Making Sense of Embodiment: Simulation Theories and the Sharing of Neural Circuitry Between Sensorimotor and Cognitive Processes

Henrik Svensson (svensson@ida.his.se)
University of Skövde, School of Humanities and Informatics
Box 408, 54128 Skövde, Sweden

Tom Ziemke (tom@ida.his.se)
University of Skövde, School of Humanities and Informatics
Box 408, 54128 Skövde, Sweden

Abstract
Although an increasing number of cognitive scientists are convinced that cognition is embodied, there still is relatively little agreement on what exactly that means. Notions of what it actually means for a cognizer to be embodied range from simplistic ones such as ‘being physical’ or ‘interacting with an environment’ to more demanding ones that consider a particular morphology or a living body prerequisites for embodied cognition. Based on experimental evidence from a range of disciplines, we argue that one of the keys to understanding the embodiment of cognition is the sharing of neural mechanisms between sensorimotor processes and higher-level cognitive processes. The latter are argued to be embodied in the sense that they make use of (partial) simulations or emulations of sensorimotor processes through the re-activation of neural circuitry also active in bodily perception and action.

Introduction
Although an increasing number of cognitive scientists are convinced that cognition is embodied (e.g. Varela et al., 1991; Clancey, 1997; Clark, 1997; Lakoff & Johnson, 1999; Ziemke, 2002), there still is relatively little agreement on what exactly that means. Notions of what it actually means for a cognizer to be embodied range from simplistic ones such as ‘being physical’ or ‘interacting with an environment’ to more demanding ones that consider a particular morphology or a living body prerequisites for embodied cognition (cf., e.g., Chrisley & Ziemke, 2002; Wilson, 2002; Anderson, 2003; Ziemke, 2003).

This lack of agreement or coherence, after more than a decade of research on embodied cognition, has unfortunate consequences. Firstly, critics commonly argue that the only thing that embodied cognitive theories have in common is in fact the rejection of traditional, computationalist and supposedly disembodied cognitive science. Secondly, there is a certain trivialization of embodiment, not least among many AI researchers who consider as embodied any physical system, or in fact any agent that interacts with some environment, such that the distinction between computationalist and embodied cognitive theories disappears since, in some sense, all systems are embodied, and thus cognitive science has always been about embodied cognition (Chrisley & Ziemke, 2002). Thirdly, there is the ‘misunderstanding’ that perhaps embodiment is only relevant to sensorimotor processes directly involving the body in perception and action, while higher-level cognition might very well be computational in the traditional sense and only dependent on the body in the sense that mental representations ultimately need to be grounded in sensorimotor interaction with the physical environment.

This paper, on the other hand, argues that one of the keys to understanding the embodiment of cognition, in an important, non-trivial sense, is to understand the sharing of neural mechanisms between sensorimotor processes and higher-level cognitive processes. Based on experimental evidence from a range of disciplines, we argue that many, if not all, higher-level cognitive processes are body-based in the sense that they make use of (partial) simulations or emulations of sensorimotor processes through the re-activation of neural circuitry that is also active in bodily perception and action (cf. Clark & Grush, 1999; Grush, in press; Hesslow, 2002). As Barsalou et al. (2003) put it, the main point is that “simulations of bodily states in modality specific brain areas may often be the extent to which embodiment is realized”.

The next section elaborates the key idea of this paper, i.e., cognition as body- and simulation-based in the above sense, in more detail. In the following sections then supporting empirical evidence from a range of disciplines is presented. The final section then presents a brief summary as well as some open questions and directions for future work.

Cognition as body-based simulation
The idea that even higher-level cognitive processes are in a strong sense grounded in bodily activity and experience is, of course, hardly new, but was developed already in the 1980s, most influentially by Maturana and Varela (1980, 1987) from a neurobiological perspective, and by Lakoff and Johnson (1980, 1999) from a linguistic perspective. Lakoff (1988) summarized the basic idea as follows:

Meaningful conceptual structures arise from two sources: (1) from the structured nature of bodily and social experience and (2) from our innate capacity to imaginatively project from certain well-structured aspects of bodily and interactional experience to abstract conceptual structures.
Back in the 1980s, however, relatively little was known about exactly how such an imaginative projection from bodily experience to abstract concepts might work. In recent years more detailed accounts of how the sensorimotor structures of the brain are involved in cognition have been developed in several disciplines, often taking into account data from neurophysiological and neuroimaging studies. These accounts show that the traditional strong division between perception and action, as well as between sensorimotor and cognitive processes, needs to be revised.

