judgment speed. This facilitation effect was replicated in another series of experiments, including one in which participants only classified the font colors of words, implying that deep processing is not necessary to observe this effect (Lachmair et al., 2011). A vertical congruency facilitation effect was found with the classification of ocean-dwelling versus sky-dwelling animals (Pecher et al., 2010) and the positive-negative classification of emotional valence words (Meier & Robinson, 2004; Zhang et al., 2015). Estes & Barsalou (2018) contains a meta-review of the methods used by these and similar studies of the influence of spatial information on word processing. These types of experiments would go on to provide a crucial foothold for the integration of symbolic language processing theory in embodiment-oriented experiments.

2.2. Symbolic theory

The symbolic approach to language processing contends that meaning is derived from the statistical relationships between words and the contexts in which they occur. These theories are generally guided by the principle that perceptual input from listening and reading activate rule governed symbolic representations, and the interaction of these symbols animates language processing. Modern distributional semantic models often include vector space representations, in which words are extracted and represented as numerical vectors in high dimensional space. The use of these types of models in cognitive research implies a degree of commitment to the theory that words are purely arbitrary symbols that are given meaning by their relational properties (Griffiths et al., 2007; Landauer et al., 2007). These models, which do not incorporate perceptual experience within representation, correlate strongly with human performance on a variety language-based tasks (Jones et al., 2015; Lenci, 2008). On these grounds, it can be argued that embodied systems are rendered insufficient and unnecessary in an explanation of human semantic processing.

Symbolic models of language processing based on language statistics begin with a large corpus of linguistic input. Once a corpus is selected, researchers program algorithms which count or predict words in context. One of the earliest models, the Hyperspace Analogue to Language (HAL), uses these data to comparatively represent words as vectors in a high dimensional space (Lund & Burgess, 1996). By comparing the distance between vectors in this semantic space, the similarity of words can be quantified. Other models, such as Latent Semantic Analysis (LSA; (Landauer & Dumais, 1997), Bound Encoding of the Aggregate Language Environment (BEAGLE; Jones & Mewhort, 2007), and Topic Model (Griffiths et al., 2007) perform similar computational processes, but with different approaches, assumptions, and parameters. Despite the algorithmic differences between these models, their success in language-based tasks and association with human performance contributes to the implication that they represent cognitively plausible models of human language processing.

One immediate benefit of the use of distributional types of symbolic semantic models is that they can readily incorporate a variety of linguistic input, regardless of grammatical class. In fact, HAL has been able to categorize parts of speech since its inception (Burgess et al., 1998). LSA has been a particularly progenitive model in this regard. Its usage of higher order co-occurrences has been demonstrated to correlate well with human performance on many psycholinguistic tasks, such as the Test of English as a Foreign Language (Landauer & Dumais, 1997), text cohesion judgments (Foltz et al., 1998; Mcnamara et al., 2007), and document classifications (Louwerse et al., 2004). LSA also boasts a strong correlation with human performance (e.g., accuracy and reaction times) in laboratory studies, particularly in semantic categorization tasks, though these types of tasks predominantly feature nouns (Siakaluk et al., 2003). As a result, LSA and similar models have been used as tools in increasingly effective and complex educational and learning technologies in many domains (Graesser et al., 2011), including adult literacy (Graesser et al., 2016). More recently, endeavors to map semantic vectors onto neurologically analogous featural concept representations have informed the degree to which vector space models may reflect human cognitive processes regarding both semantic and grammatical classes, including nouns, adjectives, and verbs (Utsumi, 2020). As researchers continue to innovate distributional semantic models (for example, Bidirectional Encoder Representations from Transformers, or BERT; Devlin et al., 2019), our ability to approximate human language processing with computational methods has increased exponentially.

While symbolic language processing models built without sensorimotor or perceptual input continue to improve, their continued use in cognitive science implies an interesting supposition: if vector space models covary so strongly with human performance, then human language processing could work in a similar way. However, Günther et al. (2019) argue that a claim of total amodality in conceptual processing is a misrepresentation of this research which is borne from the methodological limitations of distributional semantic models. Due to the lack of multimodal input channels, most distributional semantic models often cannot be used to make claims about multimodal processing in linguistic tasks. Rather, they can only account for linguistic experience, which is typically sufficient to complete the tasks involved. This implies that embodied processing is not necessary to complete some tasks, which is not the same as claiming that there is no embodiment in language processing. As a result, the issue of task specificity makes it difficult to reconcile symbolic and embodied accounts.

However, distributional semantic models do not necessarily preclude the use of non-linguistic information on either a methodological or theoretical basis. In fact, there have been successful attempts in creating distributional semantic models by merging linguistic and non-linguistic information, especially by using computer vision techniques (Bruni et al., 2014; Günther et al., 2022). While distributional models often do not account for multimodal information, it is entirely possible for them to do so. The challenge for integrated distributional models, then, is to account for the myriad and interconnected input streams of human perception. An additional challenge is to do so while performing both linguistic and non-linguistic tasks, as task goals may affect the relative dominance of symbolic and perceptual information (Dudschig & Kaup, 2017).

Unfortunately, many of the tasks employed by embodied and symbolic orbits of research have such little overlap that they provide little information about the interplay between symbolic and embodied information during language processing. For instance, manipulating bodily positions during lexical judgments has no analogue in a semantic co-occurrence model. These types of methodological gaps between the two theories are exacerbated by certain types of language, such as verbs, which often require bodily interaction in embodied experiments. Thus, independent lines of embodied and symbolic research have been unable to address the possibility that embodied and symbolic processing are interdependent in ways that are not always captured due to limitations in experimental methods. This possibility provides a key conceptual foothold for integrated theories of language processing.

2.3. Integrated theories

While substantial evidence has accumulated in support of both symbolic and embodied accounts, researchers have also advanced theories integrating the two. Integrated theories have faced the difficult challenge of reconciling hypotheses which are separated by methodological incompatibilities. These discrepancies are largely responsible for the polarization of method and theory between symbolic and embodied theories of language processing. Studies seeking to test embodied theory tend to employ embodiment-focused methods while neglecting (or suboptimally accounting for) symbolic processing, and vice versa, iteratively increasing separation between the two frameworks. Seeking a resolution to these differences, researchers have posited a few integrated theories. *Representational pluralism theory* suggests that some words are codified modally, while others are amodal (Dove, 2009). This would attribute the responsibility of embodied or symbolic effects on the words themselves, rather than the tasks one performs with them. *Language and situated simulation theory* posits that both embodied and symbolic information are used in conceptual processing, which itself is manipulated by language processing (Barsalou et al., 2012). One integrated theory that has generated particularly relevant research is Louwerse's (2007, 2011) *symbol interdependency hypothesis*, which emphasizes that language processing primarily consists of symbol manipulation that serves as a code of embodied, perceptual information.

These types of integrated theories have been generally summarized as *linguistic shortcut hypotheses* (Banks et al., 2021; Connell, 2019), in which language serves as symbols that are cognitively cheaper but rougher approximations of sensorimotor and perceptual representations. As such, symbolic processing affords fast, imprecise representations to a language user. These quick representations are then sharpened, enhanced, or specified by embodied systems when necessary. In line with Paivio's (1971) dual coding theory, the contextual demands of language use do not always necessitate the specificity of embodied representation. In these cases, the symbolic system is sufficient, and the embodied system need not be recruited. A distinguishing factor in this theory, however, is that the embodied system ultimately provides a representational grounding for the symbolic system, which is situationally recruited to facilitate processing. Linguistic shortcut hypotheses accommodate most of the evidence previously attributed to either embodied or symbolic theories because they account for the task dependent nature of language processing. Previous investigations of embodied processing theories typically involved tasks which required or elicited embodied language processing while symbolic processing studies did not. Importantly, this theory generates predictions about what may happen if embodied and symbolic processes were to be experimentally manipulated in a single methodological paradigm. Devising a way to incorporate such a manipulation across incompatible methodologies is one of the primary challenges facing integrated theories.

Some researchers have taken a data-driven, computational approach to this challenge. For example, Petilli et al., (2021) compared the relative association of language-based and computer vision-based referent similarity metrics to human semantic priming performance and found that visual similarity between referents facilitates semantic processing even in purely linguistic tasks. Other researchers have approached this challenge with behavioral experimentation, especially by adapting the aforementioned spatial iconicity and localization paradigms.

These experiments often feature nouns that were presented on a computer screen in positions congruous with their typical referent positions (e.g., *monitor* above *keyboard*), which elicit processing facilitation in semantic judgments. Results of these experiments were originally interpreted as support for an embodied account of language processing, as the facilitative effect of congruency could be explained by the application of a perceptual simulation and not by language statistics. Louwerse and Jeuniaux (2010) brought this explanation under scrutiny by applying two methodological innovations to the experimental design. First, the question being asked of participants was changed; in one condition, people were asked whether the two nouns were semantically related, while in another condition, people were asked whether the nouns were presented in their iconic arrangement (i.e., in their typical vertical orientation). By comparing a semantic judgment to an iconic one, the role of task context could be directly observed and contrasted. Second, embodied and linguistic processing factors were operationalized and quantified in order to observe their role in each task. In this case, the symbolic factor was quantified as word order frequency (e.g., how often *attic* appears immediately before *basement* in a text corpus), and the embodied factor was quantified as ratings of the iconicity of word pair presentations in a separate survey. The symbolic factor was significantly predictive of reaction times for the semantic judgment task, while the embodied factor was not. Additionally, both the embodied factor and the symbolic factor significantly predicted reaction times for the iconicity judgment task. Finally, the symbolic factor's ability to predict iconic judgment speed was weaker than its ability to predict semantic judgment speed. The results of this study support an integrated theory, where symbolic and embodied systems are recruited for different tasks and over different time courses. The semantic judgment task only requires quick, linguistic, shallow information, and thereby only recruits symbolic language processing systems. In contrast, the iconicity judgment requires perceptual representation, so embodied processing is recruited. Recently, further support of symbol interdependency has been found by innovating this paradigm through the introduction of ERP recordings (Louwerse & Hutchinson, 2012), as well as with the implementation of both concrete and abstract nouns.

The presentation of noun pairs provides an intuitive foothold for the application of both symbolic and embodied processes because word co-occurrence is easy to calculate, and nouns are relatively easy to visualize and use as stimuli in computerized experiments. Researchers have also conducted integrated spatial congruency experiments involving the presentation of single nouns. In response to pro-embodiment interpretations of these experiments, researchers have determined that linguistic factors may indeed explain some of these observations due to the relationships between the words presented and the words used to label the response options (Lakens, 2011; Louwerse, 2011). For instance, *bird* may be more quickly categorized as sky or ocean dwelling when seen in the upper part of the screen because it has stronger linguistic co-occurrence statistics with the words *sky* and *up* than it does with *ocean* and *down* because these symbolic relationships create rapid representational activation. In an empirical follow-up, Hutchinson & Louwerse (2013) adapted a similar task from Pecher et al. (2010), instructing participants to judge whether the words represented living or non-living things. Although nouns were presented singularly on the screen, bigram frequency of words used in sequential trials was recorded in order to quantify a linguistic factor for use in predictive models. While the concept-location facilitation effect was observed, bigram frequency also explained a significant portion of reaction time variance, suggesting that both embodied and symbolic systems contribute to the processing required in this task. While this study did not manipulate embodied and symbolic task conditions or explore beyond the domain of nouns, it does provide an example of how results previously considered to be in favor of either embodied or symbolic theory unilaterally can be empirically redefined as support for integrative theories.

Another potential explanation for some of these effects is polarity correspondence (R. Proctor & Cho, 2006), in which the alignment of dichotomous category valences facilitates classification processing. According to polarity correspondence, the facilitative effects in these sorts of tasks may be due to the structural overlap between binarily coded perceptual, conceptual, and response conditions (e.g., good/up/yes vs. bad/down/no). This alternative explanation appears to be substantiated specifically in tasks that may be classified as binary choice judgments, which include location-cue congruency tasks (see Lakens, 2012 for a review), but contrary evidence raises further questions about the mechanism of facilitation in these studies (Pecher et al., 2010). To illustrate, Lakens (2012, p. 8) states that polarity literature would suggest "fastest responses for +polar words presented UP, followed by +polar words presented DOWN, in turn followed by the hypothesized equally fast categorization times for –polar words presented either UP or DOWN." In other words, the word *bird* presented high may elicit processing facilitation not necessarily because of perceptual simulation, but because both it and the upward direction are positively valanced in opposition to negatively valanced objects and the downward direction. However, additional experiments by Dudschig and Kaup, (2017) further explored task dependency with this paradigm and found location congruency facilitation effects for nouns to be robust in non-binary decision tasks, further suggesting cognitive simulation as the likeliest explanation.

These experiments are important for the development of integrated language processing theories because their effects were found by successfully merging paradigmatic methodologies from each theory into designs that allow for their direct comparison. For example, in the previously mentioned study by Louwerse and Jeuniaux (2010), effects of the individual semantic and iconic judgment tasks may be interpreted separately to support either symbolic or embodied theories unilaterally. In tandem, however, inferences can be made about dual systems language processing. Other studies have used similar methods to examine processing interdependency with noun-adjective property congruency judgments (Louwerse & Connell, 2011; Tillman & Louwerse, 2018). However, the set of stimuli and tasks found in this body of experimental literature is relatively small and homogenous. So far, integrated experiments like these tend to involve semantic judgments about series of words, usually nouns and sometimes adjectives, presented on a screen. This makes sense given that nouns and adjectives can map straightforwardly onto modal representations (especially concrete ones). However, the role of other parts of speech (particularly verbs) in integrated language processing theory is underexplored.

