Mind the Brain: The Potential of Basic Principles for Brain Function and Interaction

Daniel Sjölie and Lars-Erik Janlert

UMINF-13.04
ISSN-0348-0542

Umeå University
Department of Computing Science
SE-901 87 Umeå, Sweden

Mind the Brain: The Potential of Basic Principles for Brain Function and Interaction

Daniel Sjölie[a,*], Lars-Erik Janlert[a]

[a] Department of Computing Science
Umeå University
S-901 87 Umeå, Sweden

Abstract
The prevalence and complexity of human-computer interaction makes a general understanding of human cognition important in design and development. Knowledge of some basic, relatively simple, principles for human *brain function* can significantly help such understanding in the interdisciplinary field of research and development Human-Computer Interaction (HCI) where no one can be an expert at everything. This paper explains a few such principles, relates them to human-computer interaction, and illustrates their potential. Most of these ideas are not new, but wider appreciation of the potential power of basic principles is only recently emerging as a result of developments within cognitive neuroscience and information theory. The starting point in this paper is the concept of *mental simulation*. Important and useful properties of mental simulations are explained using basic principles such as the *free-energy principle*. These concepts and their properties are further related to HCI by drawing on similarities to the theoretical framework of *activity theory*. Activity theory is particularly helpful to relate simple but abstract principles to real world applications and larger contexts. Established use of activity theory as a theoretical framework for HCI also exemplifies how theory may benefit HCI in general. Briefly, two basic principles that permeate this perspective are: the need for new skills and knowledge to build upon and fit into what is already there (grounding) and the importance of predictions and prediction errors (simulation).

Keywords: HCI theory, concepts and models; brain function; activity theory; grounded cognition; mental simulation; the free-energy principle

[*] Corresponding author.
E-mail address: daniel@cs.umu.se
Telephone: +46 90 786 92 33
Fax: +46 90 786 61 26

1. Introduction

Today, we meet computers everywhere. The design space of human-computer interaction is vast, and growing. Relatively speaking, the constant is the human, not the computer or the evolving computer-augmented everyday environment. How humans adapt to changes in their environment, and what this implies for how the environment is best changed and managed, are questions that are fundamentally about human nature. A general understanding of humans has always been considered important in HCI, and as human-computer interaction gets more deeply integrated into the everyday environment it is becoming imperative for an ever-wider range of applications. It would appear that understanding how the human brain works and what brain science can tell us about how humans interact skillfully with the world and learn new interaction tricks is important, but practical application of such knowledge is often hampered by the need for expertise to digest the scientific literature. An essential point made in this paper is that basic principles of brain function can be made accessible also to non-experts, allowing practitioners to be aware of and put to practical use the most important aspects of the relevant theories without spending a prohibitive amount of time on becoming experts. Shared knowledge of basic principles is also helpful for communication within the interdisciplinary teams characteristic of HCI and interaction design. A general understanding of the human brain is particularly important for a number of emerging applications that directly involve the brain, such as brain-computer interfaces (BCIs) (Tan and Nijholt, 2010) and computerized cognitive training (Sjölie, 2011), and to interaction paradigms that target realistic human interaction, such as virtual reality (VR) and reality-based interaction (RBI) (Jacob et al., 2008). An example of how such an understanding may be further developed for a particular application is given by Sjölie (2012), who takes the same general principles of brain function that are described in this paper as a basis for discussions relating to the sense of presence.

Activity theory is a good example of how theory can be helpful in HCI and support design and development of human-computer interaction. Over the last decades activity theory has been applied in developing a theoretical framework for understanding and reasoning about human conditions for human-computer interaction (Kaptelinin and Nardi, 2006; Kuutti, 1996; Rogers, 2005; Wilson, 2009). By helping researchers make sense of their data and ask the right questions activity theory has been instrumental in the formulation of guidelines and recommendations for interaction design, as well as for communicating such results to practitioners (Halverson, 2002; Nardi, 1996). The basic principles discussed in this paper may provide similar benefits in combination with established theories, facilitating communication and understanding. Activity theory emphasizes interaction with the environment and human development as key aspects of human cognition (Holzman, 2006; Kaptelinin and Nardi, 2006; Kaptelinin et al., 1995), which makes it particularly suitable for making connections with the basic principles of brain function discussed in this paper.

