Embodied view of language

Embodied perspectives on concepts (e.g. Barsalou, 1999; Glenberg, 1997) emphasize that cognition is shaped by the physical properties of the world (i.e. “grounded”) in multiple ways (by simulations, or, occasionally, by bodily states); that our concepts are shaped by the physical constraints of our body (i.e. “embodied”); that cognitive processing strongly depends on current constraints and task demands (i.e. “situated”; see Pezzulo et al., 2013). Behavioral and brain-imaging studies collected in the last few years converge in indicating that seeing an object activates motor information (e.g. Tucker and Ellis, 1998; Ellis and Tucker, 2000; Grèzes, Tucker, Armony, Ellis, and Passingham, 2003). Importantly the physical context, i.e. the actual possibility of reaching the object (Costantini, Ambrosini, Scorolli, and Borghi, 2011; Ambrosini, Scorolli, Borghi, and Costantini, 2012), and the social context, i.e. the presence of another agent (virtual avatar: Costantini, Committeri, and Sinigaglia, 2011; or other person: Gianelli, Scorolli, and Borghi, 2013), seem to modulate this information.

Nevertheless not only the visual perception of an object, but also language processing elicits the involvement of the motor system. As words act as surrogates for more direct interactions with the environment (Glenberg, 1997), language processing recruits brain areas typically involved in perceiving objects and interacting with them. Across psychology, neuroscience and cognitive linguistics (see Bergen, 2005; Gibbs, 2003; Pecher and Zwaan, 2005), strong empirical evidence provides support for the view that language comprehension results in embodied representations (Gallese and Goldman, 1998; Zwaan, 2004; Glenberg and Kaschak, 2002), that is in “simulations”. The notion of “simulation” is still a topic of debate, but there is general agreement in understanding it as the activation of the same sensorimotor neural correlates involved when we previously interacted with the referent of the word, i.e. the actual object (e.g. “apple”), or we executed the specific action described by the word (e.g. “grasping”) (Gallese 2008; Gallese and Sinigaglia, 2011).

The specificity of the simulation

Much recent evidence suggests that the mental simulation we run when comprehending language is not grounded in previous sensorimotor experience of a generic sort, but instead invokes rather specific sensorimotor experiences.
Object properties

Embodied simulations are sensitive to differences pertaining to intrinsic object properties (e.g. shape: Zwaan, Stanfield, and Yaxley, 2002; size: Glover and Dixon, 2002; Borghi and Riggio, 2009; color: Connell, 2007) as well as extrinsic ones (e.g. orientation: Stanfield and Zwaan, 2001; Borghi and Riggio, 2009; distance: Winter and Bergen, 2012), as shown by using the sentence-picture verification task. Participants read a sentence in which the shape/orientation/color of an object was implied (e.g. “John put the pencil in the cup” vs. “in the drawer”). Immediately after they were presented with an object’s picture that could match or mismatch the object’s property evoked by language (e.g. the pencil’s axis of symmetry in vertical vs. horizontal plane). They had to judge whether the shown object was mentioned in the sentence. Crucially the pictured object’s shape/orientation/color was irrelevant to the task. Nonetheless participants’ responses were modulated by the properties implied by the previous sentence, with a mismatch disadvantage (i.e. facilitation effect). Beyond the objects’ features, comprehenders seem to simulate also the direction of the objects’ motion (Zwaan, Madden, Yaxley, and Aveyard, 2004; Kaschak et al., 2005): participants were faster in making judgments on sentences describing a motion in an opposite direction to the one shown by a simultaneously presented visual stimuli (i.e. interference effect). To disentangle facilitation and interference effects, two accounts have been proposed: temporal overlap (i.e. the relative timing of the perceptual simulation and the conceptual task) and integratability (i.e. the extent to which the perceptual simulation can be integrated into the simulation constructed during language comprehension).

Furthermore concepts seem to be represented by modality-specific sensory-motor systems (Barsalou, Pecher, Zeelenberg, Simmons, and Hamann, 2005); therefore the described simulation processing does not involve only the visual modality, but also, for example, the auditory one (Pecher, Zeelenberg, and Barsalou, 2003).