There are several explanations for the lack of verbs in integrated studies. Despite their presence in foundational embodied and symbolic research individually, verbs have elicited the most methodological incompatibility between frameworks. As previously discussed, embodied studies of verbs often manipulate participants' bodily positions, require large movements for responses, or observe sensorimotor cortex activation while listening to action sentences (Zwaan et al., 2012). Symbolic theories simply do not have an analogue to these kinds of manipulations. Even in more physically constrained studies, such as fictive motion (Matlock, 2004) and sentence processing experiments (Zwaan & Madden, 2005), the use of verbs has typically required response behaviors that require embodied language processing, and further, multimodal sensory and response systems (e.g., a human body).

Verbs are critical to recent debates of the role of iconicity in signed and spoken language (Lupyan & Winter, 2018; Perlman et al., 2018; Perniss et al., 2010). Researchers have even devised and conducted norming surveys of animated, iconic visual representations of abstract verbs (Scicluna & Strapparava, 2020). Pertinently, distributional semantic models often provide weaker accounts of the semantic relations between verbs than some other parts of speech (Brown et al., 2023), which may be explained by a reliance on embodied processing integration. Therefore, addressing the "verb gap" methodological conundrum in embodied language processing studies is necessary for a more comprehensive account of human language processing. The purpose of the following experiments is to extend symbolic-embodied language processing research to include verbs.

3. Experiment 1

In the studies highlighted above, some researchers have collected data regarding the vertical spatial localization of noun-based concepts to use as a representation of perceptual or embodied information encoded in language for use in analyses. In fact, a trend in the creation of perceptually based lexical norming databases has included survey ratings of vertical spatial localization as a key measurement (Miklashevsky, 2018). Therefore, to establish an analogue between noun- and verb-based research in this domain, this experiment explores vertical directionality ratings for verbs for use in similar designs.

Distributional semantic models have been used to model lexical norms obtained via rating surveys, typically in common practice dimensions like concreteness, imagery, and affect. This research has been useful in validating the importance and utility of both distributional semantic models and lexical norm data in psycholinguistic research. As perceptually based norms become more popular in research, it has become important to investigate the relationships they have with other semantic models. The analysis of vertical directionality norms using distributional semantic models may provide convergent validity to the methods described here.

Additionally, making a direct comparison between different kinds of distributional semantic models may reveal information about the nature of the information involved in the processing involved in these judgments. To illustrate, we consider a set of distributional semantic models: higher order models such as LSA (Landauer & Dumais, 1997) and Word2vec (Mikolov et al., 2013), lower-order approaches such as Google's Web 1T 5-gram (Brants & Franz, 2006) and Pointwise Mutual Information (PMI, Church & Hanks, 1990), and transformer-based models such as BERT (Devlin et al., 2019).

LSA counts the frequency with which words appear in a set of documents. The result is a large matrix with rows representing unique words, columns representing documents, and word frequencies in documents within each cell. This word-by-document matrix undergoes singular value decomposition, reducing the dimensionality of this matrix while preserving as much of the original information as possible. This transformation produces a vector in the lower dimensional space, and the cosine similarity of these vectors can be used to assess the semantic relatedness of corresponding words. Due to its ability to compare words within and across flexibly defined contexts, LSA encodes global contextual information and is representative of higher-order models.

LSA has been used to great effect in a several linguistic tasks. It can account for vocabulary acquisition in children, earn a passing grade on the Test of English as a Foreign Language (Landauer & Dumais, 1997), and it correlates strongly with human performance on semantic categorization tasks (Siakaluk, Buchanan, & Westbury, 2003). It has also been used to successfully predict several lexical norms as rated by human participants, including age of acquisition, concreteness, imagery, valence, arousal, and dominance (Bestgen & Vincze, 2012; Mandera et al., 2015). LSA can be applied to any number of corpora, but the most frequently used is known as the General Reading up to 1st year college corpus, which contains about 12 million words (Zeno, Ivens, Millard, & Duvvuri, 1995).

Word2vec is an alternative high-order natural language processing algorithm that generates word embeddings by training a neural network on a large corpus of text. The algorithm uses a sliding window to look at a small context of words around each word in the corpus and learns to predict the probability of each word given its context. Its neural network is then trained to minimize the difference between the predicted probability distribution and the actual distribution of words in the context. Research using Word2vec has shown that it is often better than count-based models LSA in predicting human data, and builds on cognitively plausible principles (Hollis, 2017; Mandera et al., 2017)

The Web 1T 5-gram corpus is not an algorithmic model itself; rather it is a massive collection of text data containing over 1 trillion word tokens taken from English language web pages (Brants & Franz, 2006). The text from these web pages is collected using a 5-word sliding window, creating chunks known as 5-grams. The frequency of each unique 5-gram within the entire corpus is simply accumulated and indexed. Due to its relatively tight observational window and large corpus size, the performance of 5-gram models is representative of lower order semantic models, which encode local, contextual relationships and syntactically dependent information. Like LSA, the Web 1T 5-gram corpus has been used as a tool to approximate and explore human semantic processing in several different paradigms. Islam and Inkpen (2010) found that an unsupervised 5-gram model performed better on a near-synonym choice task than other, higher order models (like LSA). Co-occurrence statistics using 5-grams have also been used to successfully predict human reaction times on semantic relatedness tasks (Louwerse & Jeuniaux, 2010), sensory modality judgment tasks (Louwerse & Connell, 2011), and emotional feature association tasks (Tillman & Louwerse, 2018).

Another measure of word association which is primarily based on local dependences is Pointwise Mutual Information (PMI; Church & Hanks, 1990). However, PMI also accounts for the relative frequency of the individual words within a pair, which penalizes the associative strength of common words and emphasizes rarer words. PMI has been implemented in influential distributional models and corpus studies (for example, Recchia & Jones, 2009), and has demonstrated strong associations with behavioral measures of word association (Paperno et al., 2014).

BERT represents a more modern approach to Natural Language Processing and uses a multi-layer bidirectional Transformer encoder. Functionally, this means that the immediate linguistic context surrounding a word is used as information to help disambiguate multiple word senses while remaining sensitive to long range, higher order linguistic dependencies. BERT has been applied to numerous linguistic tasks, including word pair similarity, summarization, and question answering (Devlin et al., 2019). Due to its innovative approach, BERT represents an upgraded distributional semantic model that accounts for sense specificity by providing a vector for a word in context as opposed to a single vector for all occurrences of a word.

While LSA and Word2vec encode global relationships which are based more strongly on overall gist and meaning, the 1T 5-gram corpus co-occurrence technique and PMI are more proximity-driven and encode local relationships which are based more strongly on syntactic dependencies. BERT encodes higher order relations while accounting for context specificity. If any of these models exhibit an association with human performance on a given task, then any implications or assumptions of cognitive processes reflected by that model should be investigated. The results of this study may therefore reveal some information about how language encodes spatial information. Note that the goal of this study is to compare these symbolic models to the embodied norms and not robustly compare these models to each other.

Registration information, as well as materials, data, and source code for all elements of all experiments in this study can be found on the project's Open Science Framework repository found in the supplementary material.

3.1. Method

3.1.1. Participants

Thirty-five undergraduate students from the University of Memphis (M_{age} = 20.02, *SD* = 5.69) participated in exchange for course credit in an undergraduate psychology course. All participant activity was conducted online using Qualtrics.

3.1.2. Materials

Thirty-two English words were selected for their association with either upward or downward motion (sixteen for each direction). Words associated with *up* included: *add, ascend, boost, climb, elevate, escalate, float, fly, grow, increase, inflate, jump, levitate, lift, raise,* and *rise*. Words associated with *down* included: *decline, decrease, deflate, descend, diminish, dive, drop, fall, plummet, plunge, shrink, sink, slump, subtract, topple,* and *tumble*. Table 1 displays the average properties of the words for each direction, including length, age of acquisition (Kuperman et al., 2012), concreteness (Brysbaert et al., 2014), and frequency in the 1T 5-gram corpus. Independent t-tests revealed that the two groups of words did not significantly differ on any of these metrics.

Table 1

Average word properties per direction, M (SD).

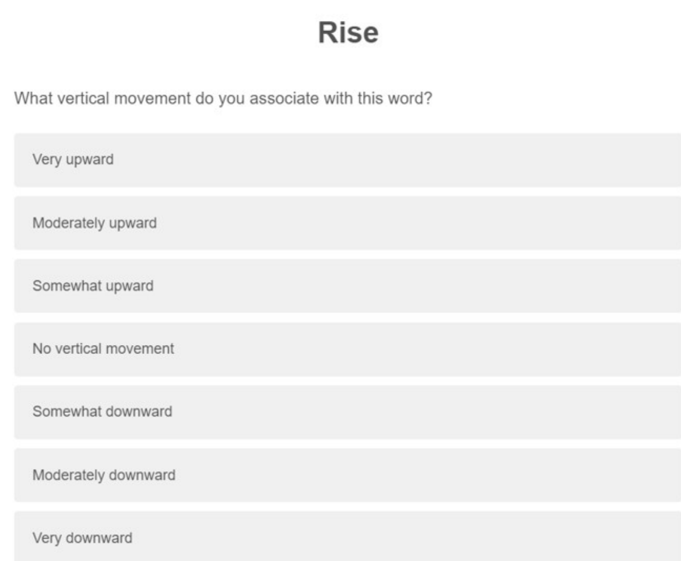
Direction	Length	AoA	Concreteness	Frequency (1T 5-gram corpus)
up	5.38 (1.75)	6.88 (2.64)	3.48 (.69)	13205304 (1312739)
down	6.06 (1.49)	7.98 (2.50)	3.47 (.62)	10778870 (2404516)

3.1.3. Procedure

Directional norms. Participants completed a survey to determine the strength of association between each word and its vertical movement. An example of the items used in the survey are included in Figure 1. To assess vertical movement, words were presented individually with the text *what vertical movement do you associate with this word?* Responses were given on a 7-point, vertically arranged Likert-type scale ranging from *very downward* to *very upward*.

Figure 1

Example of the movement association survey.



LSA. Cosine similarities between each the vector of each word and the vector of its respective direction (e.g., *rise* and *up*) were obtained using LSA's *General Reading up to 1st year college* corpus, also known as the TASA corpus (T. K. Landauer & Dumais, 1997; Zeno et al., 1995).

Web 1T 5-gram. Co-occurrence statistics between each word and its respective direction (e.g., *rise* and *up*) were obtained from the *Web 1T 5-gram* corpus, which is comprised of text samples from English language websites, and contains over one trillion word tokens (Brants & Franz, 2006). To achieve this, the corpus was analyzed by observing a sliding window of 5 words at a time. Any time both a listed word and its respective directional word were present within that window, the frequency of that 5-gram within the corpus was cumulatively added. As a result, word order does not play a role in these analyses. Log frequencies were then computed, which are typically preferred for analyses over raw frequencies due to skew (Baayen, 2001).

BERT. Cosine similarities were obtained between each word and its respective direction word using pre-trained embeddings with DistilBERT (Sanh et al., 2019). The context was specified as *The motion of [word] is what direction?*, which approximates the Likert question and provides context on both sides of the targe word. By using a contextual embedding such as this, the problems that LSA may have with multiple word senses and the problems that 5-gram may have with syntactic dependencies may each be mitigated. We also generated cosines with alternate s (available on the project's OSF page), which did not meaningfully impact our results.

Word2vec. Cosine similarities were obtained between each word and its respective direction word using Mandera, Keuleers, and Brysbaert's (2017) CBOW web interface. This model was trained on the British Web (UKWAC) and subtitles corpus.

PMI. Semantic similarity was calculated between each word and its respective direction word using PMI on the same *Web 1T* corpus used to derive 5-gram co-occurrences. To scale this metric similar to cosine values, which are bounded at [0, 1] we calculated normalized PMI (Bouma, 2009), which is bounded at [-1, 1], where -1 represents no co-occurrences, 0 represents independence, and 1 represents complete co-occurrence.

3.2. Results

Figure 2 displays a graphical representation of the norm ratings data. To adequately compare the language statistics data (i.e., within-direction groups), transformations were applied to the norm ratings data. Because antonyms like *up* and *down* appear in similar contexts, most distributional models will represent them with similar vectors. Accordingly, the response scale was broken into up and down components to avoid collinearity in cosines between target words and *up* and *down*. The *up* component is represented by 4 to 7 on the Likert scale and transformed to 0 to 3 correspondingly. The *down* component is represented by 1 to 4 on the Likert scale and transformed to 3 to 0 correspondingly (reverse scale). This transformation of the norm ratings represents the strength of the relationship between a word and its respective direction, which is more comparable than the untransformed data to the similarity values obtained using the language models. The word *levitate* was not present in the LSA TASA corpus and was excluded pairwise from these analyses.

Figure 2

Relative frequencies of responses per word.

