Casasanto, D. (2022). Embodied Semantics. In F.T. Li (Ed.) *Handbook of Cognitive Semantics.* Leiden: Brill.

CHAPTER 6

Embodied Semantics

Daniel Casasanto

Abstract: What does it mean for meaning to be *embodied*? According to a revolutionary hypothesis, part of a word's meaning is a *simulation* of its referent, implemented in neural and cognitive systems that support perception, action, and emotion. The embodied simulation hypothesis calls for a radical departure from long-held views of minds and brains. This chapter details how embodied simulation can – and cannot – be tested, and reviews evidence that meaning is embodied in 'modality-specific' simulations. Finally, some remaining challenges for the embodied simulation hypothesis are discussed.
Keywords: embodiment, simulation, psychology

1. What is embodied meaning?

What does it mean for meaning to be *embodied?* The term 'embodiment' is used in many ways, and by multiple communities of researchers (Wilson, 2002). In perception research, 'embodiment' can mean the sense of where one's own body ends and the rest of the world begins (Longo et al., 2008). In artificial intelligence, 'embodiment' is the project of endowing robots with biologically inspired mechanisms of learning and behavior (Anderson, 2003; Brooks, 1990). More broadly, 'embodiment' can mean the belief that the mind is shaped by our bodily experiences (Varela et al., 1991), or the observation that bodies and minds affect each other (Niedenthal, 2007). This chapter focuses on a notion of embodiment that is of particular relevance to theories of linguistic meaning, which has revolutionary implications for research on minds and brains: an idea that will be referred to here as the *embodied simulation hypothesis*.

According to the embodied simulation hypothesis, part of the meaning of a word (or a phrase, or a sentence) is a *simulation* of its referent, implemented in neural and cognitive systems that support perception, action, and emotion. Three variants of this hypothesis were proposed in parallel in 1999, by the psychologist Lawrence Barsalou (Barsalou, 1999), the neuroscientist Friedemann Pulvermüller (Pulvermüller, 1999)[1], and by the philosopher Mark Johnson and the linguist George Lakoff (Lakoff & Johnson, 1999). These proposals were not identical, but they overlapped to a remarkable extent. Together, these three publications have been cited more than 30,000 times, and have generated hundreds of theoretical papers and thousands of experimental studies. This chapter aims: (i.) to explain the embodied simulation hypothesis; (ii.) to review ways in which it has (and has not) been tested effectively; (iii.) to outline some remaining challenges for the embodied simulation hypothesis.

2. What is a simulation?

A simulation is a pattern of neural and cognitive activity that corresponds to our experience of an entity in the world outside of your mind (e.g., seeing a cat), but crucially, simulations occur when you are *not* currently experiencing that entity (i.e., when you are not seeing a cat, but only thinking about a cat or understanding the word *cat*). How is a simulation different from other notions of thinking or understanding? Simulations are posited to occur in neurocognitive (i.e.,

[1] Pulvermüller (1999) did not use the term *simulation*, but his proposal is nevertheless a seminal version of the embodied simulation hypothesis described here. The phrase "embodied simulation hypothesis" is intended as a generic term that encompasses key elements of the hypotheses framed by Barsalou (1999), Pulvermüller (1999), and Lakoff and Johnson (1999).

neural / cognitive) systems that are responsible for our primary experience of the external world: most notably, systems for perception and motor action, which were long believed to be separate from systems for thinking and constructing linguistic meaning (Fodor, 1983).

In order to understand the construct of *simulation*, it is first necessary to understand how information flows from the world into the mind, and what kinds of neurocognitive systems are responsible for this flow of information. People interact with their environment via multiple input and output *modalities*. A modality is a channel through which we experience or act upon the world. When we interact with a cat, we can *see* its shape, *hear* its purr, *reach out* our hand to *touch* its fur, and *feel* the happiness of communing with a pet (assuming we like cats). Each of these perceptual, motoric, and emotional components of our 'cat' experience is implemented initially in a different *modality-specific* neurocognitive system. A system is modality-specific if it is highly specialized for a single input or output modality (e.g., vision). Modality-specific regions of the brain that have spatially constrained locations, are largely segregated from each other, and are highly selective for processing information in one modality or another: *visual cortex* for sight, *auditory cortex* for sound, *somatosensory cortex* for touch, *olfactory cortex* for smell, *gustatory* cortex for taste, *motor cortex* for performing actions, and structures in the limbic system for forming emotional responses to stimuli.[2]