The rest of this paper is laid out as follows. In the next section the concept of mental simulation is introduced and related to basic principles of brain function and to activity theory. This is followed by an account of the basic principles, primarily the free-energy principle, and plausible neural implementations of mental simulations based on such principles. In the last sections of the paper the relations to existing theory in HCI and potential practical implications for HCI are briefly reviewed.

2. Mental simulations

'Mental simulation' and 'mental model' are expressions that have been used in many theories, some of which are discussed in a later section. At this point, however, *mental simulation* should be understood as a very general concept that is central for understanding how the human mind works, as described within the framework of *grounded cognition* (Barsalou, 2008). According to grounded cognition all cognitive capabilities need to be *grounded* in, i.e., based on and connected to, our bodies, our senses, and, ultimately, our external reality. Such grounding can be related to the focus on interaction with reality and internalization of activities within activity theory. This connection is developed further below but readers familiar with activity theory may wish to keep it in mind.

One of the fundamental ideas of grounded cognition is that all cognition is grounded in the brain's modal systems for perception, action and introspection (Barsalou, 2008; Barsalou et al., 2003). Mental representations and the corresponding meanings of human thoughts are stored in the areas of the brain that are directly related to perception and previous experience of interaction with associated phenomena. Thoughts are based on human interaction with the real world. If someone thinks about "ice cream" this triggers activity in areas of the brain related to tasting, seeing and touching ice cream, that is, in areas related to internalized aspects of the real "ice cream" object. In this context, the concept of mental simulation is a generalization of mental imagery (consciously imagining some experience "in the head") to include unconscious and spontaneous triggering of simulations in the brain. As with the ice cream example, the idea is that when someone thinks about something (anything) this corresponds to triggering a simulation of how this phenomenon might be experienced and how one could interact with it. These simulations are based on memories of past experience that are recombined and reenacted flexibly depending on context. The simulated experience that results when thinking of an ice cream is not the same each time. Exactly what this thought entails depends on the full context, including bodily senses, needs and emotions, such as hunger or depression.

In distinction to such concepts as mental imagery or mental model the concept of mental simulation has a strong emphasis on simulations as *running*. That is, the simulations should not be thought of (primarily) as data structures, rules, or stored snapshots of momentary experiences, but as something that works in essentially the same way as that which is simulated; something that can be interacted with. A mental simulation of a computer keyboard, for instance, is usefully imagined as an internalized version of a real keyboard that can be handled, perceived and interacted with. The rest of the brain, and other mental simulations, can access this simulated keyboard and can rely on the resulting predictions of how a real keyboard would respond to interaction. This illustrates how grounding and predictions work together to create a rich description of human cognition. Higher order skills, like fast typing on a keyboard, are grounded in and rely on predictions of how a real keyboard would behave – predictions that can in turn be described as internalizations developed through interaction with real keyboards.

Simulations and activity theory

Mental simulations can be related to activity theory in a number of ways. Most obvious is the connection to *internalization*. In activity theory, internalization is the developmental process whereby existing physical objects and activities can become internal and give rise to mental objects representing potential and desirable outcomes of activities in the real world. It is towards these potential outcomes, this vision of an object existing in a certain form, that activities are directed (Leontiev, 1978). Imagining oneself as the winner of an Olympic medal or as the owner of a new house are examples of such driving objectives and it is by considering which objects any part of an activity is oriented towards that it is possible to identify and separate activities.

These mental objects can be thought of as corresponding to mental simulations, and the grounded cognition thesis that mental simulations are the foundation of human cognition has similarities with the principle of *object-orientedness* in activity theory. A particularly important similarity is that both mental simulations and mental objects are grounded in interaction with reality. It is a fundamental tenet of activity theory that the human mind (or, as it is often formulated, consciousness) is directly dependent on human activity in the real world (Kaptelinin, 1996). “You are what you do”, and without action, without interaction with the environment, there can be no mind, no consciousness and no cognition. Grounded mental simulations also lend themselves well to a *developmental perspective*, another key element of activity theory. The development of grounded higher-level simulations depends on existing lower-level simulations; the developmental potential of a person, what they can easily learn, depends critically on what they already know, which corresponds to what mental simulations they already have internalized. The dependencies and interplay between higher and lower level simulations can also be related to the two remaining main principles of activity theory: *mediation* and *hierarchies*. The connections between mental simulations and activity theory, and the relation to *externalization*, are further developed in section 4.