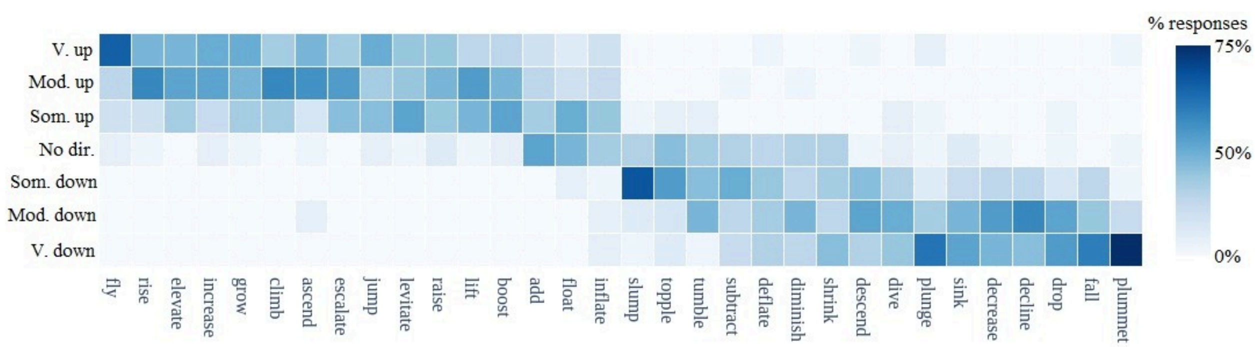


Table 2 displays Spearman's rho correlations between the norm ratings obtained and the language statistics calculated. Overall, only BERT yielded a significant positive association between with the norm ratings. When correlations for up and down components were analyzed separately, the norm ratings were significantly associated with BERT similarities for *down* words. Table 3 displays Spearman's rho correlations between all measures within direction groups. A visualization of these relationships is displayed in Figure 3.

Table 2

Spearman's rho correlations between measures (overall).

	Norms	**BERT**	**LSA**	**5-gram**	**W2V**	**PMI**
Norms						
BERT	0.44*					
LSA	0.18	0.41*				
5-gram	0.30	0.59***	0.71***			
W2V	0.20	0.21	0.73***	0.61***		
PMI	-.01	0.16	0.57***	0.29	0.62**	

Note. *p < .05, ***p < .001

Table 3

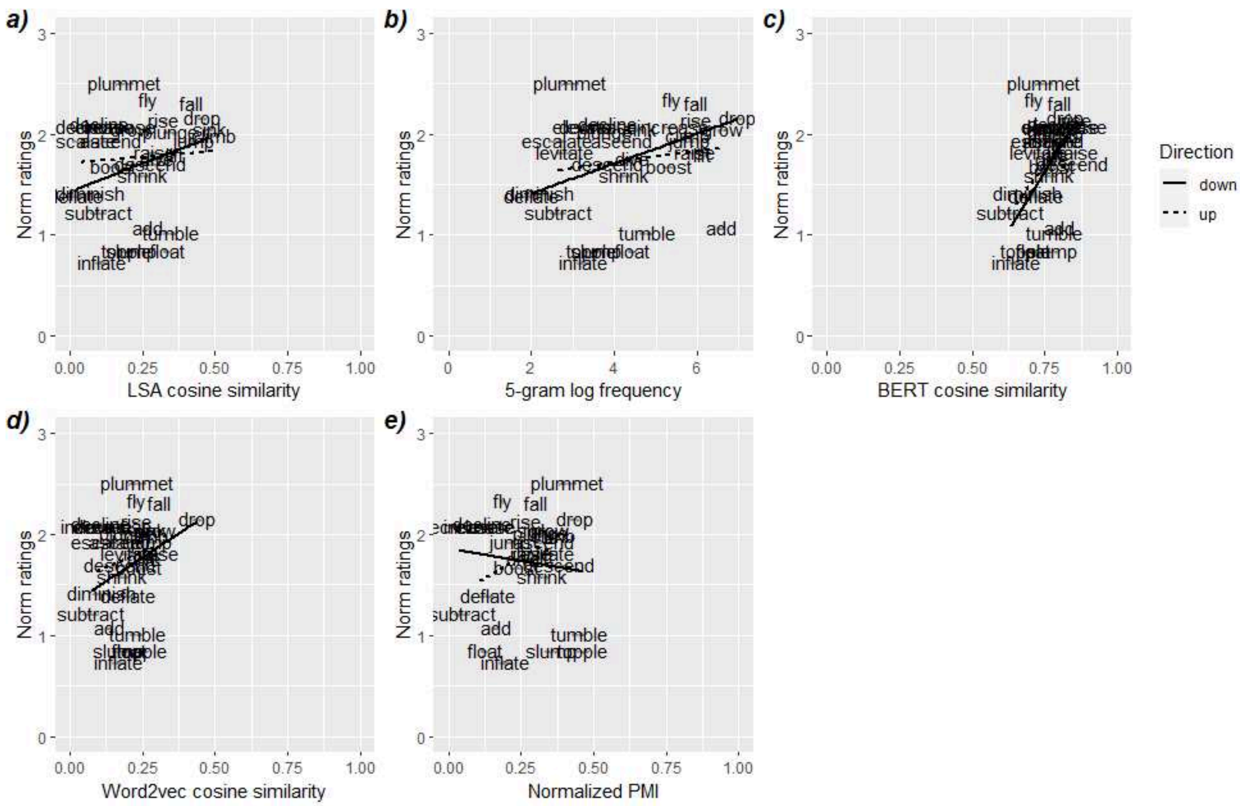
Spearman's rho correlations between measures within direction groups.

	Norms	**BERT**	**LSA**	**5-gram**	**W2V**	**PMI**
Norms		0.25	0.03	0.1	-0.02	-0.01
BERT	0.58*		0.25	0.54*	0.14	0.19
LSA	0.27	0.56*		0.63*	0.63*	0.48
5-gram	0.41	0.71**	0.85***		0.58*	0.38
W2V	0.30	0.30	0.72**	0.61*		0.73**
PMI	-0.12	0.11	0.52*	0.35	0.55*	

Note. Upper triangle contains analyses on only *up* words, lower triangle contains analyses on only *down* words.
*p < .05, **p < .01, ***p < .001

Figure 3

Relationships between norm ratings and word-direction similarities as calculated by a) LSA, b) 5-gram, c) BERT, d) Word2vec, and e) PMI.



3.3. Discussion

The cosine similarities obtained using BERT were significantly associated with the vertical movement ratings, providing convergent validity to the methods described in this study. To this point, the ratings provide a chance to observe both the overall directional grouping and the relative strength with which people associate each word with its primary vertical movement. LSA, 5-grams, Word2vec, and PMI statistics, however, were not significantly associated with the norms, though they were associated with each other and BERT. This result may indicate that the ratings are not solely explainable by simple language statistics. In other words, neither lower-order model like 5-gram nor a higher order model like LSA fully encapsulates the directional motion of the words as represented by our participants. Even BERT was not successful in explaining variability within the *up* directional group. These results are congruent with recent studies which have found verbs to be a relative weakness of distributional semantic models (Brown et al., 2023). The ratings were designed to assess vertical movement associations as determined by semantic memory, which contains perceptual and sensorimotor information (Yee et al., 2018). Since language statistics have only a loose association with these ratings, the ratings may be interpreted as uniquely representative of embodied information.

The source of the weakness of some of these language models in relation to the ratings may provide additional insight. One potential explanation of the nonsignificant effects is the use of *up* and *down* as adverbial particles in English. For example, the phrase *add up* is relatively frequent, but this is a phrasal verb which implies little, if any, physical motion. In fact, *add* is the third most frequently co-occurring listed word with *up* in the 5-gram corpus, but of the upward items, it is the third least upwardly rated word by participants. While this example applies most strongly to lower order models like 5-gram, discrepancies like this demonstrate how human rated associations differ from those derived by some distributional semantic models. Syntactic bindings (like *add up*) can cause a mismatch between distributional semantic models and semantic knowledge derived from perceptual, multimodal, embodied human experience. Accordingly, these results underscore the value of human ratings as indicative of perceptually derived semantic associations.

Correlations between the language models and the ratings were generally weaker within the *up* word group than those of the *down* word group. There are several possible explanations for this effect. First, the participants' ratings of *down* words appear to be more likely to be concentrated around a particular response option, while *up* words tend to be more broadly distributed. This may be related to a second explanation, that *up* and *down* words are subject to different linguistic tendencies. While the two words are direct opposites in their most literal sense, their many uses in the English language may skew or blur their "opposite" nature in representations as calculated by distributional semantic models, which often struggle to distinguish between contextually substitutable antonyms. Finally, this selection of *up* and *down* words is far from exhaustive, so the asymmetrical correlations may be simply due to word choice in a small sample.

The vertical movement ratings obtained in this study may provide a preliminary example of their validity and possible utility in various psycholinguistic paradigms. This is especially true in that the words used in this study are verbs, which are neglected in many laboratory studies of symbolic-embodied integrated semantic processing. These types of norms should provide a quantified factor that may be apt for use in models seeking to explain effects found in behavioral studies, particularly those testing semantic processing within an embodied cognition framework. Additionally, the verbs used in this study vary along many other dimensions, perhaps most interestingly in abstractness. The word list includes both bodily action words and words that only indicate motion in an abstract sense. These norms may encourage the underrepresented symbolic-embodied integrated study of verbs in two ways. First, these kinds of data may be used to model human performance in laboratory procedures, generating new testable hypotheses. Second, these and similar verbs that are not yet represented in laboratory paradigms may be occasioned to be included in future, innovative designs. Further, the method for obtaining these ratings may be adapted to build a database of similar norms for larger sets of words.

4. Experiment 2

One of the most fruitful lines of integrated language processing research involves vertical word positioning experiments. These studies have primarily included nouns (mostly concrete, but some abstract). To most closely match designs in previous literature, this experiment extends single-word high-low presentation experiments (Meier & Robinson, 2004; Pecher et al., 2010; Šetić & Domijan, 2007) to include verbs. These types of studies produced a word-referent congruency facilitation effect, in which words presented in a position that is congruent with their referent's associated location (e.g., *bird* presented high) are processed faster than when this relationship is incongruent (e.g., *bird* presented low). In this study, participants were asked to make similar judgments about verbs, which include words that indicate motion either upward (e.g., *rise*) or downward (e.g., *fall*). To seek replication of the congruency facilitation effect, conditions were created in which the verbs were presented either congruently or incongruently with the endpoint of their implied direction. The primary goal of this experiment is to attempt to replicate the stimulus-location congruency facilitation effect found in previous literature with nouns. This effect generally supports embodied language processing theories. Replicating this effect is a preliminary step toward using verbs to examine symbolic-embodied integrated theories because it would establish whether verbs elicit similar processing effects as nouns in comparable paradigms. If so, follow-up experimental manipulations using verbs can help untangle the relationship between embodied and symbolic processing.

4.1. Method

4.1.1. Participants

Sixty-nine participants (M_{age} = 22.90, *SD* = 8.99) were recruited from the University of Memphis undergraduate research pool by offering course credit. Participants completed the study online, in-browser, using their own computers.

4.1.2. Materials

Verbs indicating either upward or downward movement were presented to participants. The words used in this experiment are identical to those used in Experiment 1. The experiment was programmed using the jsPsych library (de Leeuw, 2015).

4.1.3. Design

The experiment utilized a single factor, two-level (high and low-presented stimuli), within-subjects design. Primary dependent measures included reaction times and error rates.

4.1.4. Procedure

After the informed consent procedure, participants received instructions, a series of practice trials, and testing trials. During testing trials, participants were presented with individual words from the list. All words were used twice, once in a congruent position (e.g., *rise* presented high), and once in an incongruent presentation (e.g., *rise* presented low). Words and congruency condition were selected randomly, without replacement. Words were presented in black, 40-point Arial font on a white full-screen background. High and low presentations were placed at the 5^{th} and 95^{th} percent of each participant's screen from the top, horizontally centered. A fixation cross appeared in the center of the screen for 500ms, followed by an ISI jitter of 500-700ms. The words remained on the screen until the response. After each trial, a blank screen intertrial jitter of 800-1200ms took place. Before testing, participants received instructions for the task and 12 practice trials using the words *up* and *down*. The practice trials contained feedback and were discarded from analyses. Participants repeated the practice trial block until they achieved at least 80% correctness on a repetition of the 12-trial sequence.

Participants were instructed at the beginning of the task to indicate whether the word was presented in a location that matches the direction represented by the word. Responses were made using f and j keys the keyboard, which corresponded to *congruent* and *incongruent* responses randomly assigned for each participant. These keys were selected due to their vertical evenness on a typical keyboard, as research shows that vertically disparate response options may confound location-cue congruency effects (Petrova et al., 2018).

4.2. Results and discussion

Consistent with similar analyses by Louwerse and Jeuniaux (2010), trials in which RT was below 200ms and more extreme than 2.5 standard deviations from an individual's mean were removed from analyses (3.0% of the data). The average error rate (9.4%) was well below exclusion thresholds used in similar studies (Hutchinson & Louwerse, 2013; Pecher et al., 2010). As there were no outliers in participants' error rates; all participants were included in these analyses.