As the 'cat' example illustrates, our primary experiences often involve more than one kind of modality-specific input. Initially, this input is segregated: The cat's visual appearance is processed in visual cortex; the sounds that the cat makes are processed in auditory cortex, etc. Information from these separate modality-specific regions is then integrated in multimodal regions of the cerebral cortex called *convergence zones* (Barsalou et al. 2003; Damasio 1989). Convergence zones are not modality-specific; rather, they receive inputs from various modality-specific systems. Each episode of interacting with a cat results in patterns of modality-specific neurocognitive activity during the process of perception (e.g., seeing the cat) or action (e.g., petting the cat). Then, with the help of non-modality-specific brain structures (which will not be discussed extensively here, e.g., the hippocampus), this transient modality-specific activity leads to the formation of longer-lasting memory traces, that are stored in non-modality-specific convergence zones. These long-term memory traces are built out of modality-specific information that accumulates over the course of an individual's experiences, but they are stored in non-modality-specific brain areas.[3]

---

[2] Primary sensory and motor cortices are essential for perception or action in a given modality (e.g., *seeing* is not possible without activity in visual cortex). However, it would be a mistake to believe that these modality-specific cortices are *sufficient* to support perception or action; each primary cortex is only one critical part of a distributed network of brain areas needed, for example, to transform sensory input into conscious perceptual experience. Here, we focus only on the primary and secondary sensory and motor cortices (e.g., motor cortex and premotor cortex) because these are the only parts of their distributed networks that are modality-specific; therefore, they are the only parts of the network that are useful for distinguishing between embodied and disembodied theories experimentally.

[3] An alternative to the term "modality-specific" is "unimodal"; there is no clear distinction between these two terms so, for most purposes, they should be treated as synonyms that both designate brain tissue that is highly specialized for processing information in a single perceptual or motor modality. The term "non-modality-specific" has several near synonyms: amodal (not associated with any modality); multimodal (processing inputs from multiple modality-specific regions); polymodal (presumably an exact synonym for multimodal); supramodal (somehow transcending modality (i.e., amodal), or incorporating different modalities (i.e., multimodal)). Some authors may intend for there to be subtle distinctions between these near synonyms; however, these distinctions are rarely made clear, and are not widely agreed upon across authors. Brain regions are sometimes designated as "bimodal" or "trimodal" (i.e., integrating input from exactly two or exactly three modalities): These regions can be considered to be a subset of non-modality-specific regions.

This sketch of how information flows from the outside world into our minds (via modality-specific systems) and produces long-term memory traces (stored in non-modality-specific brain areas) should be largely uncontroversial; this account is shared by both embodied and 'disembodied' theories, alike. These theories diverge, however, concerning how our long-term memory traces get used during thinking and linguistic meaning construction.

According to 'disembodied' 20th-century theories of concepts and semantics, not only are words' meanings *stored in* non-modality-specific brain areas, they are also *retrieved from* these same areas when we need to use our stored knowledge. One area long believed to be the locus of our 'mental dictionary' is the left Temporal cortex (Hagoort, 2005): a non-modality-specific brain area that integrates information from various modality-specific cortices. On this view, visual cortex, auditory cortex, and motor cortex *do* play a role in language, but that role is limited to processing the *forms* of words. It is uncontroversial that auditory cortex is crucial for perceiving the auditory forms of words that you hear, that visual cortex is crucial for perceiving the orthographic forms of words that you read, and that motor cortex is crucial for producing the articulatory forms of words that you speak with your mouth or sign with your hands; both disembodied and embodied theories agree upon these facts about perceiving and producing the forms of words. The theories disagree, however, about the role that modality-specific brain areas play in constructing a word's meaning. According to the 'disembodied' theories of language, these modality-specific perceptual and motor cortices are crucial for processing the forms of words, but they play no role in processing their meanings. This belief was virtually unquestioned until the turn of the 21st century.

By contrast, according to the embodied simulation hypothesis, modality-specific brain areas play important roles in processing *both* the forms of words and their meanings. The 'cat' example, above, illustrates how information about cats enters our minds, passing from various modality-specific brain areas during perception to non-modality-specific brain areas for long-term storage. Simulation involves, essentially, running this process in reverse.

When we hear the word *cat*, this acoustic input from the ear is first processed in the auditory cortex; this auditory signal is then classified as an instance of the English wordform /*kaet*/, and this wordform classification process involves left Temporal lobe structures near the auditory cortex: specifically, structures that have been traditionally associated with the 'mental dictionary' (Pulvermüller, 1999). Both embodied and disembodied theories may agree on this wordform classification process, but they diverge concerning the extent of the role that these left Temporal lobe structures play in language and thought. According to the traditional 'disembodied' view, the Temporal lobe contains complete lexical entries (i.e., entries in the mental dictionary), including words' forms, morphosyntactic roles, and – importantly – their meanings.[4] According to the embodied view, however, left temporal cortex is the locus of *some* kinds of information about words, most crucially their *forms*, but is not the locus of words' (complete) meanings. Rather, once a wordform has been identified, it cues modality-specific simulations in the relevant perceptual or motor cortices, which constitute the word's meaning. (On moderate versions of this hypothesis, simulations *partly* constitute the word's meaning.)