Far from aiming to give a full description of activity theory, the point here is to identify connections with grounded cognition and mental simulations that can help anyone familiar with either activity theory or grounded cognition to start to understand one in terms of the other. There are important similarities; similarities that can be described in general terms and that can be related to basic principles.

Evidential support for simulations in the brain

Mental simulation has been related to brain function in several ways, perhaps most famously in connection to mirror neurons and the simulation theory for social cognition (Gallese and Goldman, 1998; Rizzolatti and Fabbri-Destro, 2008). Mirror neurons were first discovered in macaque monkeys as a set of neurons that fired consistently both when the monkey executed an action and when it observed another monkey performing the same action (Pellegrino et al., 1992; Rizzolatti et al., 1996). Since then much research has been conducted to investigate this phenomenon including studies on humans. The first study demonstrating a mirror system in humans is worth mentioning. By using Transcranial Magnetic Stimulation (TMS) Fadiga et al. (1995) could measure how the motor evoked potentials (MEPs) induced by TMS were affected by watching other people move: they found marked increases in MEPs in precisely those muscles that subjects would use when performing the observed movements. The fact that the same motor circuits are activated in the brain when observing someone else performing an action as would be active when self performing the action can be directly related to the unity of perception, thought (or consciousness) and action in the world, central to activity theory (Kaptelinin and Nardi, 2006).

Another study on monkeys also speaks strongly to the connection between mental simulations and brain function. By recording the activity of single neurons in the inferior parietal lobule (IPL) it was shown that simple motor acts such as grasping were organized in chains corresponding to their place in composite actions including such acts (Fogassi et al., 2005). Interestingly, it was shown that these chains may be activated by the observation of the very first act in the chain, thus giving a direct measure of how a sequence of subsequent acts are triggered in the brain and essentially capturing the activation of a mental simulation "in the act".

Finally, the anatomy and physiology of connections in the brain (at least in the cortex) provide strong clues about general principles of brain function (Friston, 2005). The hierarchical organization of the brain fits well into functional principles about the grounding of higher functions in lower levels, and the importance of feedback, from higher to lower, suggests how high-level simulations and predictions can affect and drive lower levels (Friston, 2005; Hawkins, 2005). We return to the importance of hierarchies and feedback in section 3.

General simulations

The above conception of mental simulation specifically dispenses with any direct dependence on consciousness for mental simulations, focusing instead on the role of mental simulations as a basis for cognition as a whole, including the unconscious, spontaneous reenactment of previous experience to guide future action. This lack of dependence on human consciousness, replaced by a connection to basic neuroscience, including research on animals, marks a difference from many theories of cognition, including activity theory. Mental simulations are basic functions, placed in-between the conscious subject and the external environment. They are constantly shaped by and involved in the shaping of both the conscious experience and the real world. It is not a matter of just "playing back" previous perceptual experiences: the experience in question concerns the past connections between perception, action, body state, etc., corresponding to the history of interactions with the world, in the context of internal state – history that builds plastic systems of mental simulations in the brain. The human mind remembers what has happened, how one has acted and what one has felt before in the context provided by perception over time. This provides the foundation for predictions based on simulations of what could happen or how one could interact with the world in any familiar situation.

Mental models

The concept of mental models will be familiar to many readers. There are two main differences between mental models and mental simulations: generality and grounding. Mental simulations are conceptually simpler and more general, i.e., more basic. There have been many different conceptualizations of mental models in the literature; most focus on the details of how specific mental models might work, not on general principles for how *all* mental models work, nor on how mental models develop in general (Barsalou, 2008). It has been remarked that "Talking about mental models can be a dangerous thing", because there are so many different versions and understandings of what mental models may be (O'Malley and Draper, 1992). Communication about and wider acceptance of mental models is made difficult since "everyone who uses the term is convinced that their use of it is the appropriate one and that other uses are based on misunderstanding of the term or adaptations of the concept" (Leiser, 1992). The mental simulations described in this paper are grounded in mathematical models and plausible theories of neural implementations, which firmly constrain descriptions and capabilities of such mental simulations. By relating mental simulations to basic principles it is easier to keep track of what they can be, and thus easier to communicate clearly about them.