4.2.1. Expected effects - results

In order to determine the individual and combined effects of location-cue congruency on reaction time, a linear mixed-effects model was constructed in R (Team R Development Core, 2021) using the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages. The model comparison process and final model selection were designed to comply with the recommendations of Barr et al. (2013) to use the maximal random effects structure supported by the data, as well as guidelines from Meteyard and Davies (2020) for best practices in reporting mixed-effects modelling. Like many similar studies analyzing reaction time data, only correct trials were included in this analysis. R code for all analytic procedures can be found on the project's OSF repository in the supplementary material. Reaction times, the dependent variable, were highly skewed even after filtering for outlier trials as described above (skewness = 4.64), and were logarithmically transformed (logRT skewness = -1.02). Visual inspection of Q-Q and residual plots (available as supplemental material on the project's OSF repository) suggest that this transformation improved the model's homoskedasticity and normality of residuals. Fixed effects in this model included cue location (high or low), word direction (up or down), and their interaction. The random effect structure was determined by entering the maximal structure (including subject and item intercepts) and systematically removing slope terms until convergence was achieved. If models with equally complex structures converged, AIC was used to select between them. This analysis revealed a significant interaction, as well as significant main effects of cue location (see Table 4 for model summary). Pairwise contrasts (Tukey-HSD) indicate that this interaction is driven by the significant difference between high and low presentations of *up* words ($p < .001$). The contrasts between *up* and *down* words presented high was also significant ($p < .001$), as was the contrast between high *up* words and low *down* words ($p < .001$); no other contrasts were significant. Figure 4 graphically represents group means, using raw RT rather than log transformed RT for interpretability.

Table 4

Results of linear mixed effects model of congruency predicting log RT in correct trials

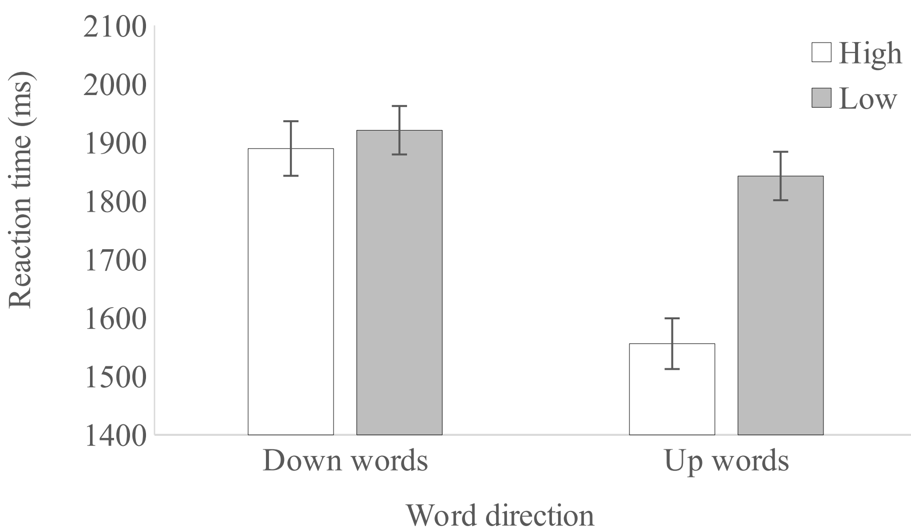
	Fixed Effects		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	3.22	3.18 – 3.26	**<0.001**
Cue location (0 = low)	-0.02	-0.04 – -0.00	**0.039**
Word direction (0 = down)	-0.01	-0.04 – 0.01	0.294
Cue location x word direction	-0.07	-0.09 – -0.04	**<0.001**
	Random Effects		
σ^2			0.03
τ_{00} subject			0.02
τ_{00} word			0
τ_{11} subject.cue_lochi			0
τ_{11} directionhi	0.00		
ρ_{01} subject			0.66

$\rho_{01\ word}$	-0.32
ICC	0.48
$N_{subject}$	69
N_{word}	32
Observations	3881
Marginal R^2 / Conditional R^2	0.027 / 0.491

Note. Model equation: logRT ~ cuelocation * worddirection + (1 + cuelocation|subject) + (1 + direction|word)
τ = estimated variance of random effects
ρ = correlation of random effects

Figure 4

Reaction times by word direction and location for correct trials



Note: Figure displays group means and standard errors.

To further determine the effects of location-cue congruency on response accuracy, we constructed a mixed-effects binomial logistic regression model predicting response accuracy using all trials. Model selection procedure was conducted similarly to the procedure used to predict reaction times. Visual inspection of residual plots did not reveal violations of the assumptions of logistic mixed-effects modeling. This analysis also revealed a significant interaction (see Table 5 for model summary and formula). Again, pairwise contrasts (Tukey-HSD) indicate that this interaction is driven by the significant difference between high and low presentations of *up* words ($p < .01$). The contrasts between *up* and *down* words presented high was also significant ($p < .01$), as was the contrast between high *up* words and low *down* words ($p < .001$); no other contrasts were significant. Group means and standard errors visualized in Figure 5.

Table 5

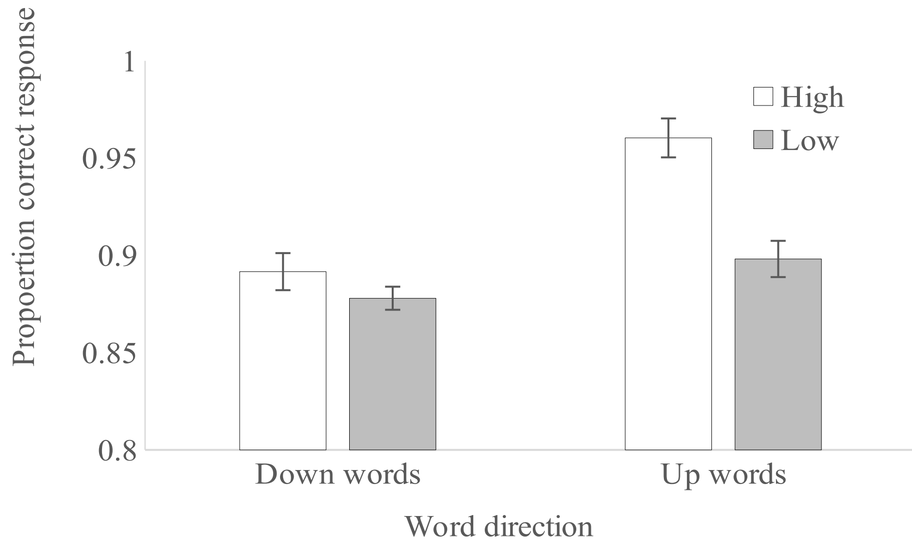
Results of mixed effects model of congruency predicting response accuracy

	Fixed Effects		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	12.61	8.03 – 19.80	**<0.001**
Cue location (0 = low)	0.98	0.67 – 1.43	0.930
Word direction (0 = down)	1.15	0.69 – 1.90	0.589
Cue location x word direction	2.45	1.54 – 3.89	**<0.001**
Random Effects			
σ^2			3.29
$\tau_{00\ subject}$			1.2
$\tau_{00\ word}$			0.35
$\tau_{11\ subject.cue\_lochi}$			0.8
$\rho_{01\ subject}$			-0.73
ICC			0.27
N_{word}			32
$N_{subject}$			69
Observations			4279
Marginal R^2 / Conditional R^2	0.039 / 0.301		

Note. Model equation: accuracy ~ cuelocation * direction + (1 + cuelocation |subject) + (1 | word), family = binomial
τ = estimated variance of random effects
ρ = correlation of random effects

Figure 5

Response accuracy by word direction and location for all trials



Note: Figure represents group means and standard errors.

4.2.2. Expected effects – discussion

These analyses represent a generally successful verb-based replication and extension of previous location-cue congruency experiments. More specifically, *up* words presented high (location-cue congruent) yielded faster reaction times and higher accuracy than *up* words were low (location-cue incongruent). These results suggest that semantic processing can be facilitated by location-cue congruence. This matches similar effects found with sets of nouns and adjectives used in previous research (including, at times, asymmetries with words presented high), providing support for embodied language processing theories (Dudschig et al., 2013; Lachmair et al., 2011). The embodied framework generally posits that semantic processing includes neurocognitive simulation of perceptuomotor information. Previous location-cue facilitation experiments using nouns support this position because the typical height of objects in the visual field appears to influence processing speed during semantic judgment tasks. Replicating this effect with verbs widens the scope of perceptuomotor information that may be responsible for these effects. While visual information is certainly still likely a strong driver of these effects, verbs presented in isolation without a subject or object provide less of a readily available visual representation than concrete nouns. As a result, information from other modalities (especially motoric and tactile) are likely involved in the effect.

Other cue-location studies have also found asymmetries between high and low or upward and downward stimuli (Lakens, 2012; Pecher et al., 2010; Dudschig et al., 2013). One possible explanation of this effect is that this study was conducted online, in web browsers, and without explicit instructions about viewing setup or posture. It is likely that participants may have completed the tasks in a variety of manners that would not reflect a typical laboratory setup. Since we can't be sure whether participants completed the study at a desk with a typical viewing angle of their monitors, "high" and "low" stimulus presentations may not have presented with equal viewing angle differences with respect to each other or the fixation cross. However, other studies that find this asymmetry have explained them with polarity correspondence.

Polarity correspondence has been offered in the literature as an explanation of experimental findings in several binary choice judgment tasks, including location-cue congruence paradigms (see Proctor & Xiong, 2015 for an overview). The simplest description of this principle is that "*For a variety of binary classification tasks, people code the stimulus alternatives and the response alternatives as + polarity and - polarity, and response selection is faster when the polarities correspond than when they do not*" (Proctor & Cho, 2006, p. 418). While the predictions generated by this principle seem straightforward (e.g., that high *up* words and low *down* words would elicit the most processing benefits), there are other interpretations that offer more complex sets of predictions. For example, Lakens (2012) offers descriptions of how +polar stimuli elicit processing benefits over -polar stimuli, which may result in all-positive polar stimuli (*up* high words) receiving more processing benefits than all-negative polar stimuli (*down* low words). This is consistent with the pattern observed in this experiment, suggesting that polarity correspondence may be an alternative explanation of these findings. However, we did not observe a facilitation for low down words predicted by other interpretations of polarity (Proctor & Xiong, 2015), and some research suggests that polarity predictions do not account for effects found in other conceptual-perceptual alignment studies (see Dudschig & Kaup, 2017; Dolscheid & Casasanto, 2015). While a polarity alignment account cannot be discounted based on the results of this experiment alone, future studies should investigate whether polarity or sensorimotor simulation is the primary driver of effects such as these by controlling for more polar dimensions or involving neurocognitive designs.

4.2.3. Exploratory analysis: symbolic and embodied factors - results

The preceding analyses serve to replicate and extend location-cue congruence effect to verbs, the results of which may be used to infer evidence for embodied language processing. However, a novel, influential approach in related research involves operationalizing the perceptual and linguistic relations in an experiment and examining how they differentially relate to cognitive processing during an experimental task. For example, Louwerse and Jeuniaux, (2010) used word order frequency (linguistic) and human iconicity ratings (embodied) to predict reaction times in various judgment tasks. This approach allows for a more direct examination of whether symbolic or embodied processes are influencing processing in a given task. Following this lead, we sought to examine how the current experiment may be situated in the literature as either symbolic or embodied-dominant. We operationalized an embodied factor as the average directional strength ratings for each word obtained in Experiment 1. We operationalized a linguistic factor as the cosine similarity between each word and its respective anchor word (*up* or *down*) using BERT (obtained as described in Experiment 1).

We then constructed a model predicting log-transformed reaction times on correct trials. The BERT and norm ratings as described above were entered as fixed effects, as was as their interaction with word direction (*up* or *down*). This interaction allows us to examine whether any facilitative effects are due to fine grain, within-direction magnitudes as defined by our embodied and linguistic factors, or due to coarse grain, between-direction categorizations. All continuous measures were centered. As with previous models, subject and item as random effects (intercepts), and the random effect structure was determined by entering the maximal structure and systematically removing slope terms until convergence was achieved. This analysis yielded a significant negative association between norm ratings and reaction times, and *up* words elicited faster reaction times than *down* words. There were no significant interactions, and no significant main effect of BERT (see Table 6). This pattern of results indicates that norm magnitude, regardless of direction, facilitates reaction time in addition to the facilitation of reaction time for *up* words. We then similarly constructed a logistic model to predict accuracy rates. In this model, only the main effect of word direction was significant. The results of this model are displayed in Table 7.