Relational properties

As conceptual knowledge guides action in the world, the simulations activated by language processing do not imply only perceptual details (e.g. visual or auditory simulations), but also motor ones. They involve the specific static or dynamic relation between the agent and the objects, as their relative positions, or the kind of interaction. Using a part verification task Borghi, Glenberg, and Kaschak (2004) found that response latencies were affected by the perspective (inside vs. outside) induced by language (e.g. “You are eating in a restaurant”: advantage for “table”; “You are waiting outside a restaurant”: advantage for “sign”). Varying participants’ motor responses (upward vs. downward movement) they also found an advantage for responses’ direction compatible with the object part’s typical location (e.g. responding upward to verify that “a doll” does have “hair”).

Still manipulating the actions required to perform the task, Glenberg and Kaschak (2002) directly tested if the comprehension of sentences referring to oppositely directed actions affected response times. Even if they were never instructed to consider the direction implied by the sentence, participants were faster and more accurate in judging the sensibility of sentences when there was a match between the movement direction implied by the sentence (e.g. “Put your finger under your nose” vs. “under the faucet”) and the one required to respond (button near vs. far from the body) rather than in the case of mismatch. This action-sentence compatibility effect further supports the embodied claim that language understanding is grounded in bodily action (see Table 13.1).
Table 13.1 Examples of evidence on the specificity of simulation activated by language understanding.

Referential aspect of language:
Specificity of the simulation activated by concrete language

<table>
<thead>
<tr>
<th>Semantic features</th>
<th>Object properties</th>
<th>Relational properties</th>
<th>Grammatical categories and constructions</th>
<th>What part of the event to focus on</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intrinsic object features</td>
<td>Extrinsic object features</td>
<td>Direction of object motion</td>
<td>Perspectives</td>
</tr>
<tr>
<td>SHAPE</td>
<td>ORIENTATION</td>
<td>Zwaan et al., 2004; Zwaan, 2001; Kaschak et al., 2005</td>
<td>Borghi et al., 2004; Brunyé et al., 2009; Sato and Bergen, 2013</td>
<td>Glenberg and Kaschak, 2002; Glenberg et al., 2008; Bergen and Wheeler, 2010</td>
</tr>
<tr>
<td>Zwaan et al., 2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIZE</td>
<td>ORIENTATION</td>
<td>Stanfield and Zwaan, 2001; Kaschak et al., 2005</td>
<td>Borghi et al., 2004; Brunyé et al., 2009; Sato and Bergen, 2013</td>
<td>Glenberg and Kaschak, 2002; Glenberg et al., 2008; Bergen and Wheeler, 2010</td>
</tr>
<tr>
<td>Borghi and Riggio, 2009</td>
<td>ORIENTATION</td>
<td>Stanfield and Zwaan, 2001; Kaschak et al., 2005</td>
<td>Borghi et al., 2004; Brunyé et al., 2009; Sato and Bergen, 2013</td>
<td>Glenberg and Kaschak, 2002; Glenberg et al., 2008; Bergen and Wheeler, 2010</td>
</tr>
<tr>
<td>COLOR</td>
<td>ORIENTATION</td>
<td>Stanfield and Zwaan, 2001; Kaschak et al., 2005</td>
<td>Borghi et al., 2004; Brunyé et al., 2009; Sato and Bergen, 2013</td>
<td>Glenberg and Kaschak, 2002; Glenberg et al., 2008; Bergen and Wheeler, 2010</td>
</tr>
<tr>
<td>MASS</td>
<td>ORIENTATION</td>
<td>Stanfield and Zwaan, 2001; Kaschak et al., 2005</td>
<td>Borghi et al., 2004; Brunyé et al., 2009; Sato and Bergen, 2013</td>
<td>Glenberg and Kaschak, 2002; Glenberg et al., 2008; Bergen and Wheeler, 2010</td>
</tr>
<tr>
<td>Scorolli et al., 2009; Scorolli, 2011</td>
<td>ORIENTATION</td>
<td>Stanfield and Zwaan, 2001; Kaschak et al., 2005</td>
<td>Borghi et al., 2004; Brunyé et al., 2009; Sato and Bergen, 2013</td>
<td>Glenberg and Kaschak, 2002; Glenberg et al., 2008; Bergen and Wheeler, 2010</td>
</tr>
</tbody>
</table>
The effector and the final goal