3. Basic principles of brain function

Concepts such as mental simulations can be used to discuss internalized objects and processes in relation to brain function, and related neuroscientific research lend these ideas some credibility. Still lacking, however, are the details on how these functions might be implemented in the brain. It is by digging a little bit deeper into these issues that we may find practical principles that can guide real applications in HCI and interaction design. Relating the basic principles to the basic conditions for life and the mathematical models, showing how they are grounded in modern cognitive neuroscience, physics, and information theory, also serves to build confidence in their general validity and value.

The free-energy-principle

The most basic principle presented in this paper is the free-energy principle. It has recently been put forward as a possible universal theory for brain function with the potential to subsume and further explain a number of previously suggested theories (Friston, 2010). These theories include several descriptions in terms of neuronal implementation and computational models that satisfy the requirement of mental simulations to be grounded in the real world. The free-energy principle thus has the potential to provide explanations for many phenomena that are described but not explained by theoretical frameworks such as activity theory; and it introduces simple, derivative but still basic, principles that can be used to get a general understanding of the conditions of human cognition.

The free-energy principle is based on a mathematical formulation of how living systems must work in order to resist the fundamental tendency towards disorder in the universe. This formulation refers to the second law of thermodynamics and has its basis in information theory (Friston, 2010). The concept of free energy may be hard to grasp without a background in mathematics, but the core idea can be explained by considering the importance of *surprises*. For any living organism it is of critical importance to avoid surprises and to minimize the risks they represent. The free-energy principle describes how such minimization can actually work based on known laws of physics and information. True surprise cannot be calculated, but the free energy in a system is an upper bound on surprise, and it can be calculated. Several theories of global brain function build on this specific mathematical approach to surprise avoidance (Friston, 2010; Huang, 2008). For the purpose of the discussion in this paper the focus will be on the importance of predictions and prediction errors, corresponding to expectations and surprises. These terms are simpler and clearer as they do not depend on a subject having an expectation or being surprised. Prediction errors correspond to the surprises that need to be minimized in order to interact competently with the world and thrive.

The basic motivation for the free-energy principle is the requirement on all living organisms to avoid surprising states in order to stay alive (Friston and Stephan, 2007; Friston, 2010). The states in which any given organism can thrive are limited; a fish, for example, needs to be in water. Throughout their evolution, surviving organisms exist primarily in these favorable states and the resulting invariance of states enables the species to build expectations about which states will be beneficial over time. Surviving fishes can come to expect and depend on aspects of the environment that are related to being in water and they can begin to predict which states are good. This corresponds to the evolution of innate preferences (Friston, 2010) or the needs that drive all action according to activity theory. Deviations from the expected wholesome states should be avoided: the *prediction error*, the difference of the current state from the predicted, optimal state, should be *minimized*. The fish expects to be in water, its brain predicts that the environment will behave like water, and if a deviation from the predicted is detected it needs to be resolved. Put differently, the brain dreams of the perfect world and reacts to all differences between this “prediction” and the currently perceived reality by taking action, striving to satisfy the organism’s needs. This corresponds to the notion of objects in activity theory as visions of desirable outcomes towards which human activities are directed. Thinking of human needs as predictions related to an optimal environment also facilitates the integrated consideration of brain function and humans’ ultimate motivations in the context of HCI, similarly to how activity theory has often been used to analyze HCI activities in a wider context of human needs.