Table 6

Linguistic and embodied factors predicting reaction times

Fixed Effects			
Predictors	*Estimates*	*CI*	*p*
(Intercept)	3.2	3.16 – 3.25	**<0.001**
Norms	-0.04	-0.07 – -0.00	**0.032**
Word direction (0 = down)	-0.05	-0.07 – -0.03	**<0.001**
BERT	0.02	-0.33 – 0.38	0.911
Norms x direction	0.02	-0.03 – 0.08	0.359
BERT x direction	-0.34	-0.79 – 0.11	0.135
Random Effects			
σ^2			0.03
$\tau_{00\ subject}$			0.03
$\tau_{00\ word}$			0

$\tau_{11\ \text{word.BERT}}$	0.04
$\rho_{01\ \text{word}}$	-1
ICC	0.46
N_{subject}	69
N_{word}	32
Observations	3881
Marginal R^2 / Conditional R^2	0.015 / 0.468

Note. Model equation: logRT ~ BERT*worddirection + norms*worddirection + (1|subject) + (1 + BERT|word)
τ = estimated variance of random effects
ρ = correlation of random effects

Table 7

Linguistic and embodied factors predicting accuracy rates

	Fixed Effects		
Predictors	*Std. Odds Ratio*	*CI*	*p*
(Intercept)	12.14	8.44 – 17.45	**<0.001**
BERT	1.27	0.90 – 1.78	0.175
Norms	1.15	0.83 – 1.59	0.392
Word direction (0 = down)	1.56	1.02 – 2.40	**0.04**
BERT x direction	0.88	0.54 – 1.44	0.617
Norms x direction	0.94	0.58 – 1.53	0.808
	Random Effects		
σ^2			3.29
$\tau_{00\ \text{subject}}$			0.69
$\tau_{00\ \text{word}}$			0.27
ICC			0.23
N_{word}			32
N_{subject}			69
Observations			4279
Marginal R^2 / Conditional R^2	0.029 / 0.248		

Note. Model equation: accuracy ~ BERT*worddirection + norms*worddirection + (1|subject) + (1|word)
τ = estimated variance of random effects
ρ = correlation of random effects

4.2.4. Exploratory analysis: symbolic and embodied factors – discussion

To examine the task specificity of embodiment in language processing, a growing body of literature suggests that we can determine whether symbolic or perceptual processing is dominant in a task by observing the relative association between language statistics or embodied information on reaction times and accuracy (Louwerse, 2018). We sought to determine how embodied or symbolic the task in this experiment was to determine its place in the broader literature. In short, the embodied factor was significantly associated with reaction times, while the linguistic factor was not. This contrasts with several similar studies (Hutchinson & Louwerse, 2013; Louwerse & Hutchinson, 2012; Louwerse & Jeuniaux, 2012) which have found that language statistics tend to either dominate or evenly share dominance with embodied factors in comparably structured experiments. There are a few possible explanations for this contrast. First, the operationalization of each factor differs slightly between each study (e.g., the use of BERT versus LSA, word order frequency versus semantic similarity scores, etc.). Second, slight differences in task demands may place more emphasis on one system over the other. In other words, this particular task might have been more demanding of the perceptual system than other tasks in this literature. Third, the use of verbs as stimuli may elicit the involvement of embodied processing more than nouns. Additionally, distributional semantic models have been demonstrated to construct weaker representations for verbs in comparison to other parts of speech (Brown et al., 2023), so linguistic factors operationalized in this way may be at a disadvantage. Future research should seek to examine the sensitivity of similar effects to these methodological elements (and their combination).

While the embodied factor was significantly associated with faster reaction times, it was not associated with increased accuracy. This is unusual, as speed and accuracy are often found to be tightly linked in studies with similar paradigms (Pecher et al., 2010). However, the word direction factor was significantly associated with both reaction time and accuracy, suggesting that, in this case, support for the judgments being made in the task may have been due to between-direction category distinctions more than within-direction directional magnitude, contrary to what was found in the reaction time model. Future research may need to examine cases where judgment speed and accuracy yield asymmetric patterns of results to probe for a more generalizable explanation.

4.3. Experiment 2 discussion

This experiment extends upon previous work in location-cue congruence paradigms by using verbs to replicate effects found typically using nouns. It further extends this line of inquiry by introducing analyses that probe the cognitive explanation for the observed effects. We found that a verb presented in a manner congruent with its directional meaning yielded faster, more accurate judgments for *up* words presented high. Further analyses suggest that polarity alignment is not a comprehensive explanation for this effect. Rather, we found that the embodied factor (perceptually based human ratings) is more strongly associated with reaction times than the linguistic factor (language statistics), which supports accounts of language processing involving perceptual simulation for comprehension.

One plausible explanation for similar studies involving nouns is that the height of physical referents in the visual field creates a priming effect which facilitates processing during lexical judgment tasks. Extending this explanation to include verbs expands the scope of perceptuomotor information that may be responsible for these effects. In this experiment, verbs are likely eliciting motoric, tactile, or other interoceptive information since they were presented without subjects or objects. Embodied language processing theory readily accounts for non-visual perceptual information in this way.

However, task specificity cannot be discounted as an explanation for the results of a single experiment. Whether these results are due to the demands of the task or due to the psycholinguistic processing of verbs must still be disentangled. Further, studies which compare the relative influence of linguistic and embodied factors across tasks often do so in between-subjects designs (Louwerse & Jeuniaux, 2012; Louwerse & Jeuniaux, 2010). However, within-subjects designs investigating the same hypotheses would provide more powerful evidence regarding this theory. Accordingly, we devised an additional experiment that leverages the unique properties of verbs to create a within-subjects design to further examine the influence of task on symbolic and embodied language processing,

5. Experiment 3

One challenge to studying verbs is that they generally signify an event with takes place over time using relational information, which necessitates complex dynamic processing (Richter et al., 2021). This poses a problem to the methods described thus far, which present words in static positions. Research in signed languages suggests that signed verbs are often particularly rich with iconicity because several types of motion may be represented over time in a single sign (Perniss et al., 2010). Representing the meaning of motion verbs in a computer-based study may therefore require animations that represent motion more directly than static pictures can.

In this study, using the same verbs, participants made semantic judgments of simple animations of vertically moving objects. Conditions were created in which the motion of the objects were either congruent or incongruent with the meaning of each verb. Previous studies using similar paradigms have manipulated the task in some way in order to occasion differences in the influence of symbolic and embodied processing (Dudschig & Kaup, 2017; Lachmair et al., 2011).This study offers a novel task manipulation that is designed to accomplish this goal while accommodating the unique nature of verbs. This animated object (or movement cue) consisted of a “scrambled” version of each word, consistent with recommendations of luminance control in pupillometry research (Mathôt, 2018). This process renders the image of each word illegible, but with the same luminance properties as the word itself. By decoupling the word and movement cues and presenting them serially, the sequence of whether the word or movement cue appears first was manipulated. In both conditions, participants make their judgments (whether the cue is moving in the implied direction of the verb) during the final phase of each sequence. The purpose of this manipulation is to influence how either symbolic or embodied processes are recruited during participants' judgments.

In the word-first condition, participants are allowed time to process the word linguistically before having to make a judgment about a simple movement cue. According to the linguistic shortcut hypothesis, linguistic processes are recruited first in this condition and are sufficient to make a judgment, the slower embodied system is preempted, and reaction times should be explained by linguistic, statistical relationships between words. In the cue-first condition, participants view only the movement cue until the revelation of the word, delaying linguistic processing (and judgment timing) until after perceptual processing is underway by means of the movement cue. Therefore, reaction times in this condition should be explained by embodied, perceptually derived associations.

5.1. Method

5.1.2. Participants

Seventy-three participants (M_{age} = 22.04, *SD* = 4.89) were recruited from the University of Memphis undergraduate research pool by offering course credit. Participants completed the study online, in-browser, using their own computers.

5.1.3. Materials

Verbs indicating movement either upward or downward were presented to participants. The words used in this experiment are the same as in Experiments 1 and 2.

To create the movement cues, an image of each word was horizontally mirrored and divided into 100 equally sized squares. These squares were randomly rearranged and rotated to one of four orientations (0-270 degrees at 90-degree intervals). An example of the word *rise* scrambled in this way is presented in Figure 6. Four different cues were created for each word and implemented at random during the experiment so that participants could not associate a cue with its word. This process was carried out in MATLAB. The rest of the experiment was programmed using the jsPsych library (de Leeuw, 2015).

Figure 6

Example stimulus (rise) used for movement cues.

To determine the relative influence of symbolic and embodied processing in this experiment, each was operationalized using methods similar to Louwerse and Jeuniaux (2010). The embodied factor was operationalized as the movement ratings for each word obtained in Experiment 1. The symbolic (linguistic) factor is operationalized as the cosine similarity between each word and its respective anchor word (*up* or *down*) using BERT as defined in Experiment 1.

5.1.4. Design

The experiment utilized a 2x2 within-subjects design. The first factor was the sequencing manipulation, which contains word-first and cue-first levels. The second factor was the directional movement of the cue manipulation, which contains up and down levels. Primary dependent measures included reaction times and error rates.

5.1.5. Procedure

After the informed consent procedure, participants were randomly assigned to one of two starting conditions: word-first or cue-first. Each participant completed both conditions. All words were used twice in each condition, once with congruent movement, and once with incongruent movement. The order of word and movement congruency were chosen at random without replacement. Words were presented in black, 40-point Arial font on a white background. In each condition, instructions for the task and 12 practice trials using the words *up* and *down* commenced. The practice trials contained feedback and were discarded from analyses. Participants repeated the practice trial block until they achieved at least 80% correctness on a repetition of the 12-trial sequence. Test trials appeared as follows.

In the word-first condition, a fixation cross appeared at the center of the screen for 500ms. Then, a word and movement congruency condition were randomly chosen as described above. The word appeared in the center of the screen, stationary, for 500ms, followed by a blank screen ISI jitter of 500-700ms. Then, one of the movement cues created for that word was chosen at random. Because the experiment is designed to adapt to multiple screen sizes and resolutions, the word was programmed to move from the midpoint of the screen to an endpoint located at 95% of the total screen height either upward or downward in exactly 2000ms. The speed at which the cue moved remained constant from its appearance until it reached its destination (i.e., there is no apparent acceleration or deceleration). Participants were instructed at the beginning of the block to indicate whether the cue is moving in the direction represented by the word. Responses were made using the keyboard using the *f* and *j* keys, which corresponded to *congruent* and *incongruent* responses randomly for each participant. These keys were selected due to their vertical evenness on a typical keyboard, as research shows that vertically disparate response options may confound location-cue congruency effects (Petrova et al., 2018). Response timing began when the movement cue appeared. Stimulus presentation speeds were based on related semantic similarity judgment tasks that use serial stimulus presentation (Pu et al., 2019), though this novel task manipulation does not provide a perfect analogue to such designs. After each trial, a blank screen intertrial jitter of 800-1200ms took place.

In the cue-first condition, a fixation cross appeared at the center of the screen for 500ms. Then, a word and movement congruency condition were randomly chosen. One of the movement cues created for that word was randomly chosen. The cue's movement is designed exactly like the movement in the word-first condition. Once the cue reached its destination, it disappeared, and a blank screen ISI jitter of 500-700ms occurred. Then the selected word appeared in the center of the screen and remained stationary there for 2000ms. Responses were made using the f and j keys on the keyboard, assigned randomly for each participant. Response timing began when the word appeared. Feedback and practice trials occur in the same manner as the word-first condition, though the practice trials are adapted to reflect the presentation of this condition. After each trial, an intertrial jitter of 800-1200ms took place in the form of a blank screen before the next trial began.

5.2. Results and discussion

Consistent with similar analyses by Louwerse and Jeuniaux (2010), trials in which RT was below 200ms and more extreme than 2.5 standard deviations from an individual's mean were removed from analyses (2.5% of the data). The error rate (M = 8.0%) was well below exclusion thresholds used in similar studies by Louwerse and Jeuniaux (2010) and Zwaan and Yaxley (2003). As there were no outliers in participants' error rates, all participants were included in these analyses.

5.2.1. Expected effects - results

In order to determine the individual and combined effects of directional congruency and word-movement presentation order on reaction time, a linear mixed-effects model was constructed in R (Team R Development Core, 2021) using the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages. Similar to Experiment 2, the model construction and selection processes were designed to comply with the recommendations of Barr et al. (2013) and Meteyard and Davies (2020). Again, for the analysis of reaction time data, only correct trials were included. Reaction times, the dependent variable, were highly skewed even after filtering for outlier trials as described above (skewness = 5.92) and were logarithmically transformed (logRT skewness = 0.85). Visual inspection of Q-Q and residual plots (available as supplemental material on the project's OSF repository) suggest that this transformation improved the model's homoskedasticity and normality of residuals. Fixed effects included cue movement direction, word meaning direction, and order (cue-first or word-first). The random effect structure was determined by entering the maximal structure and systematically removing terms until convergence was achieved. R code for all analytic procedures can be found on the project's OSF repository in the supplementary material.

Visual inspection of residual plots revealed no obvious violations of the assumptions of linear mixed-effects modeling. This analysis revealed significant interactions between cue movement and order, as well as cue movement and word direction, with significant main effects of order and cue movement (see Table 8 for model summary and formula). Post-hoc contrasts (Tukey-HSD) indicate that the interactions are driven by the significant difference between *up* words presented with upward moving cues and *up* words with downward moving cues, as this contrast was significant in both the cue first ($p < .001$) and word first ($p < .001$) conditions, while the same was not true for *down* words. No other contrasts were significant. Figure 7 graphically represents group means and standard errors, using raw RT rather than log transformed RT for interpretability.

Table 8

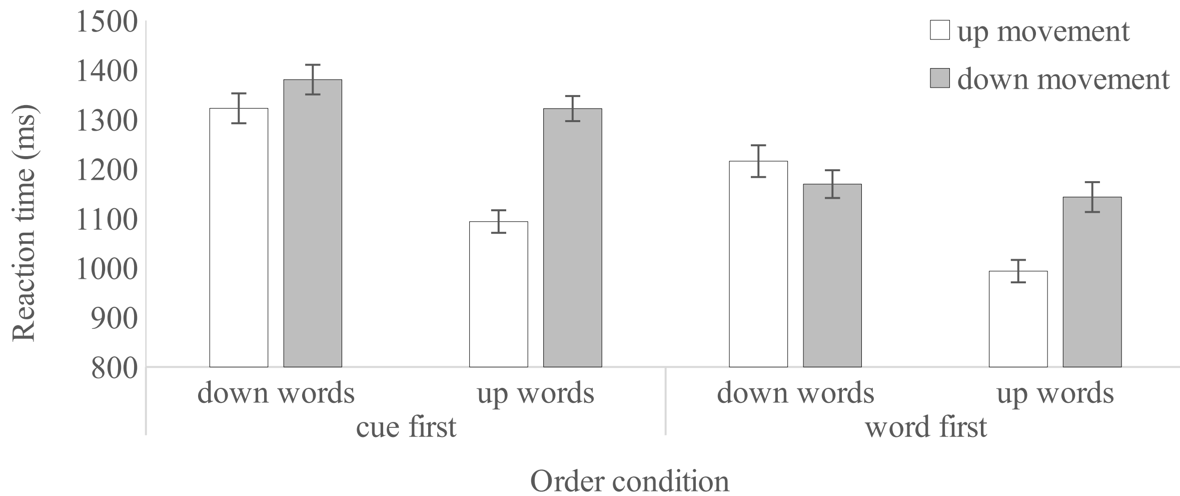
Results of linear mixed effects model of experimental conditions predicting log RT.