Once the wordform /*kaet*/, has been identified, various aspects of the context determine which particular simulations will (or will not) be run: The relevant 'context' may include the linguistic context, social context, the physical context in which the word is perceived, as well as

[4] Adherents to this long-held majority view of language in the brain acknowledge that semantic information is likely to be distributed over various non-modality-specific brain areas, but maintain that the left temporal lobe is particularly "crucial for lexical-semantic processing" (Hagoort, 2005, pg. 421).

the language user's own history of using this wordform previously (Willems & Casasanto, 2011). In the linguistic context, "My pet cat is a white Angora," naming a breed known for having exceptionally soft fur, the simulations that get triggered may involve somatosensory areas that perceive touch with the fingers, and motor areas that allow us to plan and execute the action of petting a cat with our hands, as well as low-level visual areas that allow us to perceive the animal's shape and higher-level visual areas that are involved in perceiving color. If, instead, the linguistic context were, "The black cat dashed across the road," then in addition to simulations in visual areas for perceiving shape and color, simulations may also be triggered in high-level visual areas involved in perceiving motion and speed (Wallentin et al., 2011).[5] Alternatively, if the context were a jazz musician commenting that, "Our drummer is a cool cat," it remains an open question whether any simulations related to a small furry animal would be triggered, at all; instead, a different set of simulations would likely be triggered, since the referent of this figurative expression is a drum-playing human.

Typically, simulations are *implicit*, meaning that the simulations, per se, are not available to consciousness (even though language users are usually conscious of hearing and understanding the words whose referents are being simulated). Listeners would not actually perceive the attributes of a cat, even when hearing this wordform causes them to simulate these attributes in perceptual cortices: Experiencing a percept in the absence of any perceptual input is a hallucination. Likewise, even though listeners' understanding of *cat* may include motor simulations of what their hands do when they pet a cat, they are unlikely to actually perform these hand actions. It is unclear, at this time, *why* we don't perform the actions that we simulate (e.g., whether the actions are never fully programmed by the motor cortex, or whether they are programmed but inhibited). What is clear, however, is that even when listeners (or readers) understand a sentence like, "I pet the cat," they do not typically start involuntarily petting a cat (or petting the air, if no cat is present). Furthermore, in most instances of rapidly understanding ongoing speech, listeners do not form a conscious mental image of the cat (see Willems et al., 2010 for a discussion of mental simulation vs. mental imagery).

3. Testing the embodied simulation hypothesis

In order to design an experimental test of the embodied simulation hypothesis, two fundamental criteria must be met. First, it must be clear what *competing hypothesis* embodied simulation is being tested against: If the experimental results turned out to confirm the embodied simulation hypothesis, what alternative hypothesis (or null hypothesis) would be disconfirmed?[6]

---

[5] As these 'cat' examples illustrate, simulations are posited to occur in all of the contextually-relevant modality-specific cortices, in parallel. This process is sometimes referred to as a "multimodal simulation." This label, however, is potentially confusing: Most likely, authors using the term "multimodal simulation" *do not* mean that the simulation is being implemented in multimodal (i.e., non-modality-specific) brain areas. Rather, this term is used to indicate that multiple, distinct, modality-specific simulations are being run in parallel (e.g., in visual cortex, auditory cortex, etc.)

[6] The 'disembodied' theory of semantics described in Section 1 can be considered to be an *alternative hypothesis* against which embodied simulation is being tested in many studies. However, the disembodied theory can also be considered to be the *null hypothesis* in these studies: that which the relevant community of researchers would continue to believe if the experiment in question had never been done, or if it produced no interpretable results. Often, when testing a new hypothesis (e.g., embodied semantics) against an established hypothesis (e.g., disembodied semantics), the alternative hypothesis is coextensive with the null hypothesis. Here, what the relevant community of researchers would continue to believe in the absence of evidence that semantics is embodied would be: that semantics is disembodied (i.e., linguistic meaning relies on neurocognitive systems that are distinct from the systems that support our direct perceptual, motoric, and affective interactions with the environment).

Second, the competing theories must make *contrasting predictions* about the behavior or brain activity in the experimental participants. Designing experiments that satisfy both of these criteria – where competing hypotheses make contrasting predictions – has been more challenging than many researchers realized (Dove, 2009; Machery, 2007).