Implementation of simulations

Predictive coding is a theory of brain function closely related to the free-energy principle (Rao and Ballard, 1999). It has been used to describe how the mirror neuron system might actually be implemented in the brain (Kilner et al., 2007) and thus provides another tie between the notion of internalized objects/actions, and brain function. Predictive coding explains the mirror neuron system in terms of so-called *generative models*: a top-down activation of sensory and motor representations driven by high-level goals and intentions. A generative model corresponds to running a mental simulation, driven by a high-level intention and potentially including (dynamic) reenactment of both sensory experience and motor action associated with the intention. Running a model involves predicting the consequences of actions since they are simulated together with the expected sensory experience. Generative models can also be *inverted* to recognize patterns in sensory input and infer their higher-level causes, which correspond to more persistent phenomena in the world. An apple in the real world can be said to *cause* the perception of the apple; the ability to imagine the apple (the generative model) is directly related to the ability to recognize the apple. It is essentially the same model run in two different directions. When applied to the simulation of actions and the corresponding recognition of goals, the implementation of such models in the brain can explain behavior and measurements of the mirror neuron system, thus constituting a possible implementation of internalized objects and actions in the brain.

One biologically plausible suggestion for how generative models may be implemented in the brain rests on the use of empirical Bayes within the hierarchical structure of the brain (Friston, 2005). Bayesian inference is a statistical method used to calculate the probability of a hypothesis given an observation and prior expectations related to the observation and the hypothesis. For example, what is the probability that an engine sound is caused by a car depending on the expected distribution of cars, engine sounds and cars with engine sounds? In empirical Bayes the necessary data is learned through experience and in the hierarchical structure of the cortex different levels can represent different hypotheses and different contexts. Higher levels, tuned to more persistent phenomena, may provide empirical priors for Bayesian inference at lower levels. For example, the expectations about cars and engine sounds above are different depending on whether you are in a rural area or in a city, information that in turn is a (more persistent) hypothesis that is supported (or not) by observations over time. Such a hierarchy of contextual empirical Bayes expectations corresponds to a dynamic hierarchical model with a plausible mapping to the anatomy of the brain. A mapping like this makes it possible to reason about how measured brain activity in different brain areas is connected with internalized aspects of activities (Friston, 2005; George and Hawkins, 2009; Hawkins, 2005), connections which can be used, for example, to inform the interpretation of brain measurements while the user learns (through internalization) to interact with new or changing interaction environments. Hierarchical models of this kind (dynamically developed based on persistency through experience) also meet the requirement from the grounded cognition theory of cognition being grounded in the modalities. The information at different levels can be represented as recognition densities, i.e., probability density distributions describing the likelihood of different inputs, for which plausible neural implementations have been suggested (Friston, 2010; George and Hawkins, 2009). The recognition densities at any point are relative to the context provided by the level above: predictions made at each level depend on prior probabilities given from above, predictions which in their turn provide the contexts for predictions at the level below, leading to a chain of predictions all the way down through the hierarchy. Sensitivity to the current context is thus built into the mechanism and predictions made at a high level of the hierarchy can shape and direct predictions throughout the entire underlying hierarchy. For example, if you are at a dog show you are more likely to assume that an animal is a dog, at a high level, that a rear end appendage is a dog's tail or that a sudden sound is a dog barking, at a lower level, and that the visual and tactile details of an animal's back are consistent with the structure and color of dog hairs, at an even lower level.

Prediction errors

According to the account of brain function above most of the brain's activity concerns the resolution of prediction errors. If the input matches the most probable given the context from above, predictions are confirmed and the brain is apparently well prepared for what should come next, so not much brain activity is to be expected. Several well-known phenomena in brain measurements such as repetition suppression and the P300 signal in electroencephalography (EEG) (Friston, 2005) then become useful means of determining whether something is going according to plan or not, which can be of great value to BCI and related HCI developments.