Fixed Effects			
Predictors	*Estimates*	*CI*	*p*
(Intercept)	3.09	3.06 – 3.13	**<0.001**
Cue movement (0 = down)	-0.02	-0.04 – -0.01	**0.001**
Word direction (0 = down)	-0.02	-0.04 – 0.00	0.092
Order (0 = cue first)	-0.09	-0.10 – -0.07	**<0.001**
Cue movement x word direction	-0.06	-0.08 – -0.04	**<0.001**
Cue movement x order	0.05	0.03 – 0.07	**<0.001**
Word direction x order	0	-0.02 – 0.02	0.986
Cue movement x word direction x order	-0.01	-0.04 – 0.01	0.312
Random Effects			
σ^2			0.03
$\tau_{00\ subject}$			0.02
$\tau_{00\ word}$			0
$\tau_{11\ subject.word\_dirup}$			0
$\rho_{01\ subject}$			-0.34
ICC			0.37
$N_{subject}$			73
N_{word}			32
Observations			8418
Marginal R^2 / Conditional R^2	0.057 / 0.408		

Note. Model equation: logRT ~ cuemovement*worddirection*order + (1 + worddirection| subject) + (1| word)

Figure 7

Reaction time by order, word meaning, and cue movement conditions



Note: Figure represents group means and standard errors.

To further determine the individual and combined effects of these factors on response accuracy, we constructed a similar mixed-effects logistic regression model using all trials. Model selection procedure was conducted similarly to the procedure used to predict reaction times. This analysis revealed significant interactions between cue movement and word direction, as well as word direction and order (see Table 9 and Figure 8). Post-hoc contrasts (Tukey-HSD) indicate that the significant interactions are driven by the significant difference between upward and downward cue movement for *up* words in the cue-first condition ($p < .001$), but not the word-first condition. No other contrasts were statistically significant.

Table 9

Results of logistic mixed effects model of experimental conditions predicting accuracy

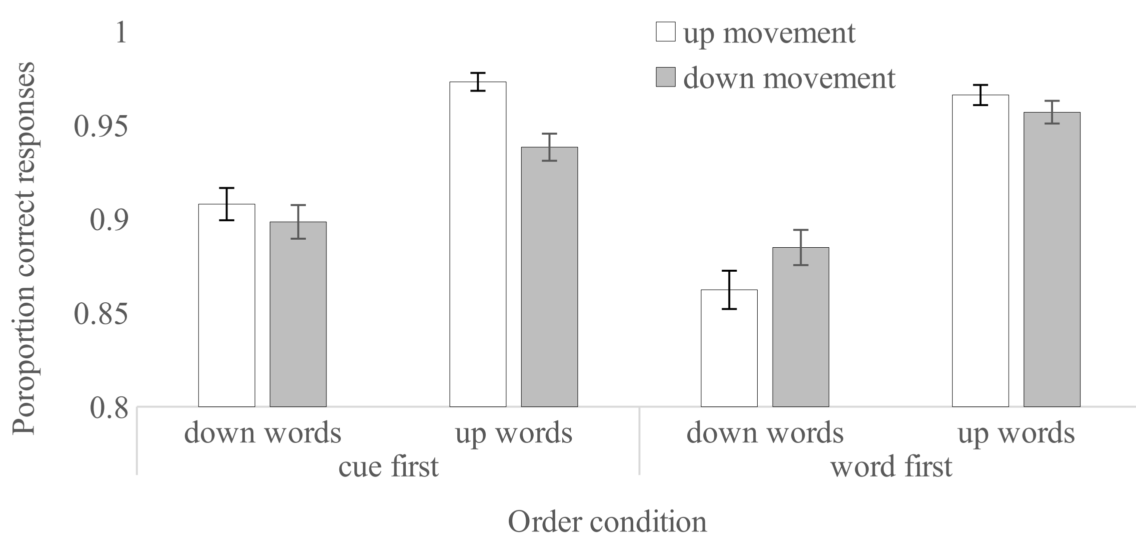
	Fixed Effects		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	12.88	8.49 – 19.53	**<0.001**
Cue movement (0 = down)	1.14	0.85 – 1.54	0.369
Word direction (0 = down)	1.59	0.91 – 2.78	0.104

Order (0 = cue first)	0.86	0.65 – 1.14	0.305
Cue movement x word direction	2.17	1.29 – 3.65	**0.003**
Cue movement x order	0.68	0.46 – 1.02	0.06
Word direction x order	1.72	1.08 – 2.74	**0.023**
Cue movement x word direction x order	0.76	0.37 – 1.57	0.466
Random Effects			
σ^2			3.29
$\tau_{00\ subject}$			0.45
$\tau_{00\ word}$			0.43
ICC			0.21
N_{word}			32
$N_{subject}$			73
Observations			9115
Marginal R^2 / Conditional R^2	0.079 / 0.274		

Note. Model equation: accuracy ~ cuemovement*worddirection*order + (1 |subject) + (1 |word), family = binomial
τ = estimated variance of random effects
ρ = correlation of random effects

Figure 8

Accuracy by order, word meaning, and cue movement conditions



Note: Figure represents group means and standard errors.

5.2.2. Expected effects - discussion

We observed reaction time facilitation for movement-meaning congruence, primarily up-meaning words with upward cue movement, and we observed an accuracy facilitation for up-meaning words with upward cue movement. These effects generally support the previous location-cue congruency experiments (see Estes et al., 2015; Estes & Barsalou, 2018). Specifically, the facilitative congruency effect has generalized from static location cues to dynamic movement cues. As this effect generalizes (at least so far in spatial domains), other features may be considered for psycholinguistic investigations of sign-referent congruency effects.

An additional objective of this experiment was to determine whether the decoupling and manipulated sequencing of a verb and its movement cue could reveal information about embodied and symbolic language processing. The order manipulation yielded a significant effect – participants were faster to make their judgments during the word-first condition than the cue-first condition. This effect supports linguistic shortcut hypotheses. In the word-first condition, reading the target word allows quick, associative, linguistic processes to quickly form a directional hypothesis, preempting the need for embodied semantic processing and allowing for a quicker judgment. The embodied or perceptual language processing, in this case, is cut short, and appears to only be needed to check for the movement cue's consistency with the narrow directional hypothesis. In the cue-first condition, perceptual information does not encode any particular word. Rather, multiple possible words may be partially preactivated by a movement cue, so when the word appears, the remainder of word recognition processes must still run to completion, causing slower judgments. However, as the accuracy results show, these slower judgments are also more accurate. The contrasting effect of order condition on reaction time (word first better) and accuracy (cue first better) adds a new dimension to the linguistic shortcut hypothesis: linguistic shortcuts may have a cost. Taken together, these results provide a detailed depiction of the nature and constraints of the dynamic and interconnected roles of embodied and symbolic language processing systems.

As discussed in Experiment 2, polarity correspondence may bear on these results. However, while some of the observed effects are consistent with an interpretation of polarity correspondence, other contrasts were not. Specifically, *up* words with upward moving cues in the word-first condition did not significantly differ in accuracy from those with downward moving cues, though polarity correspondence would suggest that the former would yield higher accuracy. This is especially curious in relation to the same comparison in the cue-first condition, where accuracy up words was affected by cue direction. Since these contrasts differ by cue-first and word-first condition, it is possible that the demands of each condition elicit different processing by linguistic and embodied systems, which, which may be outside the scope of the usual cognitive principles that align with polarity effects. This is not to say that polarity alignment plays no role in these tasks, but a polarity correspondence account is, at most, not fully supported by our data. This conceptually replicates the findings of Pecher et al. (2010), who conducted a similar study with some of the factors as between-subjects manipulations. In that study, reaction times did not align well with polarity predictions, or at best, did so inconsistently, and the authors concluded that mental simulation was a likelier explanation of their results than polarity. Our data suggest a similar conclusion, but again, incorrect responses were much more scarce than correct responses in our study, resulting in large variances and rendering contrasts with incorrect responses unreliable when attempting to draw conclusions in this way.

5.2.3. Symbolic and embodied factors - results

We also sought to determine whether symbolic and embodied processing factors influence processing speed differently in different order conditions. Here, the embodied factor is operationalized as the movement ratings for each word obtained in Experiment 1. The symbolic (linguistic) factor is operationalized as the cosine similarity between each word and its respective anchor word (*up* or *down*) using BERT (obtained as described in Experiment 1). The model construction process was similar to previously described models. Reaction time analyses were performed on correct trials. The BERT and norm ratings as described above were entered as fixed effects, as was their interaction with word direction (*up* or *down*) and the 3-way interaction between these two factors and order. This interaction allows us to examine whether any facilitative effects are due to fine grain, within-direction magnitudes as defined by our embodied and linguistic factors, or due to coarse grain, between-direction categorizations (i.e. up/down words), as well as whether the order manipulation differentially demanded the involvement of these factors. The results of the model are presented in Table 10. This analysis revealed significant main effects of order, word direction, and norms. No significant interactions were observed. We then similarly constructed a logistic model to predict accuracy rates. In this model, the main effects of order and word direction, a significant two-way interaction between order and word direction, and a significant three-way interaction between BERT, order, and word direction. The results of this model are displayed in Table 11.

Table 10

Results of linear mixed effects model predicting reaction times

	Fixed Effects		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	3.08	3.05 – 3.11	**<0.001**
Norms	-0.03	-0.05 – -0.00	**0.025**

Word direction (0 = down)	-0.05	-0.06 – -0.03	**<0.001**
Order (0 = cue first)	-0.06	-0.09 – -0.03	**<0.001**
BERT	-0.09	-0.32 – 0.15	0.464
Norms x direction	0.02	-0.01 – 0.06	0.19
Norms x order	-0.01	-0.03 – 0.01	0.422
Word direction x order	-0.01	-0.02 – 0.01	0.266
BERT x word direction	-0.16	-0.48 – 0.17	0.337
BERT x order	0.15	-0.05 – 0.35	0.146
Norms x word direction x order	0	-0.03 – 0.03	0.859
BERT x word direction x order	-0.07	-0.34 – 0.20	0.615
Random Effects			
σ^2			0.02
$\tau_{00\ subject}$			0.02
$\tau_{00\ word}$			0
$\tau_{11\ subject.word\_dirup}$			0
$\tau_{11\ subject.orderwordfirst}$			0.01
$\rho_{01\ subject.word\_dirup}$			-0.45
$\rho_{01\ subject.orderwordfirst}$			-0.32
ICC			0.44
$N_{subject}$			73
N_{word}			32
Observations			8418
Marginal R^2 / Conditional R^2	0.046 / 0.466		

Note. Model formula: logRT ~ norm*worddirection*order + BERT*worddirection*order + (1 + worddirection + order|subject) + (1|word)
τ = estimated variance of random effects
ρ = correlation of random effects

Table 11

Logistic model using linguistic and embodied factors to predict accuracy rates

Fixed Effects			
Predictors	*Std. Odds Ratio*	*CI*	*p*
(Intercept)	14.68	10.25 – 21.04	**<0.001**
Norms	1.32	0.94 – 1.86	0.109
Order (0 = cue first)	0.67	0.54 – 0.82	**<0.001**
Word direction	1.95	1.22 – 3.12	**0.006**
BERT	1.23	0.86 – 1.77	0.247
Norms x order	0.83	0.68 – 1.03	0.086
Norms x word direction	1.02	0.59 – 1.74	0.947

Order x word direction	1.86	1.29 – 2.70	**0.001**
BERT x order	0.94	0.76 – 1.18	0.616
BERT x word direction	0.87	0.51 – 1.48	0.584
Norms x order x word direction	0.98	0.65 – 1.48	0.925
BERT x order x word direction	1.6	1.04 – 2.45	**0.028**
Random Effects			
σ^2			3.29
$\tau_{00\ subject}$			0.45
$\tau_{00\ word}$			0.32
ICC			0.19
$N_{subject}$			73
N_{word}			32
Observations			9115
Marginal R^2 / Conditional R^2	0.099 / 0.269		

Note. Model equation: accuracy ~ norms*worddirection*order + BERT*worddirection*order + (1|subject) + (1|word)
τ = estimated variance of random effects
ρ = correlation of random effects

5.2.4. Symbolic and embodied factors - discussion

The models employing embodied and linguistic factors yielded complex and interesting results. When predicting reaction time, the embodied factor, word direction, and order condition were statistically significant factors, while the linguistic factor and the interactions were not significant. While we predicted significant interactions due to the relative time course of perceptual and linguistic processing during this task, there are several possible explanations for this pattern of results that are consistent with the literature. These results suggest that much like in Experiment 2, between-direction categorization and embodied, within-direction magnitudes were more influential of processing during this task than languages statistics (as operationalized here). A lack of effect for the linguistic factor on reaction time could be because the present task is more demanding of perceptual system involvement, or because verbs favor perceptual processing or are weakly represented by symbolic systems in general (Brown et al, 2023, though the results of Experiment 1 show that BERT maps onto these word senses relatively well).

However, these explanations for lack of a linguistic factor effect are not consistent with the model for accuracy, discussed next.