3.1 First successful tests: *Where* does simulation happen in the brain?

The first successful tests of embodied simulation, that satisfied both of these criteria, were conducted by Freidemann Pulvermüller and colleagues (Hauk et al., 2004; Pulvermüller, 2005). Pulvermüller's team identified a prediction that follows from the embodied simulation theory but does not follow from the disembodied alternative: Understanding sentences that refer to actions should selectively activate modality-specific parts of the motor system that are necessary for planning and / or executing these actions. To test this prediction, Hauk, Johnsrude, and Pulvermüller (2004) introduced a new experimental paradigm using Functional Magnetic Resonance Imaging (fMRI). With fMRI, researchers can track changes in cerebral blood flow that accompany neural activity, and can determine which parts of the brain are associated with which mental activities (e.g., understanding a sentence). Hauk et al. (2004) tested participants while they were engaged in two separate tasks: a cognitive task, and then a motor task. The cognitive task was silently reading sentences with action verbs referring to either foot actions (e.g., *kick*), mouth actions (e.g., *lick*), or hand actions (e.g., *pick*). In the first part of the study participants read these sentences while lying still in the fMRI scanner, *not* moving any of these body parts. In the second part of the study, they were instructed to move their feet, mouth, and hands. The cognitive task was of primary interest: The goal was to determine what brain areas were active while participants read about hand, foot, and mouth actions. The motor task was not testing any hypothesis, per se. Rather, it served as a *functional localizer*: a task that allows researchers to identify, in individual subjects' brains, regions that are known to serve a behavioral function. In this case, the goal was to identify the areas of the motor cortex that enable foot, mouth, and hand actions. In the context of this study (and subsequently many others), these motor areas served as *regions of interest* (often abbreviated *ROI*s): brain regions that are already known to serve a particular function (e.g., programming motor actions), which guide researchers' search for the locus of a cognitive function of interest (e.g., the semantics of action verbs).

The motor cortex is organized in a way that makes this brain area particularly useful as a testbed for embodied simulation. Stretched out over part of the cerebral cortex is a *somatotopy* (i.e., a body map). Each part of our bodies corresponds to a patch of motor cortex that controls this body part's movements, and neighboring body parts (e.g., thumb, index finger) correspond to neighboring parts of the cortex. A secondary motor area, the premotor cortex (involved in planning motor actions), has roughly the same somatotopic organization as the primary motor cortex (necessary for executing motor actions). This somatotopy allows for precise prediction of where in the brain simulations should and should not be found: Foot verbs should preferentially activate foot areas of motor and premotor cortex (more than they activate mouth or hand areas); hand verbs should preferentially activate hand areas of motor and premotor cortex (more than they activate mouth or foot areas); etc.

Hauk, Johnsrude, and Pulvermüller (2004) found that action verbs cued a somatotopic pattern of brain activity. Activity cued by the foot, mouth and hand sentences overlapped with activity found in the foot-, mouth-, and hand-action regions of interest. This result was predicted by the embodied simulation hypothesis but not by the disembodied alternative, according to

which the somatotopic motor areas should play no role in understanding sentences or representing the meanings of verbs.

This result has been replicated and extended by multiple studies testing for "semantic somatotopy" cued by action verbs in the motor and premotor cortices (for reviews see Fischer & Zwaan, 2008; Willems & Casasanto, 2011). Beyond the motor system, analogous results have been found for perceptual simulations driven by language about perceptible things: Understanding color words activates visual areas specialized for color perception (Simmons, et al., 2007); understanding words for fragrant things activates olfactory perception areas (González et al., 2006); understanding sentences about motion events activates visual areas specialized for motion perception (Wallentin et al., 2011). Together, these fMRI studies provide an initial body of evidence that supports the embodied simulation hypothesis by showing that modality-specific perceptual and motor areas are selectively activated by language referring to our perceptual and motor experiences.

3.2 *When* does simulation happen in the brain?

Studies that use fMRI to show *where* language-driven activity in the brain is happening have answered a first important question: Does activity in modality-specific perceptual and motor areas correlate with the process of computing the meanings of words (and sentences) that refer to perceptuo-motor experiences? But these studies leave open a further question: *When*, precisely, is this language-driven modality-specific activity happening? FMRI is not a suitable tool for addressing this question. After a stimulus (e.g., a word) is presented to an experimental participant, neurons that are receptive to that stimulus begin responding almost instantaneously. However, as mentioned above, the fMRI signal does not index neural activity directly: It indexes changes in cerebral blood flow that occur in response to stimulus-driven neural activity. These changes in blood flow may peak four to five seconds after the neural event that necessitated them. Given this lag in the blood flow response (which is somewhat variable and depends on many factors), fMRI does not allow researchers to determine precisely how much time has passed between the presentation of a stimulus and the brain's response to it.