The details of how brain function might be implemented according to the free-energy principle provide some further guidance on how learning and internalization works. Such details can, for example, be exploited to optimize interaction in computer applications for cognitive training or education. Most importantly, changes in the hierarchical model, and the corresponding plasticity of the brain, are driven by prediction errors (Den Ouden et al., 2008). Research on how humans acquire expertise has demonstrated the critical importance of challenging and unfamiliar interactions, which, of course, lead to prediction errors. It also emphasizes the importance of keeping such challenges at the right level (Ericsson and Charness, 1994). The size of a prediction error depends on the precision of the prediction, corresponding to the degree of confidence in the prediction. If a phenomenon is unfamiliar the brain cannot make any precise predictions: nothing is truly surprising (as there are no expectations to violate) and prediction errors remain small. Effective training needs to build on phenomena and situations that are relatively familiar, in order to enable predictions with enough precision to risk significant prediction errors. The importance of pre-existing skill for determining the level at which learning is optimal has been recognized by activity theory, where this level is known as the zone of proximal development (Kaptelinin and Nardi, 2006).

It should be noted that the relationship between prediction, precision and prediction errors exists simultaneously on all levels of the hierarchy and that the only information going upward in the hierarchy is reports of prediction errors from below. The fact that humans are supplied with a continuous stream of information from the external world reflects the imperfection of all predictions. There are always some prediction errors. In this context the effect of *attention* can be described as a modulation of the precision of certain predictions. By directing attention to a phenomenon the precision of prior expectations is increased, giving rise to more prediction errors, resulting in more information being fed upwards in the hierarchy, potentially reaching higher levels. Among other things, this may explain how humans can bring details from perception into consciousness by directing attention to them.

Another aspect of the free-energy principle of specific interest for the considerations in this paper is the critical importance of action for minimizing prediction errors. Humans are, like all biological entities, open systems in exchange with their environment. The environment acts on the organism, giving rise to sensory input, and the organism acts on the environment, changing its states. Biological entities need to be able to control their relation to the environment in order to avoid damaging surprises (Friston and Stephan, 2007). The better they are at controlling the environment the greater their ability to construct the environment to be predictable, understandable and safely distant from dangerous surprises. Thus humans act in the long run to make perceptions (experiences) more predictable and in the short run to make predictions come true. We act in accordance with our expectations. This serves the dual purpose of changing the environment to match our expectations when possible and efficiently detecting deviations from those expectations when necessary. Illustrative examples can be found in relation to active sensing, that is, the active control of sensory organs based on predictions. According to studies on eye movements, one tends to move the gaze in specific patterns depending on what one is looking at, checking particular predictions based on what one expects to see. When humans look at a face they tend to look first at the eyes, the nose and the mouth (Barton et al., 2006), and when they look at scenes they tend to examine doors closely but almost never look at the ceilings (Richardson, 2010). This reflects that predictions focus on specific aspects of the environment. In short, action and perception is tightly interwoven based on human experience of interacting with our environment and human desires (rooted in needs) for which predictions one wants to come true.

Summary of the basic principles

There are several layers to the descriptions above, and it might not be immediately clear what should be considered as basic principles. For use in HCI and interaction design, the principles should be basic in the sense of providing a useful common ground that can be expected to be stable over a long time, while at the same time being easy to keep in mind. The most general principle presented here is the free-energy principle, but it may be a bit too technical and abstract for the present purpose. As an alternative, one may take as basic principles 1) prediction and simulating how something may be is the basis for cognition, and

2) everything humans learn must fit into existing structures that are ultimately grounded in reality. These principles can be directly derived from the free-energy principle, providing a solid foundation, but they may be considerably easier to get an intuition for, in particular when interpreted in terms of mental simulations. Strongly condensed, these principles may be formulated as *simulation* and *grounding*, or together as *grounded simulations*. Human cognition and the brain can be considered to be all about grounded simulations. To put it more formally, in the words of Karl Friston, many theories of brain function can be united under the perspective of "the brain as a generative model of the world it inhabits" (Friston, 2010).

For practical purposes the most useful principle may be the fundamental importance of *prediction errors*, particularly together with a realization of how prediction errors depend on how a particular prediction fits into a large hierarchy of mental simulations. This is developed further in section 5, but, very briefly, consideration of the importance of prediction errors leads directly to the general impact of randomness or regularity in interaction, and to the importance of existing mental simulations and familiarity with relevant interaction phenomena.