A particular kind of “embodiment” theory that has emerged in different contexts are so-called emulation or simulation theories\(^1\) (e.g., Barsalou et al., 2003; Decety, 1996; Frith & Dolan, 1996; Grush, in press; Hesselow, 2002; Jeannerod, 1994, 2001). The basic idea is that neural structures that are responsible for action and/or perception are also used in the performance of various cognitive tasks. As Hesselow (2002) pointed out, this idea is not entirely new; e.g. Alexander Bain suggested in 1896 that thinking is basically a covert form of behavior that does not activate the body and thus remains invisible to external observers. Today simulation theories, based partly on data from neuroscience, can further clarify the possible role of simulation in cognition, thus explaining in a more concrete way than before the embodiment of cognition.

One of the more comprehensive descriptions of the idea has been presented by Grush (in press; see also Clark & Grush, 1999). Based on the control theoretic concept of forward models (emulators), previously used to account for motor control (e.g., Wolpert & Kawato, 1998), Grush developed an emulation theory for several types of cognitive processes, including perception, imagery, reasoning and language. In a nutshell, he argued that emulation circuits are able to calculate a forward mapping from control signals to the (anticipated) consequences of executing the control command. For example, in goal-directed hand movements the brain has to plan parts of the movement before it starts. To achieve a smooth and accurate movement proprioceptive/kinesthetic (and sometimes visual) feedback is necessary, but sensory feedback per se is too slow to affect control appropriately (Desmurget & Grafton, 2000). The ‘solution’ is an emulator/forward model that can predict the sensory feedback resulting from executing a particular motor command.\(^2\)

The following section summarizes a number of the many empirical studies that support the idea that cognition is body-based, especially as predicted by simulation theories.

---

\(^{1}\) The terms simulation and emulation are used somewhat interchangeably in this paper, as in much of the literature, but it should be noted that they are sometimes used differently (e.g., Grush, in press).

\(^{2}\) According to Blakemore, Frith and Wolpert (1999), this is also why it is not so easy to tickle oneself; the forward model produces predicted sensory feedback that ‘prepares’ the agent.

---

**Empirical Evidence**

Several sources of evidence support the basic tenet of the simulation account, viz., that perceptual and motor areas of the brain can be covertly activated either separately or in sequence for use in cognitive processes. In particular, several studies have indicated that there are extensive similarities between the neural structures activated during preparation (and execution) of an action and mentally simulating an action (i.e., motor imagery), as well as between visual perception and visual imagery. The similarities are so striking that some have argued that internally activated actions and perceptions are the same as overt ones, except that the overt execution or sensory input is missing (e.g., Hesselow, 2002; Jeannerod, 2001).

The following subsections review some of the empirical evidence that suggest that sensorimotor structures of the brain are deeply involved in the generation of cognitive phenomena, such as imagery and problem solving.

**Motor imagery**

There is an extensive literature on the neural and behavioral similarities between actions and motor imagery (e.g., Decety, Jeannerod, & Prablanc, 1989; Jeannerod & Decety, 1995; Jeannerod & Frak, 1999; for reviews see Decety, 1996, 2002; Jeannerod, 1994, 2001).\(^3\)

Motor imagery is the recreation of an experience of actually performing an action, e.g., the person should feel as if he/she was actually walking (Decety, 1996; Jeannerod, 1994). The evidence cited in support for the equivalence of performing an action and simulating an action comes mainly from three different sources: mental chronometry, autonomic responses, and measurements of brain activity.\(^4\)

In mental chronometry experiments, it has been found that the time to mentally execute actions closely corresponds to the time it takes to actually perform them (Jeannerod & Frak, 1999). For example, Decety and Jeannerod (1996) found that Fitt’s law (i.e., the finding that execution times increase with task difficulty) also holds for motor imagery. Decety et al. (1989) compared the durations of walking towards targets (with blindfolds) placed at different distances and mental simulation of walking to the same targets. In both conditions times were found to increase with the distance covered.

Autonomic responses, such as the adaptation of heart and respiratory rates, which are beyond voluntary control have been shown to be activated by motor imagery to an extent proportional to that of actually performing the action, and as a function of mental and actual effort (Decety, 1996; Jeannerod, 1994; Jeannerod & Decety, 1995).