Why does timing matter with respect to the embodied simulation hypothesis? Simulations are posited to be the stuff of thought, and the stuff of linguistic meaning (or, at a minimum, *some of* the stuff of thinking and meaning construction). Much of our thinking happens fast, on the order of tens or hundreds of milliseconds, as does much of ordinary language understanding. Therefore, to be the stuff of cognition and semantics, simulations would need to happen *fast.*

How fast do people understand words? A neurolinguistics literature that predates the embodied simulation hypothesis provides some precise information, and sets a lower 'speed limit' for how fast simulations would need to happen in order to fulfill the role in our mental lives that embodied simulation theorists posit. For more than four decades, researchers have used Electroencephalography (EEG) to measure electrical signals generated by the brain in response to linguistic stimuli. According to this literature, readers can understand a word, and determine whether a newly-presented word is sensible in its linguistic context, in about 400 milliseconds (Kutas & Hilyard, 1980). Therefore, for simulations to be the stuff of meaning, they would need to occur in less than 400 milliseconds. Do they? In order to answer this question, researchers need a tool that has better temporal precision than fMRI and also greater spatial precision than EEG (which can indicate precisely when a neural event occurred but does not typically give precise information about where it occurred).

To determine how long it takes to generate a modality-specific simulation, Pulvermüller and colleagues used a third kind of brain imaging, which combines high spatial resolution (like

fMRI) with high temporal resolution (like EEG): Magnetoencephalography (MEG). MEG measures the weak magnetic fields that are induced by electrical currents flowing through neurons. Pulvermüller and colleagues determined the instant when spoken action verb stimuli could be uniquely identified, and measured how much time elapsed between wordform identification (in the left Temporal lobe, near auditory cortex) and the appearance of somatotopic activity in the motor system (e.g., activity in foot-motor areas for verbs like *kick*). Results showed that somatotopic motor simulations could be detected within tens of milliseconds after wordform identification, suggesting these simulations were indeed happening fast enough to play the role in linguistic meaning construction that is posited by the embodied simulation hypothesis (for a review of relevant MEG studies see Pulvermüller, 2005).

3.3 Do simulations play a causal role in understanding words?

Results from fMRI studies show that modality-specific brain areas are activated in response to linguistic stimuli, and MEG studies show that this activation happens fast enough to be relevant to the process of linguistic meaning construction. Do these results demonstrate that meaning is represented, at least in part, in neurocognitive systems for perception and action? No, these results lay the groundwork for such a conclusion, but they do not license this conclusion, per se. Why not? Because typical brain imaging studies can only demonstrate a *correlation* between patterns of brain activity and patterns of thinking, and *correlation does not imply causation.*

Inferring a causal relationship, for example, between somatotopic motor activity shown in fMRI and MEG studies and the comprehension of action verbs would be an error in statistical reasoning. In order to test for a causal relationship, a different kind of experiment is needed. An illustration of what kind of studies can only show correlation (like the brain imaging studies reviewed above), and what kind of further study is needed to support causal inferences: Medical studies often test for correlations between a behavior and a medical outcome. Imagine a study that tested thousands of people who drink caffeine and found that the more caffeine they reported drinking the more likely they were to have a heart attack. A tempting conclusion would be: Caffeine causes heart attacks. But this imaginary study would not license this conclusion: There could be a strong correlation between caffeine and heart attacks even if there were no causal relationship, at all. For instance, the real causal factor could be: Stress. Maybe people with more stressful jobs: (a) drink more caffeine to keep themselves motivated, and also (b) have more heart attacks because *stress* causes heart attacks, not caffeine.

In order to determine whether there is a causal relationship between caffeine and heart attacks a second type of medical study would need to be conducted: a randomized controlled trial (RCT). Whereas correlational studies measure naturally-occurring relationships between two variables (e.g., caffeine consumption, heart attacks), RCTs intervene on naturally-occurring relationships, manipulating one variable in order to determine how it influences the other. To test the caffeine-heart attack relationship, our imaginary medical researchers could randomly assign groups of participants to different *treatments* (e.g., caffeine drinking, no caffeine drinking) and then measure whether there was a difference in the number of heart attacks between the two groups, post-treatment.

How can a RCT study help researchers test for a causal relationship between modality-specific brain activity and linguistic meaning construction? To determine whether activity in modality-specific areas plays any causal role in language understanding, it is necessary to manipulate activity in these brain areas and measure whether this intervention on the brain influences how participants process the meanings of words. One tool for directly manipulating