The semantics of verbs does not imply only a specific direction of action (e.g. “opening” vs. “closing”) but also a specific part of the agent’s body (e.g. “grasping” vs. “kicking”), and possibly a specific final goal (e.g. “grasping the toffee to eat it” vs. “grasping the aubergine to cook it”). In order to further investigate the kind of relationship existing between language and the motor system recent studies used single verbs (Pulvermüller, Harle, and Hummel, 2001; Buccino et al., 2005), or verbs embedded in sentences (Scorolli and Borghi, 2007; Borghi and Scorolli, 2009), that could refer to hand-arm/foot-leg/mouth actions (see Table 13.1). Depending on the methodology used, response latencies from different effectors (behavioral paradigms) or event-related current source densities from brain electrical activity (electroencephalography, EEG) and motor-evoked potentials from hand muscles (transcranial magnetic stimulation, TMS) were collected. Neurophysiological data converged in demonstrating that action words elicit neurophysiological activity with different cortical topographies (e.g. Pulvermüller et al., 2001; Buccino et al., 2005), reflecting the words’ meanings. Consistently, behavioral findings (e.g. Buccino et al., 2005; Scorolli and Borghi, 2007) highlighted a modulation determined by the match or mismatch10 between the specific effector referred to by language and the one used to respond (e.g. sentences: “to throw the ball” vs. “to kick the ball”; response-effector: hand vs. foot).

Finally, action words seem to be represented and encoded not just at a proximal level but also at a distal level (see Hommel, Müsseler, Aschersleben, and Prinz, 2001), as the simulation entailed by language processing is sensitive also to the final goal expressed by the sentence. Coherently Borghi and Scorolli (2009) found an advantage for verb–noun pairs referring to manual actions (e.g. “to pick up – grapes”) and mouth actions (e.g. “to eat – grapes”), but not for pairs referring to foot actions (e.g. “to press – grapes”) in case of dominant-right-hand responses. This effect is consistent with the fact that usually an action with the mouth (e.g. eating a slice of bread), but not with the foot, implied a previous action with the hand (e.g. slicing the bread and bringing it to the mouth); it is also consistent with evidence indicating that at the neural level hand and mouth actions activate contiguous regions (e.g. Pulvermüller et al., 2001). It is worth noting that these findings support striking proposals suggesting that language could have evolved from gestures (e.g. Corballis, 2002; Arbib, 2005).

Grammatical features

The reviewed evidence suggests that the simulation triggered by language is affected by content words, like nouns and verbs. Nonetheless another aspect to be considered is whether, and possibly how, grammar affects language-driven mental simulations. The meaning of noun/verb specifies which kind of object/action to simulate, i.e. their properties. However, given a scene to be simulated, it is important to identify also higher-order characteristics of simulation (see Glenberg and Gallese, 2012; Sato, Schafer, and Bergen, 2013; see Table 13.1): grammar serves this important function. Bergen and Wheeler (2010) found that progressive aspect drives comprehenders to mentally simulate the central process of a described motor event, while perfect aspect does not. These effects (“second-order effects” of grammar) are qualitatively different from the effects of content words, as grammatical aspect markers operate over the representations evoked by content words, for example modulating the specific part of simulation on which to focus (e.g. “Virginia is brushing her teeth” vs. “Virginia has brushed her teeth”: progressive aspects accentuate the internal structure of an event, while perfect aspects highlight the resulting end state, see Dowty, 1977).

The issue of abstractness

The evidence above strongly supports the main claim of standard embodied accounts: concepts are grounded in perception and action systems, and therefore they are modal. However, all these
empirical studies focus on words with concrete referents (e.g. highly imageable words), and have limited reach with respect to abstract concepts (Louwerse and Jeuniaux, 2010; Borghi and Cimatti, 2009; 2012; Dove, 2009; Mahon and Caramazza, 2008). While many studies have been devoted to this important topic, the issue of abstractness remains puzzling (for a recent review, see Pecher, Boot, and van Dantzig, 2011; see also the special issues edited by Scorolli, 2012, and by Borghi and Pecher, 2012).