---

\(^{3}\) There are also similarities between actions and other cognitive tasks, such as observing an action and prospective action judgments, which differ from motor imagery in that they do not produce a conscious motor image of performing an action and the brain activation is not as similar to overt actions as in motor imagery (cf., Jeannerod, 2001).

\(^{4}\) Further evidence comes from effects on action performance after mental training using motor imagery and also from lesion studies.
Since the first study that investigated motor imagery using regional cerebral blood flow (rCBF) to indicate active brain areas (Ingvar & Philipson, 1977) there have been many neuroimaging experiments that confirm the first study’s indication that similar brain areas are activated during action and motor imagery. The general conclusion is that there is a functional equivalence between performing an action and mentally simulating it (Decety, 2002; Jeannerod, 2001).

Together, these three types of evidence point to an explanation of motor imagery that implicates sensorimotor structures. Although the focus above has been on actions, it should not be forgotten that the motor system also integrates sensory information when planning and executing an action (e.g., Grush, in press; cf. Desmurget & Grafton, 2000; Jeannerod, 1997). Thus, simulating an action might also involve an emulator mechanism that predicts the sensory feedback that would have resulted from the executed action (Decety, 2002; Grush, in press; Jeannerod, 1997, 2001).

### Mental visual imagery

The discussion of motor imagery can be extended to the visual modality in that the same type of studies report that perceptual structures can and are internally reactivated when, e.g., visually recreating a previous perception. Many studies in cognitive psychology have found similarities between visual perception and visual imagery (Farah, 1988; Finke, 1989). For example, Shepard and Metzler (1971) performed a number of mental chronometry type experiments, where they compared the concrete manipulation of physical objects and corresponding manipulations performed mentally. In their experiments, the time between the two conditions was closely correlated, which suggests that mental imagery uses the same mechanisms as the visual system. Although alternative explanations are difficult to rule out, neuropsychological and neuroimaging studies offer more conclusive evidence of the involvement of sensorimotor areas of the brain in mental visual imagery (Farah, 1988; Hesslow, 2002).

Note that these findings do not necessarily influence debates on representational format, since both perception and imagery may be said to use the same format (Block, 1983). On the other hand, explaining cognition as reactivation of sensorimotor structures does not (at least in some cases) rely on the computer metaphor of symbol manipulation, and thus may offer a novel view that does not see the vehicle and the content of representations as separate entities but as constitutive of each other (Gallese, 2003; cf. Dreyfus, 2002; Thomas, 1999). Gallese argued that canonical neurons in the monkey brain illustrate how the interaction between an agent and its environment provides an example of such representations.

### Canonical neurons

In the macaque monkey, neurons located in the rostral part of the inferior premotor cortex (area F5) of the monkey brain discharge during goal directed movements, such as grasping, holding, or tearing. However, they do not respond to similar movements, but only actions that have the same “meaning” (di Pellegrino et al., 1992; Rizzolatti et al., 1996; Rizzolatti et al., 2002), which is why they are often interpreted as internal representations of actions, rather than motor or movement commands (Jeannerod, 1994; Rizzolatti et al., 1996; Rizzolatti et al., 2002). Gallese (2003) emphasized seeing them as coding not physical parameters of movement, but a relationship between agent and object.

Some of the neurons in area F5, so-called canonical neurons, also have sensory properties and discharge both during the action they code and when an object that affords that action in the Gibsonian sense is perceived. Canonical neurons have a strict congruence between the type of grasping action and the size or shape of the object they respond to (Gallese, 2003). This implies that they implement affordances, e.g., code things that are graspable-in-a-certain-way, specifying not only perceptual and action aspects but a particular relationship between agent and environment (cf. Gallese, 2003, cf. also Dreyfus, 2002).

### Problem solving using covert perception and action

Instead of advertizing to symbol manipulation, a flexible inner world, in which an agent might try out possible action sequences, can be explained by internal activation of perceptions and actions (Clark & Grush, 1999, Grush, in press; Hesslow, 2002). That means, an agent can sustain such an inner world by letting an internally activated action elicit through an anticipatory mechanism internally generated perceptions that would be the likely result of executing that action in the particular external situation. As discussed above, the internal activation of sensorimotor structures is well supported, but the neural underpinnings of the anticipation mechanism are still an open issue. Some suggest the involvement of the cerebellum (e.g., Hesslow, 2002).