brain activity in healthy experimental participants is Transcranial Magnetic Stimulation (TMS). With TMS, researchers can place a magnetic coil on the scalp and pulsate a magnetic field in order to selectively increase or decrease neural activity in the cortical area beneath the coil. Willems et al. (2011) used TMS to modulate activity in the hand areas of participants' left or right premotor cortex. After treatment, participants performed a standard task known to elicit activation of words' meanings called a *lexical decision task*: For each stimulus shown, participants judged whether the stimulus was a real English verb or a meaningless pseudo-verb (e.g., *to wunger*), as quickly as possible. Only responses to the real verbs were of interest. Each verb either referred to a manual action that is typically performed with one's dominant hand (e.g., to write) or a non-manual action (e.g., to wander). The experimenters reasoned that, if motor simulation consists in partially preparing the brain's motor system to perform the action named by the verb, then only the *left* premotor hand area would be involved in simulating the manual actions (since the left hemisphere of the brain controls the right hand, and all of the participants were right-handed). Furthermore, neither premotor hand area would be involved in simulating the non-manual actions, since the hand area only programs hand actions (not actions with other parts of the body). The results supported these predictions: Stimulating the left premotor cortex (but not the right premotor cortex) influenced how quickly participants could respond to the manual verbs (but not the non-manual verbs). This finding provided some of the first evidence that somatotopic motor activity plays a causal role in processing action verbs' meanings.[7]

3.4.1 How embodied semantics cannot be tested

This chapter focuses almost exclusively on studies using methods from Cognitive Neuroscience, which allow questions about embodied semantics to be addressed by examining relationships between neural and cognitive activity. Another related body of studies has emerged over the first two decades of the 21st century which uses behavioral methods from Cognitive Psychology. This body of research, which includes thousands of experiments, has been omitted from this review of foundational evidence for embodied semantics because, in most cases, these studies do not accomplish their goal of testing the embodied simulation hypothesis.

In one influential study from this behavioral literature, Zwaan and Yaxley (2003a) showed participants pairs of words, one presented above the other, which named pairs of objects that have canonical vertical positions with respect to each other (e.g., *cup*, *saucer*). The experimenters compared how long participants took to judge whether the words were related to each other, depending on whether the arrangement of the words on the computer screen matched the canonical positions of their referents (e.g., cup above saucer) or mismatched their canonical positions (e.g., saucer above cup). Participants judged words faster when their relative locations on the screen matched their referents' typical locations in the world. This result is often interpreted as evidence that participants understand the meanings of words referring to visible objects, at least in part, using modality-specific simulations of the referents, implemented in visual cortex. Yet, in addition to this embodied explanation for Zwaan and Yaxley's (2003a) match-mismatch effect, there are also clear disembodied alternative explanations.

---

[7] Pulvermüller et al. (2005) conducted a similar TMS study prior to Willems et al. (2011), but the pattern of data they obtained made the results hard to interpret with respect to the embodied simulation hypothesis. Other studies have used a different kind of TMS protocol in which single magnetic pulses are applied to motor cortex while participants process limb-specific action verbs (e.g., write, kick), and muscle activity in the relevant limb is measured (e.g., Papeo et al., 2009). These single-pulse TMS studies do not support direct inferences about a causal role for the motor system in processing language because the dependent measure (that which is influenced directly by the TMS treatment) is a measure of muscle activity, not of language processing.

To Zwaan and Yaxley's (2003a) credit, the authors acknowledged these alternatives, suggesting the following as one plausible disembodied explanation for their match-mismatch effect, which would not implicate any modality-specific brain areas in the process of constructing and comparing the words' meanings. For the pair of stimulus items (1) BRANCH presented above ROOT, or (2) ROOT presented above BRANCH, the authors suggested that:

> [A] spatial "tag" is attached to each word. In a semantic network, a concept like BRANCH would have a link with concepts such as *top*, given that branches are typically found in the top parts of trees. As a consequence, for the pair in (1), TOP would be attached to BRANCH and BOTTOM to ROOT, whereas the reverse would happen in (2). In the case of (2), this would yield a conflict between the spatial tags and the information in semantic memory. This conflict would delay the activation above threshold of the concept pair, thus delaying the response (p. 957).

Zwaan and Yaxley (2003a) also acknowledged that similar results had been obtained years before the embodied simulation hypothesis had been formulated (MacLeod, 1991), which were necessarily motivated and explained by disembodied theories of language. To the authors' further credit, Zwaan and Yaxley (2003b) conducted a subsequent study in which they attempted to resolve the ambiguity of these results by grounding their experimental predictions in patterns of brain activity that would be compatible or incompatible with an embodied account.