The related model predicting accuracy yielded a different pattern of results, including evidence for the more direct involvement of language statistics. In this model, we observed a significant 3-way interaction between the linguistic factor, word direction, and order. For upward meaning words, stronger within-direction linguistic associations yielded higher accuracy rates in the word first condition than the cue first condition. This result can be interpreted as consistent with our other findings in that activating stronger linguistic connections while only being able to see the word may create a much narrower hypothesis of potentially congruent movement, demanding less disambiguation from the perceptual system upon viewing the cue, thereby rendering speeded judgments easier.

Why this might be true only of upward meaning words, however, demands further explanation. While these asymmetric results are somewhat common in similar paradigms, with location-cue congruence effects often being driven by up/high words (Dudschig et al., 2013; Lachmair et al., 2011; but not always, see Pecher, 2010), this particular combination of analytic approaches is in novel territory. Future research should seek to redefine and/or disentangle the contributions of linguistic and embodied processing in these situations, especially considering the frequency with which task specificity is raised as a concern for the field.

5.3. Experiment 3 discussion

One of the primary purposes of this experiment was to extend the integrated language processing literature to include verbs in a new stimulus presentation paradigm. By decoupling lexical stimulus presentation from the perceptual cue used to make judgments in a cue-congruency paradigm, we hoped to create an experimental manipulation that could be used within-subjects that might allow us to examine embodied language processing in greater detail using new stimulus sets. This endeavor was successful, but not without limitations and unexpected results.

Regarding expected effects (section 5.2.1), We found that words with upward meanings paired with upward-moving cues facilitate reaction time and accuracy, consistent with previous location-cue studies. Since movement-word congruency aided processing in both conditions, these results suggest a degree of generalizability in using dynamic stimulus presentations to examine embodied and symbolic language systems. Further, there was a significant effect of the novel order condition; participants were faster in the word-first condition. In this condition, reading the word first triggers linguistic processes which quickly create a narrow judgment hypothesis, which shortcuts need for slow, embodied language processing, thereby reducing reaction time. In contrast, participants were more accurate for upward words and upward movement in the cue-first condition, suggesting embodied processing before linguistic processing. These effects generally support language shortcut hypotheses, but warrant further investigation, especially to examine whether the order manipulation is truly eliciting differences in symbolic and embodied language processing as expected.

We also examined whether polarity alignment could explain our effects (section 5.2.3). As in Experiment 2, we found that our results generally did not align with the predictions made by polarity correspondence research. However, this does not completely rule out polarity as a factor that influences reaction times. Rather, it suggests that perceptual simulation is a more likely explanation for more of the variance observed in our data (as in Pecher et al., 2010). We recommend that researchers in the future seek to disentangle the possibly plural role of polarity and mental simulation more directly in similar experimentation, for example, by incorporating more difficult tasks or reconsidering the response dimension of polarity.

The analyses using operationalized symbolic and embodied factors to predict reaction time and accuracy (section 5.2.5) yielded complex results. We found that the embodied factor (human ratings) was significantly associated with reaction times, while the symbolic factor (language statistics) was not. Unexpectedly, this was true across both order conditions. There are many possible explanations for this finding. It could be that verbs elicit more perceptual processing than other parts of speech, causing the embodied factor to predominantly influence reaction times. Alternatively, the judgments being asked of participants might be more perceptually demanding than other lexical judgments often found in related studies. Along these lines, distributional semantic models have been shown to account more weakly for verbs than other parts of speech (Brown, 2023), so the linguistic factor may have been disadvantaged from the start. However, the linguistic factor did yield a significant interaction three-way interaction with the order manipulation and word direction when predicting accuracy. More specifically, stronger within-direction linguistic associations led to higher accuracy rates in the word-first condition for words with upward meanings. This outcome aligns with our previous findings, suggesting that when stronger linguistic connections are activated while only the word is visible, participants form narrower hypotheses about potentially congruent movement. This reduces the need for disambiguation by the perceptual system when the cue is observed, making quicker judgments more possible. However, this effect may apply only to upward meaning words, which is not readily explained by mental simulation alone. As previously discussed, future research may need to investigate polarity alignment as an explanation in conjunction with cognitive simulation, rather than as a replacement for it.

This study is constrained by several limitations that may be addressed with continued research. The time course of the sequencing manipulation may not be optimal for the observation of these types of effects. Participants had roughly 500ms between stimuli, which may overestimate the time course involved in the interplay between symbolic and embodied systems. Future research should investigate similar manipulations over multiple time courses to determine how they impact shifts in processing between linguistic and embodied systems under the sequencing manipulation in this task. Other psycholinguistics explanations of our data may involve the relative contextual demands of the words in each list; it could be that certain words have more plausible axes or degrees of motion than others. This may lead to asymmetric results if the differences are not evenly distributed between *up* and *down* lists. Future research using similar measures should consider the motoric associations of words in stimulus lists carefully, because this type of sensorimotor information may play an important role in similar experiments and tasks.

6. General Discussion

This study was motivated by the lack of verbs in symbolic-embodied integrated language processing research. Accordingly, a primary objective of these experiments was to determine an empirically valid way of integrating verbs into this literature. The results of these studies suggest that verbs are viable for examination in several ways.

In Experiment 1, a judgment task of the vertical movement meanings of individual verbs yielded norms that were both related to and importantly distinct from related language statistics. This selection of verbs was small and designed to vary along one dimension (the vertical movement axis), but future research could explore a great number of verbs in higher scale norming studies involving the unique motion properties of verbs. The results of this experiment suggest that such norms might capture perceptually grounded information in semantic memory that is not fully represented by current computational linguistic measures.

In Experiment 2, these verbs were implemented in a location-cue congruency paradigm common in embodied and integrative language processing research. Verbs presented in positions that were congruent with the endpoints of their implied directions were processed more quickly than those presented in incongruent positions – this effect was driven by upward meaning words presented high. At a basic level, this finding indicates a successful verb-based replication of embodied language processing experiments that have previously neglected verbs. Further, the norms obtained in Experiment 1 were significant predictors of response speeds in this experiment, while language statistics were not. This somewhat contrasts with previous research in this domain that demonstrates that language statistics are often a significant predictor of reaction times for these kinds of tasks (Hutchinson & Louwerse, 2013; Tillman & Louwerse, 2018). However, distributional semantic models have been demonstrated to construct weaker representations for verbs in comparison to other parts of speech (Brown et al., 2023, though contextual embedding models like BERT should mitigate this concern somewhat). Further, it is likely that our task was more demanding of deep, embodied information than something like a surface-level lexical judgment (e.g., word or non-word), and the use of motion/directionality norms for verbs may also be a stronger embodied factor than the embodied factors operationalized in these noun-based studies. Future research should seek to collect and examine the properties of perceptually oriented norms, and further explore how task demands create demands for embodied and linguistic processing at a fine grain level, including at the level of perception (Lupyan et al., 2020).

In Experiment 3, these verbs were investigated using a novel manipulation designed as an extension of previous location-cue congruency and semantic/iconic judgment paradigms. This manipulation involved animated stimuli and the temporal decoupling of word and meaning presentations. The manipulation used in this study was found to affect reaction times in a manner consistent with linguistic shortcut hypotheses because faster reaction times were observed when words were presented before physical movement. Consistent with Experiment 2, norms were again associated with faster response times across conditions while linguistic factor yielded no significant effect. However, the linguistic factor was involved in a significant 3-way interaction predicting response accuracy, such that stronger language statistics elicited more accurate responses for *up* words in the word-first condition, while the embodied factor did not yield a significant main effect. This complex set of results may be cautiously interpret as generally supportive of embodied language processing theories and linguistic shortcut hypotheses, but much more work must corroborate, disentangle, or otherwise explain some inconsistent effects, such as the asymmetrical effects between upward and downward words, or the difference in patterns of results when predicting speed versus accuracy. Future research could explore additional time course manipulations or utilize ERP or eye-tracking methods to reveal more fine-grain information about the scale at which embodied and symbolic systems activate and interrelate.

In addition to their individual results, this series of experiments may provide insight to embodiment-integrative language processing research when considered in context. These findings provide evidence that, for verbs, language processing is indeed both symbolic and embodied, and that language statistics will obviate the need for embodied processing in certain contexts. Further, findings related to accuracy suggest that embodied processing done in advance of linguistic processing may reinforce or enhance judgments made using linguistic information in some contexts. However, our knowledge about the nature and constraints of these interdependent systems is far from exhaustive. At a surface level, this study provides an example of how verbs may be adapted in the study of symbolic and embodied language processing. At a deeper level, verbs have been used in a series of incremental, novel tasks to reveal further nuance about the interdependent nature of linguistic and embodied information.

7. Acknowledgements

This material is based upon work supported by the National Science Foundation under Grants 1918751 and 1934745 by the Institute of Education Sciences under Grant R305A190448. Thanks to Roger Kreuz and John Sabatini for their helpful comments on this work.

References

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling Embodied and Distributional Accounts of Meaning in Language. Topics in Cognitive Science, 6(3), 359–370. https://doi.org/10.1111/tops.12096

Aravena, P., Hurtado, E., Riveros, R., Cardona, J. F., Manes, F., & Ibáñez, A. (2010). Applauding with Closed Hands: Neural Signature of Action-Sentence Compatibility Effects. *PLoS ONE*, *5*(7), e11751. https://doi.org/10.1371/journal.pone.0011751

Baayen, R. H. (2001). *Word frequency distributions*. Kluwer. https://books.google.com/books?hl=en&lr=&id=xUSM69ZkjHoC&oi=fnd&pg=PR9&ots=Y11egdDm76&sig=XVdJXeKIM0ud94o0PSJaNyE5Wys#v=onepage&q&f=false

Banks, B., Wingfield, C., & Connell, L. (2021). Linguistic Distributional Knowledge and Sensorimotor Grounding both Contribute to Semantic Category Production. *Cognitive Science*, *45*(10). https://doi.org/10.1111/cogs.13055

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barsalou, L., Santos, A., Simmons, W., & Wilson, C. (2008). Language and simulation in conceptual processing. In *Symbols, embodiment, and meaning* (pp. 245–283). http://barsaloulab.org/Online_Articles/2008-Barsalou_et_al-chap-language_situated_simulation.pdf

Barsalou, L. W. (2010). Grounded Cognition: Past, Present, and Future. *Topics in Cognitive Science*, *2*(4), 716–724. https://doi.org/10.1111/j.1756-8765.2010.01115.x

Barsalou, L. W. (2016). On Staying Grounded and Avoiding Quixotic Dead Ends. In *Psychonomic Bulletin and Review* (Vol. 23, Issue 4, pp. 1122–1142). Springer New York LLC. https://doi.org/10.3758/s13423-016-1028-3

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2012). Language and simulation in conceptual processing. In *Symbols and Embodiment: Debates on Meaning and Cognition*. https://doi.org/10.1093/acprof:oso/9780199217274.003.0013

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex*, *19*(8), 1905–1914. https://doi.org/10.1093/cercor/bhn217

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL*, 31–40. https://books.google.com/books?hl=en&lr=&id=Ksv5DwAAQBAJ&oi=fnd&pg=PA31&ots=Ti0XMUL6jc&sig=-Cc3TdfyeQy6BGD_p9LN6SWNCjc

Brants, T., & Franz, A. (2006). Web 1t 5-gram. *Linguistic Data Consortium*.

Brown, K. S., Yee, E., Joergensen, G., Troyer, M., Saltzman, E., Rueckl, J., Magnuson, J. S., & McRae, K. (2023). Investigating the Extent to which Distributional Semantic Models Capture a Broad Range of Semantic Relations. *Cognitive Science*, *47*(5), e13291. https://doi.org/10.1111/COGS.13291

Bruni, E., Tran, N. K., & Baroni, M. (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, *49*, 1–47. https://doi.org/10.1613/jair.4135

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Burgess, C., Livesay, K., & Lund, K. (1998). Explorations in context space: Words, sentences, discourse. *Discourse Processes*, *25*(2–3), 211–257. https://doi.org/10.1080/01638539809545027

Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22–29. https://aclanthology.org/J90-1003.pdf

Connell, L. (2019). What Have Labels Ever Done For Us? The Linguistic Shortcut in Conceptual Processing. *Language, Cognition, and Neuroscience*, *34*(10), 1308–1318. https://doi.org/10.1080/23273798.2018.1471512

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

De Vega, M., Glenberg, A., & Graesser, A. (2008). Symbols and embodiment: Debates on meaning and cognition. In *Symbols and Embodiment: Debates on Meaning and Cognition*. https://doi.org/10.1093/acprof:oso/9780199217274.001.0001

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*, 4171–4186.