Some support for the involvement of this type of simulation of behavioral chains comes from the problem solving and planning involved in the Tower of London (ToL) problem (Shalllce, 1982). Dagher et al. (1999) found that planning and problem solving activated higher motor areas (premotor cortex, prefrontal cortex) and the basal ganglia, and that they seemed to interact with visual and posterior parietal areas (cf. Schall et al., 2003). This gives some support to the idea that the subjects solved the problem by simulating the action of moving one ball to another location through the use of reactivated perceptions and actions (Hesslow, 2002).
Social cognition and language
The use of internally reactivated sensorimotor structures has also been suggested to play a crucial role in social cognition, especially emotive states (cf. Barsalou et al., 2003; Nielsen, 2002). The evidence that is reported in this subsection emphasizes that perceptual and motor processes are not different in nature at the neural and behavioral level, but seem to be intimately linked in social cognition possibly through simulation mechanisms.

Embodiment effects Barsalou et al. (2003) argued that there are at least four types of well known phenomena in social cognition which can be explained by simulation of bodily states. The types of effects they mention are: a) perceived social stimuli produce bodily states, b) perceiving bodily states produce bodily mimicry, c) bodily states can produce and affect emotion states, and d) compatibility between bodily states and emotional states increases performance (see also Nielsen, 2002). These effects seem to arise automatically without any conscious mediating knowledge structures (Barsalou et al., 2003; Nielsen, 2002).

Firstly, social stimuli do not only produce cognitive responses, but at the same time automatically and unconsciously cause bodily responses (e.g., movement patterns, facial expressions). For example, subjects induced with an elderly stereotype (e.g., by watching elderly people) perform in a manner more similar to the elderly stereotype than control subjects (Barsalou et al., 2003).

Secondly, bodily responses sometimes are the same or similar to the eliciting social stimuli as in the many cases of facial mimicry. For example, watching somebody yawning often causes oneself to yawn too (cf. Barsalou et al., 2003).

Thirdly, bodily states can directly induce effects on affective states (Barsalou et al., 2003). One such effect is the facial feedback hypothesis, which states that a person’s own facial expressions can (either directly from the brain areas responsible for the facial expression or through processing of proprioceptive feedback) produce or modify his/her experience of the emotional state (Nielsen, 2002).

Finally, when there is a mismatch between bodily state and affective state cognitive performance is degraded. Barsalou et al. (2003) argued that this is perhaps the most important effect because it indicates that embodied states are directly involved in higher cognition. That is, the affective states are thought to involve sensorimotor states and when there is an incompatible bodily state co-present it produces a competing sensorimotor state, which reduces performance. For example, Wells and Petty (1980) showed that head movements were faster when compatible with the message (e.g., nodding vertically to an agreeable message) than when incompatible (cf. Barsalou et al., 2003).

The findings discussed in this section suggest that bodily states are involved in social cognition and that they might constitute the very foundations of the particular social cognitive phenomena in question. An example of how perception, action, and social cognition come together at the level of single neurons is so-called mirror neurons in macaque monkeys (Decety & Sommerville, 2003).

Observation execution matching system Besides canonical neurons, area F5 of the monkey brain contains so called mirror neurons, which are neurons with sensory properties that become activated both when performing a specific action and when observing the same goal-directed hand (and mouth) movements of an experimenter (di Pellegrino et al., 1992; Rizzolotti et al., 1996). Mirror neurons provide a key example of sensorimotor brain structures also involved in (social) cognitive processes.

Although different hypotheses exist, many of the theories of the function of mirror neurons emphasize their role in social cognition (e.g., Decety & Chaminade, 2003; Gallese & Goldman, 1998; Rizzolotti & Arbib, 1998; Rizzolotti et al. 2002). These researchers acknowledge that area F5 and mirror neurons can be interpreted as a kind of observation-execution mechanism or resonance mechanism, which links the observed actions to actual actions of the subject’s own behavioral repertoire. That is, it enables the monkey to understand the meaning of the observed action. Thus, mirror neurons can be interpreted as representations of actions, used both for performing and understanding actions (e.g., Rizzolotti et al., 1996; Rizzolotti & Arbib, 1998). 6

Gallese and Goldman (1998) hypothesized that mirror neurons might be a basic mechanism necessary for “mind-reading”, i.e., attributing mental states to others. They further argued that such mechanisms can explain how an agent determines what mental states of another agent have already occurred. When mirror neurons are externally activated by observing a target agent executing an action (allowing the subject to evaluate the meaning of the other’s action), the subject knows (visually) that the observed target has already occurred. When mirror neurons are externally activated by observing a target agent executing an action (allowing the subject to evaluate the meaning of the other’s action), the subject knows (visually) that the observed target is currently performing this very action and thereby “tags” the “experienced” action as belonging to the target. However, how the subject can distinguish its own actions from those performed by others is relatively unknown (cf. Blakemore, Wolpert & Frith, 2002).