Zwaan and Yaxley's (2003a) study, and thousands of behavioral studies like it, failed to test the embodied simulation hypothesis effectively *not* because of idiosyncratic aspects of the particular experiments, but rather because of their violation of a general principle of experimental design. In order for an experiment to distinguish between two competing hypotheses (e.g., embodied semantics, disembodied semantics), the competing hypotheses must make contrasting predictions. If two hypotheses both predict the same pattern of results (for different reasons), then obtaining this pattern is uninformative: The result supports both the hypothesis that the experimenters hoped to confirm *and* the hypothesis that they hoped to disconfirm. Although it is possible to design purely behavioral studies that test the embodied simulation hypothesis effectively (e.g., Shebani & Pulvermüller, 2013; Escámez et al., 2020), most behavioral tests of embodied semantics to date share the same fatal flaw: The competing hypotheses do not make contrasting predictions. In principle, it is possible that nearly any cognitive function could be implemented in *either* modality-specific systems or non-modality-specific systems in the mind; behavioral results are generally compatible with either of the in-principle possibilities. Therefore, in general, studies using methods from Cognitive Neuroscience have been more successful than behavioral studies at generating predictions that are capable of confirming one account of semantics (either embodied or disembodied) and disconfirming the alternative, on the basis of whether the predicted patterns of neural activity are found in modality-specific brain areas or only in non-modality-specific areas.

3.4 Summary and open questions

Together, studies using fMRI, MEG, and TMS established a body of evidence that confirmed key predictions of the embodied simulation hypothesis, and challenged the 'disembodied' alternative that dominated theories of language in the mind and brain through the end of the 20th century. Modality-specific brain areas including the visual cortex, auditory cortex, and motor cortex, which have long been known to play a crucial role in perceiving and producing the forms of words, now appear to be involved in instantiating the meanings of words, as well. Beyond showing that these modality-specific brain areas are active during language understanding via fMRI studies, researchers have used MEG to show that activity in modality-

specific areas happens quickly enough to meet the requirements of rapid, online language processing, and therefore to play the role in linguistic meaning construction that simulation theorists posit. Beyond showing a correlation between modality-specific brain activity and language understanding, studies using TMS provide preliminary evidence that modality-specific areas play a causal role in processing word meaning.

The majority of the studies reviewed here were conducted during roughly the first decade of the 21st century. These studies validated the most fundamental tenets of the embodied simulation hypothesis. During the subsequent decade, in addition to replicating and incrementally extending these foundational studies, researchers have turned to questions for which no complete answer has yet emerged. For example: *How much* of meaning is embodied in modality-specific simulations? Intervening on modality-specific brain areas has been shown to produce small changes in response times (Willems et al., 2011) or accuracy (Gijssels et al., 2018) in judging relevant words, but there is little evidence that modulating modality-specific brain activity has any substantial influence on the process of meaning construction (cf., Escámez et al., 2020).

Another open question concerns the extent to which semantic representations differ between individuals and groups as a result of their differing bodily experiences. If thinking and language understanding is (partly) constituted by simulations of our own perceptuo-motor experiences, then do people with different kinds of bodies who perceive or act upon the world in systematically different ways, also think differently in corresponding ways? A body of research has shown that, indeed, people with different kinds of bodies construct predictably different feelings, object representations, mental images, and word meanings, and these thoughts are implemented in predictably different modality-specific brain areas (Casasanto, 2011). Yet, this research on the 'body-specificity' of language and thought remains in its early stages, in part because only one bodily difference between individuals (i.e., their handedness) has been extensively explored.

Perhaps chief among the open questions about embodied simulation is: How can modality-specific simulations represent abstract concepts, like time, justice, or happiness? Concrete objects like cats can be represented via perceptual simulations; concrete actions like throwing can be represented via motor simulations. But how can we use perceptual or motor simulations to represent abstract ideas of things we can never perceive with the senses or act upon with the muscles? Three potential solutions have been pursued by researchers, but none of these pursuits has been particularly successful. First, perhaps abstract ideas can be embodied via metaphorical mental representations (Lakoff & Johnson, 1999)? For example, suppose the abstract notion of *understanding* were conceptualized, in part, via the concrete action of *grasping*, as suggested by expressions like "she grasped the idea." If so, then at least part of the semantics of *understanding* could be a motor simulation of grasping, in the hand motor areas that support literal grasping. Although this possibility remains plausible and well-motivated, numerous experimental tests have failed to provide any clear support for it. Whereas literal sentences like "she grasped the knife" reliably activate somatotopic hand areas, metaphorical sentences like "she grasped the idea" do not (for a review see Casasanto & Gijssels, 2015).

According to another proposed solution, perhaps abstract ideas can be embodied via simulation of the complex situations in which these ideas are experienced and used (Barsalou, 1999). For example, perhaps *justice* can be understood via simulating a courtroom scene? Yet, this proposal faces both in-principle and empirical challenges. In principle, a courtroom scene could be simulated in rich multi-modal detail (like an audio-video recording in one's brain and

mind), but *still* the person experiencing this courtroom simulation would not necessarily understand justice. Empirically, the study that has tested the embodiment of 'situation models' most directly did not report any clear modality-specific brain activity corresponding to situation model construction, even for concrete concepts (Simmons et al., 2008).