Some studies show that abstract words refer metaphorically to concrete referents (Lakoff and Johnson, 1999; Casasanto and Boroditsky, 2008), that abstract sentences recruit the motor system (Glenberg et al., 2008; see also Glenberg and Gallese, 2012) and that abstract concepts elicit situations, as well as simulations of internal states (Barsalou and Wiemer-Hastings, 2005). This research, though compelling, refers only to a limited subset of phenomena (see Borghi and Cimatti, 2009).

In an effort to generalize these results, a first step could be to study the abstract-concrete dimension in a continuum (Wiemer-Hastings, Krug, and Xu, 2001). In a recent cross-cultural study Scorolli, Binkofski, et al. (2011) combined the same concrete verb (action-related verb, e.g. “to grasp”) with a concrete noun (graspable object, e.g. “a flower”) and with an abstract noun (non-graspable object, e.g. “a concept”), and then an abstract verb (verb not directly related to action, e.g. “to describe”) with the nouns previously used. Participants had to evaluate sensibility of the combinations. To take into account also grammatical features, two syntactically different languages were considered: German and Italian (in the former the nouns precede the verbs). Compatible combinations (concrete-concrete, abstract-abstract) were processed faster than mixed ones, and did not differ, whereas processing of mixed combinations (concrete-abstract, abstract-concrete) was modulated by the specific language: when concrete words preceded abstract ones responses were faster (regardless of the word grammatical class). Results on compatible combinations can be explained by modal theories, but also by amodal ones. Conversely findings from the mixed combinations are consistent with the claim that both abstract and concrete words are represented modally, and they activate parallel systems: one relying more on perception and action areas, the other more on sensorimotor linguistic areas (switching between systems implies a cost: Pecher et al., 2003). The higher difficulty of abstract words could reflect their later acquisition (Borghi and Cimatti, 2009, 2012), given that linguistic experience is particularly relevant for concrete words.

The same paradigm was used also in studies involving both TMS (Scorolli, Jacquet, et al., 2012) and fMRI (functional magnetic resonance: Sakreida et al., 2013). They converged in showing that both concrete and abstract multiword expressions engage core areas of the sensorimotor neural network (Scorolli, Jacquet, et al., 2012; Sakreida et al., 2013). But while phrases containing concrete verbs imply a direct early activation of the hand-related motor system, the activation of the same system is delayed in the case of phrases containing abstract verbs (Scorolli, Jacquet, et al., 2012): the processing of abstract verbs could first engage mouth-related motor areas, that later affect the contiguous areas (hand areas). Furthermore, processing of abstract noun–abstract verb combinations compared with concrete language content shows a pronounced activation in the left anterior middle temporal gyrus (Sakreida et al., 2013), an area close to the language-processing system (see Price, 2010).

**Multiple types of representations**

How to explain these results? Attempts to sketch a framework to address both concrete and abstract words representation suggest that multiple representational systems\(^{11}\) are activated during conceptual processing. Some of these proposals adopt an embodied perspective\(^{12}\) and do not
assume a transduction process from sensorimotor experience to abstract–amodal symbols. Barsalou, Santos, Simmons, and Wilson (2008) have proposed the Language And Situated Simulation theory (LASS), claiming that linguistic forms and situated simulations interact continuously: different mixtures of the two systems underlie a wide variety of tasks. The linguistic system (left-hemisphere) is involved mainly during superficial linguistic processing, whereas deeper conceptual processing requires the activation of the sensorimotor system (bilateral posterior areas). In a similar vein, Borghi and Cimatti (2009, 2012; Words As Tools theory, WAT) propose to extend the embodied view of cognition in order to consider not only language grounding but also the role that language plays in shaping our experience. In contrast to LASS, WAT claims that the linguistic system does not simply involve a form of superficial processing. WAT makes specific predictions concerning abstract and concrete word representations: while both engage the sensorimotor neural network (Scorolli, Jacquet, et al., 2012; Sakreida et al., 2013), abstract words specifically activate the linguistic neural network. Why? According to the authors the reason is to be found in the different acquisition mechanisms for concrete and abstract words (Borghi and Cimatti, 2009, 2012).