Dolscheid, S., & Casasanto, D. (2015). Spatial congruity effects reveal metaphorical thinking, not polarity correspondence. *Frontiers in psychology*, 6, 1836.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, *110*(3), 412–431. https://doi.org/10.1016/j.cognition.2008.11.016

Dudschig, C., & Kaup, B. (2017). Is it all task-specific? The role of binary responses, verbal mediation, and saliency for eliciting language-space associations. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(2), 259–270. https://doi.org/10.1037/xlm0000297

Dudschig, C., Souman, J., Lachmair, M., Vega, I. de la, & Kaup, B. (2013). Reading “Sun” and Looking Up: The Influence of Language on Saccadic Eye Movements in the Vertical Dimension. *PLoS ONE*, *8*(2), e56872. https://doi.org/10.1371/journal.pone.0056872

Estes, Z., & Barsalou, L. W. (2018). A Comprehensive Meta-Analysis of Spatial Interference From Linguistic Cues: Beyond Petrova et al. (2018). *Psychological Science*, *29*(9), 1558–1564. https://doi.org/10.1177/0956797618794131

Estes, Z., Verges, M., & Adelman, J. S. (2015). Words, objects, and locations: Perceptual matching explains spatial interference and facilitation. *Journal of Memory and Language*, *84*, 167–189. https://doi.org/10.1016/j.jml.2015.06.002

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, *25*(2–3), 285–307. https://doi.org/10.1080/01638539809545029

Graesser, A. C., Cai, Z., Baer, W. O., Olney, A. M., Hu, X., Reed, M., & Greenberg, D. (2016). Reading comprehension lessons in AutoTutor for the Center for the Study of Adult Literacy. In *Adaptive educational technologies for literacy instruction* (pp. 1-288–293). https://doi.org/10.4324/9781315647500

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing Multilevel Analyses of Text Characteristics. *Educational Researcher*, *40*(5), 223–234. https://journals.sagepub.com/doi/pdf/10.3102/0013189X11413260

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244. https://doi.org/10.1037/0033-295X.114.2.211

Guan, C. Q., Meng, W., Yao, R., & Glenberg, A. M. (2013). The Motor System Contributes to Comprehension of Abstract Language. *PLoS ONE*, *8*(9), e75183. https://doi.org/10.1371/journal.pone.0075183

Günther, F., Petilli, M. A., Vergallito, A., & Marelli, M. (2022). Images of the unseen: extrapolating visual representations for abstract and concrete words in a data-driven computational model. *Psychological Research*, *86*(8), 2512–2532. https://doi.org/10.1007/S00426-020-01429-7

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-Space Models of Semantic Representation From a Cognitive Perspective: A Discussion of Common Misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033. https://doi.org/10.1177/1745691619861372

Hollis, G. (2017). Estimating the average need of semantic knowledge from distributional semantic models. *Memory and Cognition*, *45*(8), 1350–1370. https://doi.org/10.3758/s13421-017-0732-1

Hutchinson, S., & Louwerse, M. M. (2013). What ' s Up can be Explained by Language Statistics. *Escholarship.Org*, 2596–2601. https://escholarship.org/content/qt4h97698p/qt4h97698p.pdf

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37. https://doi.org/10.1037/0033-295X.114.1.1

Jones, M. N., Willits, J., & Dennis, S. (2015). Models of Semantic Memory. *Oxford Handbook of Mathematical and Computational Psychology*, 232–254. https://books.google.com/books?hl=en&lr=&id=_dbFBgAAQBAJ&oi=fnd&pg=PA232&ots=-mteE5QEUX&sig=d_--DsM9LGE0CIGGTOYb2z6kStw

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lachmair, M., Dudschig, C., De Filippis, M., de la Vega, I., & Kaup, B. (2011). Root versus roof: Automatic activation of location information during word processing. *Psychonomic Bulletin and Review*, *18*(6), 1180–1188. https://doi.org/10.3758/s13423-011-0158-x

Lakens, D. (2011). High skies and oceans deep: Polarity benefits or mental simulation? In *Frontiers in Psychology* (Vol. 2, Issue FEB). https://doi.org/10.3389/fpsyg.2011.00021

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(3), 726–736. https://doi.org/10.1037/A0024955

Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Landauer, T., McNamara, D., Dennis, S., & Kintsch, W. (2007). Handbook of Latent Semantic Analysis. In *Handbook of Latent Semantic Analysis*. https://doi.org/10.4324/9780203936399

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, *20*(1), 1–31. https://www.researchgate.net/publication/237672963

Louwerse, M. (2011). Stormy seas and cloudy skies: Conceptual processing is (still) linguistic and perceptual. In *Frontiers in Psychology* (Vol. 2, Issue MAY). https://doi.org/10.3389/fpsyg.2011.00105

Louwerse, M., & Hutchinson, S. (2012). Neurological Evidence Linguistic Processes Precede Perceptual Simulation in Conceptual Processing. *Frontiers in Psychology*, *3*, 385. https://doi.org/10.3389/fpsyg.2012.00385

Louwerse, M., & Jeuniaux, P. (2008). Language comprehension is both embodied and symbolic. In *Symbols and embodiment: Debates on meaning and cognition* (pp. 309–326). https://www.academia.edu/download/30813794/LouwerseJeuniauxBookChapter08.pdf

Louwerse, M., & Jeuniaux, P. (2012a). Language comprehension is both embodied and symbolic. In *Symbols and Embodiment: Debates on Meaning and Cognition*. https://doi.org/10.1093/acprof:oso/9780199217274.003.0015

Louwerse, M., & Jeuniaux, P. (2012b). Language comprehension is both embodied and symbolic. In *Symbols and Embodiment: Debates on Meaning and Cognition*. https://doi.org/10.1093/acprof:oso/9780199217274.003.0015

Louwerse, M. M. (2007). Symbolic or embodied representations: A case for symbol interdependency. *Handbook of Latent Semantic Analysis.*, *November*, 107–120. https://www.researchgate.net/publication/253609370

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*(2), 273–302. https://doi.org/10.1111/j.1756-8765.2010.01106.x

Louwerse, M. M. (2018). Knowing the Meaning of a Word by the Linguistic and Perceptual Company It Keeps. *Topics in Cognitive Science*, *10*(3), 573–589. https://doi.org/10.1111/tops.12349

Louwerse, M. M., & Connell, L. (2011). A Taste of Words: Linguistic Context and Perceptual Simulation Predict the Modality of Words. *Cognitive Science*, *35*(2), 381–398. https://doi.org/10.1111/j.1551-6709.2010.01157.x

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*(1), 96–104. https://doi.org/10.1016/J.COGNITION.2009.09.002

Louwerse, M. M., McCarthy, P. M., McNamara, D. S., & Graesser, a C. (2004). Variation in language and cohesion across written and spoken registers. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, *1988*, 843–848. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.123.1183&rep=rep1&type=pdf

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Lupyan, G., Rahman, R., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in Cognitive Sciences*, *24*(11), 930–944. https://www.sciencedirect.com/science/article/pii/S1364661320302138

Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic? In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 373, Issue 1752). Royal Society Publishing. https://doi.org/10.1098/rstb.2017.0137

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://www.sciencedirect.com/science/article/pii/S0749596X16300079?

casa_token=W0spy2cPrg8AAAAA:ZAKAp9lvDnMhT6w4C-SqsVCIb6VnnV5Az0_xGqflGZ_EkdhEH1acgVkejElRXv2WXmu2cdG1X-k

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1), 1–23. https://doi.org/10.5334/joc.18

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory and Cognition*, *32*(8), 1389–1400. https://doi.org/10.3758/BF03206329

Mcnamara, D., Cai, Z., & Louwerse, M. M. (2007). *Optimizing LSA Measures of Cohesion Complementary systems for conceptual processing: Modality norms and ERPs*.

Meier, B. P., & Robinson, M. D. (2004). Why the Sunny Side Is Up: Associations Between Affect and Vertical Position. *Psychological Science*, *15*(4), 243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*. https://doi.org/10.1016/j.jml.2020.104092

Miklashevsky, A. (2018). Perceptual Experience Norms for 506 Russian Nouns: Modality Rating, Spatial Localization, Manipulability, Imageability and Other Variables. *Journal of Psycholinguistic Research*, *47*(3), 641–661. https://doi.org/10.1007/s10936-017-9548-1

Paivio, A. (1971). Imagery and Verbal Processes. In *Imagery and Verbal Processes* (1st ed.). Holt, Rinehart and Winston. https://doi.org/10.4324/9781315798868

Paperno, D., Marelli, M., Tentori, K., & Baroni, M. (2014). Corpus-based estimates of word association predict biases in judgment of word co-occurrence likelihood. *Cognitive Psychology*, *74*, 66–83. https://www.sciencedirect.com/science/article/pii/S0010028514000541?casa_token=zBsllwGOrx4AAAAA:tXP4SL8XsRb4zfb7a7ZdswsEsEbcQXPm6X4a6wlVm7YjG-OLPl07MrqZ5pZ12tc4MtW9ZhDBBlg

Pecher, D., Van Dantzig, S., Boot, I., Zanolie, K., & Huber, D. E. (2010). Congruency between word position and meaning is caused by task-induced spatial attention. *Frontiers in Psychology*, *SEP*. https://doi.org/10.3389/fpsyg.2010.00030

Perlman, M., Little, H., Thompson, B., & Thompson, R. L. (2018). Iconicity in signed and spoken vocabulary: A comparison between American Sign Language, British Sign Language, English, and Spanish. *Frontiers in Psychology*, *9*(AUG), 1433. https://doi.org/10.3389/fpsyg.2018.01433

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. In *Frontiers in Psychology* (Vol. 1, Issue DEC, p. 227). Frontiers. https://doi.org/10.3389/fpsyg.2010.00227

Petilli, M., Günther, F., Vergallito, A., Ciapparelli, M., & Marelli, M. (2021). Data-driven computational models reveal perceptual simulation in word processing. *Journal of Memory and Language*, *117*. https://www.sciencedirect.com/science/article/pii/S0749596X2030108X

Petrova, A., Navarrete, E., Suitner, C., Sulpizio, S., Reynolds, M., Job, R., & Peressotti, F. (2018). Spatial Congruency Effects Exist, Just Not for Words: Looking Into Estes, Verges, and Barsalou (2008). In *Psychological Science* (Vol. 29, Issue 7, pp. 1195–1199). SAGE Publications Inc. https://doi.org/10.1177/0956797617728127

Proctor, R., & Cho, Y. (2006). Polarity Correspondence: A General Principle for Performance of Speeded Binary Classification Tasks. *Psychologican Bulletin*, *132*(3), 416–442. https://doi.org/10.1037/0033-2909.132.3.416

Proctor, R. W., & Xiong, A. (2015). Polarity Correspondence as a General Compatibility Principle. *Current Directions in Psychological Science*, *24*(6), 446–451. https://doi.org/10.1177/0963721415607305

Pu, H., Medina, Y. E., Holcomb, P. J., & Midgley, K. J. (2019). Testing for nonselective bilingual lexical access using L1 attrited bilinguals. *Brain Sciences*, *9*(6). https://doi.org/10.3390/brainsci9060126

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, *17*(6), 884–892. https://doi.org/10.1162/0898929054021111

Recchia, G., & Jones, M. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*, *41*(3), 647–656. https://doi.org/10.3758/BRM.41.3.647

Richardson, D., & Matlock, T. (2007). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition*, *102*(1), 129–138. https://doi.org/10.1016/j.cognition.2005.12.004

Richter, M., Lins, J., & Schöner, G. (2021). A Neural Dynamic Model of the Perceptual Grounding of Spatial and Movement Relations. *Cognitive Science*, *45*(10), 13045. https://doi.org/10.1111/cogs.13045

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. http://arxiv.org/abs/1910.01108

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

Scicluna, S., & Strapparava, C. (2020). *VROAV: Using Iconicity to Visually Represent Abstract Verbs*. http://vroav.online,

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

Shapiro, L. (2014). The routledge handbook of embodied cognition. In *The Routledge Handbook of Embodied Cognition*. https://doi.org/10.4324/9781315775845

Siakaluk, P. D., Buchanan, L., & Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Memory and Cognition*, *31*(1), 100–113. https://doi.org/10.3758/BF03196086

Team R Development Core. (2021). A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, https://www.R-project.org. https://ci.nii.ac.jp/naid/20001689445/

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. https://doi.org/10.1162/0898929053124965

Tillman, R., & Louwerse, M. (2018). Estimating Emotions Through Language Statistics and Embodied Cognition. *Journal of Psycholinguistic Research*, *47*(1), 159–167. https://doi.org/10.1007/s10936-017-9522-y

Utsumi, A. (2020). Exploring What Is Encoded in Distributional Word Vectors: A Neurobiologically Motivated Analysis. *Cognitive Science*, *44*, 12844. https://doi.org/10.1111/cogs.12844

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247. https://doi.org/10.1515/LANGCOG.2009.011

Willems, R. M., & Francken, J. C. (2012). Embodied cognition: Taking the next step. *Frontiers in Psychology*, *3*(DEC). https://doi.org/10.3389/fpsyg.2012.00582

Yee, E., Jones, M. N., & McRae, K. (2018). Semantic Memory. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–38). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119170174.epcn309

Zeno, S., Ivens, S., Millard, R., & Duvvuri, R. (1995). *The educator's word frequency guide*.

Zhang, Y., Hu, J., Zhang, E., & Wang, G. (2015). The influence of spatial representation on valence judgements: An event-related potential study. *Journal of Cognitive Psychology*, *27*(2), 218–226. https://doi.org/10.1080/20445911.2014.995667

Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. In *Trends in Cognitive Sciences* (Vol. 18, Issue 5, pp. 229–234). Elsevier Ltd. https://doi.org/10.1016/j.tics.2014.02.008

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking* (pp. 224–245). https://doi.org/10.1017/CBO9780511499968.010

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430

Zwaan, R. A., van der Stoep, N., Guadalupe, T., & Bouwmeester, S. (2012). Language comprehension in the balance: The Robustness of the action-compatibility effect (ACE). *PLoS ONE*, *7*(2), e31204. https://doi.org/10.1371/journal.pone.0031204

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin and Review*, *10*(4), 954–958. https://doi.org/10.3758/BF03196557