Language Some researchers have argued that conceptualization and language understanding cannot be achieved through the manipulation of amodal, arbitrary symbols alone, but has to be grounded in bodily interaction with the environment. In particular, Glenberg and Kaschak (2003) have outlined an explanation of language in line with the idea of cognition as body-based simulation as expressed in this paper, suggesting that language is partly achieved through the same neural structures used to plan and guide action.

---

6 For practical and ethical reasons it is so far not possible to investigate the existence of mirror neurons (and canonical neurons) at the single neuron level in humans. However, many researchers have presented strong arguments for the existence of a similar system in humans (e.g., Fadiga et al., 1995; Grafton et al., 1996; Rizzolotti, & Arbib, 1998; Rizzolotti et al., 1996).
Under the heading of the indexical hypothesis they developed an account of language comprehension partly based on simulation of action. They argued that the meaning of a sentence is achieved by a process that indexes words to perceptual symbols which in turn retrieves the available affordances in the situation and determines their relevance through the particular sentence construction. Thus, the understanding of a sentence is essentially achieved through a simulation of action using the same neural systems active in overt behavior.

An empirical result that supports the close coupling between language and action is the “action-sentence compatibility effect” (Glenberg & Kaschak, 2002). It was found that the sensibility of a sentence is modified by physical actions. Reaction times increased when subjects read “toward sentences” that implied action toward the reader, such as “Open the drawer” and had to give the answer through an incongruent action, i.e., moving the hand away from the body. Conversely, when subjects answered through an action congruent with the sentence, reaction times decreased. Glenberg and Kaschak interpreted the result as indicating that understanding a sentence is dependent on structures usually used for action. Readers interested in more comprehensive reviews of the coupling between language and action/perception are referred to Glenberg and Kaschak (2003) or Zwaan (2004).

Summary and conclusions

This paper has presented an emerging framework of simulation based on terminology and ideas from control theory and data from neurophysiological and neuroimaging studies that explain higher-level cognitive processes as at least partly based on reactivations of sensorimotor structures of the brain. By reactivating mechanisms used in perception and action together with a predictive mechanism a flexible inner world emerges that can be used for many different higher-level cognitive tasks (cf. Grush, in press; Hesslow, 2002). Crucial to the embodiment of cognition, according to this account, is perhaps not so much the physical nature of a cognizer’s body, or its interaction with the environment as such, but the relation between sensorimotor and higher-level cognitive processes, more specifically, the way that the latter are fundamentally based on and rooted in the former.

Although corroborating evidence comes from several disciplines, the simulation account is not yet a well established or coherent theory of cognition in general, and there are many questions still to be answered. For example, in current accounts it is unclear exactly what constitutes the difference between an executed, overt action and a simulated/imagined, covert one. Can this be accounted for in terms of simulation theories or are other, presumably higher-level, mechanisms required after all to selectively trigger one or the other?

Moreover, there is a level of granularity problem (cf. Meltzoff & Prinz, 2002) that seems to apply to many simulation accounts. That is, at what level of abstraction does the simulation occur? During imagery it seems that the simulation occurs on a low-level including very many of the aspects of actually perceiving or acting, as indicated by neuroimaging studies (e.g., Jeannerod, 2001), whereas in problem solving, as in the ToL task, more abstract aspects of actions may be employed, which might be supported by the finding that problem solving activity in the ToL seem to activate only higher motor centers, such as prefrontal and premotor cortex (Dagher et al., 1999). However, this is a speculative interpretation of the neuroimaging results.

The level of granularity is also an important issue in robotic models of simulation theories (e.g. Ziemke, Jirenhed & Hesslow, in press). Previous work in our lab has dealt with internal simulations at the lowest level, but current work also addresses simulation at more abstract levels of granularity. We believe that robotic models offer a fruitful approach to further investigating this and other open questions that need to be answered in the future development of simulation theories.

References

Decety, J. (2002). Is there such a thing as functional equivalence between imagined, observed, and executed action. In M. A. Meltzoff & W. Prinz (Eds.), The Imitative Mind. Cambridge: Cambridge University Press.