**The acquisition of new words**

In acquiring concrete words, we first experience the concrete entities (e.g. pencil), and then we tag them using linguistic labels (we learn the name “pencil”). With abstract words, we might initially learn a word (the label) and then “tag” it with our sensorimotor experience, that is we use the word to assemble a set of experiences (e.g. we probably put together different experiences of freedom once we have learned the word “freedom”). Therefore abstract words refer to more varied experiences than concrete words, and language might be necessary to keep all these (bodily) experiences together. Consistently general abstract word meanings rely more than concrete word meanings on the social experience of language. The novelty of this proposal basically consists in conceiving words not as mere signals of something, but also as “social tools” (Borghi and Cimatti, 2009, 2012).

If the core difference between concrete and abstract words is to be found in their peculiar acquisition mechanisms, simulating the acquisition of new concrete and abstract words should lead to findings symmetrical to the ones obtained with existing words. With the aim to verify this prediction Borghi and colleagues (Borghi, Flumini, Cimatti, Marocco, and Scorolli, 2011) performed four experiments in which they simulated the acquisition of new words. First, participants acquired new concepts by manipulating novel objects (concrete concepts) or by observing groups of objects interacting in novel ways (abstract concepts). Then they were provided with a new category label for each concept. To assess if the novel abstract and concrete words actually differed along the same dimensions as existing abstract and concrete ones, participants performed a feature production task, in which they were asked to list properties for each concept-label. Crucially the association score analyses showed that concrete words evoked more perceptual properties, as typically found with existing concrete words (Borghi and Caramelli, 2001). Participants were also asked to judge which one of two exemplars corresponded to a given (previously learnt) verbal label. Results on errors showed that it was more difficult to form abstract than concrete categories. Finally participants performed a color verification task (on the experienced objects / groups of objects) with manual or verbal responses. Interestingly, for abstract words microphone use was easier than keyboard use; symmetrical results were found for concrete words. Consistently with the predictions, (new) concrete words evoked more manual information, while (new) abstract words elicited more verbal information (Borghi and Cimatti, 2009; for similar results see also Scorolli, Granito, and Borghi, 2013).
New perspectives on language: language to act

Cultural and social features

As we have seen, a critical problem embodied views face concerns the issue of how abstract words are represented. According to recent proposals, both sensorimotor and linguistic information play a role in conceptual representation (Barsalou et al., 2008; Borghi and Cimatti, 2009, 2010). The WAT proposal advances the idea that language is not only affected by our previous experience, but it also actively shapes speakers’ perceptions of the world (see also Boroditsky and Ramscar, 2002; Gentner, 2003; compare the Whorfian hypothesis: Whorf, 1956).

In keeping with this hypothesis, some recent evidence suggests that different languages can differently carve up our experience. For example, in the behavioral study, Scorolli, Binkofski, et al. (2011) found differences between speakers of two languages that differ grammatically: German participants were faster with abstract verbs while Italian ones were slower with the same kind of verbs (regardless of the kind of noun that preceded or followed the verb). In a similar vein, Bergen and Chan Lau (2012) tested native speakers of English, Mandarin Chinese from mainland China, and Mandarin Chinese from Taiwan: the writing direction for the first two groups is left to right and then top to bottom; for the third group, writing direction is predominantly top to bottom and then right to left. When asked to order cards describing different stages of temporal development, participants basically replicated differences in the native writing systems, supporting the idea that the axis used to represent time in terms of space is affected by language experience. In a related study Siyanova-Chanturia, Pesciarelli, and Cacciari (2012) tested people belonging to the same culture (Italian) to verify if gender stereotypes conveyed by language affected the processing of third-person pronouns. Results on event-related brain potentials confirmed their hypothesis; importantly they also found that female (e.g. teacher) and male (e.g. driver) stereotypes affected the processing of pronouns differently.

The attention on the social aspects of language has suggested a focus on not only cultural differences, but also on the effects of having another person in the scene (Knoblich and Sebanz, 2008; Goldman and de Vignemont, 2009), with or without a collaborative attitude (see Tomasello, 2009; Scorolli, Miatton, Wheaton, and Borghi, 2012; Becchio et al., 2012). Gianelli, Scorolli, and Borghi (2013) investigated whether the reach-to-grasp movement towards an object was influenced by the presence of a second person. This person could be either a friend or a non-friend, and was either invisible (behind) or located in different positions with respect to the agent and to the object. While only the agent performed the physical action, both the participants could be (in turn) the speaker. Before movement initiation, the speaker pronounced a sentence referring to her own action (e.g. “I grasp”) or to the same action performed by the other (e.g. “You grasp”). Interestingly, the agent’s grasping component of the movement was influenced by the kind of relationship between her and the other person, as well as by the relative physical position. Most crucially, the overall reaching time showed an interaction between the speaker and the used pronoun: participants reached for the object more quickly when the other spoke, particularly if she used the “I” pronoun.