According to a third proposal, perhaps abstract concepts are embodied via simulations of affective (i.e., emotional) experiences (Meteyard et al., 2012)? Consistent with this proposal, abstract words are statistically more likely than concrete words to have affective content as part of their semantics (Kousta et al., 2009). Yet, this proposal also faces a priori and empirical challenges. A priori, this proposal is not likely to be a complete answer to the problem of embodying abstract concepts (and the corresponding word meanings) because many abstract ideas have no clear emotional charge (e.g., time, neutrality, multiplication, quark, etc.) Empirically, this proposal is hard to evaluate with respect to the embodied simulation hypothesis because the brain structures that support our primary experience of emotions (e.g., the amygdala) are multifunctional and non-modality-specific.

Ultimately, the resolution to these and other outstanding questions will determine whether the discovery of modality-specific simulations simply modifies 20th-century theories of concepts and semantics in the brain and mind, or revolutionizes them.

References

Anderson, M. L. 2003. Embodied cognition: A field guide. *Artificial Intelligence*, *149*(1), 91-130.

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

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. 2003. Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84-91.

Brooks, R. A. 1990. Elephants don't play chess. *Robotics and Autonomous Systems*, *6*(1-2), 3-15.

Casasanto, D. 2011. Different bodies, different minds: the body specificity of language and thought. *Current Directions in Psychological Science*, *20*(6), 378-383.

Casasanto, D., & Gijssels, T. 2015. What makes a metaphor an embodied metaphor? *Linguistics Vanguard*, *1*(1), 327-337.

Damasio, A. R. 1989. Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1-2), 25-62.

Dove, G. 2009. Beyond perceptual symbols: A call for representational pluralism. *Cognition*, *110*(3), 412-431.

Escámez, O., Casasanto, D., Vigliocco, G., & Santiago, J. 2020. Motor interference changes meaning. In S. Denison., M. Mack, Y. Xu, & B.C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*.

Fischer, M. H., & Zwaan, R. A. 2008. Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology*, *61*(6), 825-850.

Fodor, J. A. 1983. *The modularity of mind*. MIT press.

Gijssels, T., Ivry, R. B., & Casasanto, D. 2018. tDCS to premotor cortex changes action verb understanding: Complementary effects of inhibitory and excitatory stimulation. *Scientific Reports*, *8*(1), 1-7.

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

Hagoort, P. 2005. On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, *9*(9), 416-423.

Hauk, O., Johnsrude, I., & Pulvermüller, F. 2004. Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*(2), 301–7.

Kousta, S.T., Vinson, D.P., & Vigliocco, G. 2009. The role of emotional valence in the processing of words. *Cognition*, *112*, 473-481.

Kutas, M., & Hillyard, S. A. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205.

Lakoff, G. & Johnson, M. 1999. *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Basic Books.

Longo, M. R., Schüür, F., Kammers, M. P., Tsakiris, M., & Haggard, P. 2008. What is embodiment? A psychometric approach. *Cognition*, *107*(3), 978-998.

Machery, E. 2007. Concept empiricism: A methodological critique. *Cognition*, *104*(1), 19-46.

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.

Niedenthal, P. M. 2007. Embodying emotion. *Science*, *316*(5827), 1002-1005.

Papeo, L., Vallesi, A., Isaja, A., & Rumiati, R. I. 2009. Effects of TMS on different stages of motor and non-motor verb processing in the primary motor cortex. *PloS One*, *4*(2), e4508.

Pulvermüller, F. 1999. Words in the brain's language. *Behavioral and Brain Sciences, 22*(2), 253–336.

Pulvermüller, F. 2005. Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576-582.

Shebani, Z., & Pulvermüller, F. 2013. Moving the hands and feet specifically impairs working memory for arm- and leg-related action words. *Cortex*, *49*, 222-231.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. 2007. A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802-2810.

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. 2008. fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology-Paris*, *102*(1-3), 106-119.

Varela, F. J., Thompson, E., & Rosch, E. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. MIT press.

Wallentin, M., Nielsen, A. H., Vuust, P., Dohn, A., Roepstorff, A., & Lund, T. E. 2011. BOLD response to motion verbs in left posterior middle temporal gyrus during story comprehension. *Brain and Language*, *119*(3), 221-225.

Willems, R. M., & Casasanto, D. 2011. Flexibility in embodied language understanding. *Frontiers in Psychology*, *2*, 116.

Willems, R. M., Toni, I., Hagoort, P., & Casasanto, D. 2010. Neural dissociations between action verb understanding and motor imagery. *Journal of cognitive neuroscience*, *22*(10), 2387-2400.

Wilson, M. 2002. Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625-636.

Zwaan, R. A., & Yaxley, R. H. 2003a. Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review*, *10*(4), 954-958.
Zwaan, R. A., & Yaxley, R. H. 2003b. Hemispheric differences in semantic-relatedness judgments. *Cognition*, *87*(3), B79-B86.