This evidence supports the idea of two forms of social perspective: the physical perspective, conveyed by both the other’s body position and the distance from the object, and the perspective induced by language. Speaking, and particularly assuming the first-person perspective (linguistically conveyed by the first-person pronoun), evokes a potential action, consistently with the claim that words can be intended as kinds of actions. In a similar vein, Scorolli and co-authors (Scorolli, Daprati, Nico, and Borghi, 2011) showed that effective use of words, like
an effective use of physical external auxiliaries such as tools (e.g. Farnè, Iriki, and Lădavas, 2005), can determine an extension of the peripersonal space (Borghi and Cimatti, 2009; 2010; Clark, 2008).

**Extended mind**

Nowadays we badly need further evidence to complement this big picture, particularly empirical data on normal language development (see Bergelson and Swingley, 2013) as well as on developmental disorders (for a good review see Marcus and Rabagliati, 2006).

Nevertheless evidence collected so far clearly argues in favor of (1) a mind that is not “brainbound” (e.g. Clark, 2008; Noë, 2009; Wilson and Golonka, 2013), but distributed also beyond body’s edges; (2) a language that cannot be conceived only in its referential aspects, but also in its social and public features (e.g. Borghi and Cimatti, 2009); (3) body edges that are not static, but that can be plastically rearranged (e.g. Tsakiris, 2010; Longo and Serino, 2012). This suggests that an embodied-grounded view of cognition should be integrated with the extended mind view (Clark and Chalmers, 1998): the combination of these two perspectives promises to shed new light not only on language processing, but also on the actual potentialities of language (Borghi, Scorolli, Caligiore, Baldassarre, and Tummolini, 2013).

**Notes**

1 See Chapter 4 on “Complex Dynamical System and Embodiment”, by Michael J. Richardson and Anthony Chemero.
2 For an alternative perspective see also Tomasino and Rumiati, 2013.
3 As to “Neuroscientific Bases of Embodiment”, see Chapter 5, by Laila Craighero.
4 I.e. properties that depend on object’s relationship with the agent and/or other objects.
5 For recent replication studies see Zwaan and Pecher, 2012.
6 For controversial results as to color see Zwaan and Pecher, 2012.
7 For a detailed discussion on interference and facilitation effects see Borrego and Kaschak, 2006; Connell and Lynott, 2012.
8 It is worth noting that both modalities seem to use similar metrics, as found (2012) by contrasting the visual modality (e.g. “You are looking at the milk bottle across the supermarket” vs. “You are looking at the milk bottle in the fridge”) with the auditory one (e.g. “Someone fires a handgun in the distance” vs. “Right next to you, someone fires a handgun”: Winter and Bergen, 2012).
9 For studies measuring also kinematic parameters of the motor responses see for example Gentilucci et al., 2000; Gentilucci, 2003; Scorolli, Borghi, and Glenberg, 2009.
10 For a discussion on embodied models explaining interference and facilitation effects see Borghi, Caligiore, and Scorolli, 2010; Borghi, Gianelli, and Scorolli, 2010.
11 See also the dual-coding theory, Paivio, 1986.
12 For a non-embodied version of this view see Dove, 2009.
13 For a close examination of the mode of acquisition construct see Wauters, Tellings, van Bon, and van Haafken, 2003; as to “Concepts Acquisition”, see also Chapter 11, by Daniel Casasanto.
14 For an in-depth discussion, see the Chapter 20 on “Cultural Differences”, by Tamer Soliman and Art Glenberg.
16 See Chapter 14 on “Language Acquisition”, by Chen Yu.

**References**


Whorf, B. L. (1956). Language, thought, and reality; selected writings, Cambridge, Technology Press of Massachusetts Institute of Technology.