4. Basic principles and HCI theory

In what ways may the basic principles presented in this paper be of value when designing interaction for specific applications? Basic principles of grounded simulation may play a role similar to that of activity theory in HCI. They can be used as descriptive theory, to describe how different aspects of interaction may be grounded in general brain function. They can be used as explanatory theory, by clearly laying out the basics of efficient interaction and explaining why different interaction solutions work well or not. They can be used as generative theory, by connecting interaction design to a multitude of active research areas that may inspire further development of both tools and theory for improved solutions. These roles are the same as are identified for activity theory by Kaptelinin and Nardi (2006) based on the roles and uses of theories as defined by Shneiderman (2002). One specific benefit with shared basic principles of brain function is that there is a large amount of empirical evidence in several different research fields to draw on in designing or analyzing specific interaction solutions.

Shneiderman defines two additional roles for theories: they may be predictive, and/or prescriptive. These roles may have to await the maturation of the referenced theories of brain function and a fuller theoretical and methodological framework based on the presented principles. Still, granted that one is prepared to accept the basic principles as a working hypothesis, it is even now possible to derive several predictions and prescriptions, at a general level. A few examples are given in section 5.

The use of theory within HCI faces a number of challenges related to the diverse and dynamic nature of the field (Rogers, 2012). Human-computer interaction concerns people with a wide range of backgrounds and the range of applications is arguably even wider. The general diversity in the field is also reflected in how theory has been integrated into and used within HCI, drawing on several related fields and reflecting different views on human cognition. The most prominent dividing line is probably between cognitivist theories focusing on modeling cognitive processes and on how the brain works, and theories focusing on the role of human interaction in the world as the basis for cognition. The principles presented here provide a starting point for a theoretical perspective that combines contextually framed interactions in the real world with plausible neural implementation of individual human brain function.

There have certainly been earlier attempts to bridge this divide, e.g., by discussing how mental models may be grounded in interaction (O'Malley and Draper, 1992), but these accounts have had limited impact on HCI, most likely since it can be very hard to change definitions in a diverse community. One of the primary purposes of this paper is to suggest how basic principles that can be recognized in many existing theories of cognition and interaction may be used to find a solid common ground without explicitly pushing a modified version of one or another existing theory within HCI. Basic principles are also particularly valuable in a dynamic field such as HCI since they can be expected to be more stable and reliable over time. As the world changes and new interaction methods become common some of the more specific theories may become irrelevant. Specific theories may be very helpful in practice, but in the long run it makes sense to keep the basic principles in mind, in particular for practitioners that are not experts in cognitive science. The common ground provided by stable basic principles is valuable for facilitating communication between experts and non-experts, and between experts with different theoretical foundations. One does not have to be an expert on cognitive neuroscience or activity theory to grasp general principles that can really impact the design process.

Relation to other theory within HCI

Many parallels can be drawn to existing HCI theory. Connections to activity theory and mental models are described elsewhere in this paper. Below follows some brief examples of how well known theoretical concepts and ideas can be related to grounded simulation principles.

The concept of *affordance* is arguably one of the more widely known theoretical concepts within the HCI community. The concept originates from the ecological psychology approach developed by Gibson (1986, 1983) but the common interpretation within HCI is somewhat simpler than the original idea. Donald Norman described affordance as the properties of an object that allows users to know how to interact with them, that is, as the clues users pick up in order to connect the object to a possible interaction (Norman, 1988). In a grounded simulation perspective this corresponds to the activation of mental simulations as the clues are recognized; mental simulations that make it possible to efficiently recognize higher-level interaction possibilities grounded in such lower level "clue" simulations. Affordances have also been described as possibly being either perceived or actual (Norman, 1999). Perceived affordances are essentially imagined possible interactions while actual affordances are actually possible interactions. This view fits well with a conception of mental simulations as continually simulating and predicting what could happen. Some of these predictions correspond to perceived affordances and consideration of the hierarchy of grounded simulations in the brain give some guidance as to how such perceived affordances are generated. The probability of each simulation depends on the context provided by higher levels and the most likely possibilities will result in expectations to compare against input from lower levels. For example, it is very common that one is able to jump in a platform game. The nature of the game comes with a perceived affordance for jumping corresponding to a mental simulation that the player might try to push into the real world through action, an endeavor that will fail if the affordance is just perceived and no actual jumping function exists.

The importance of the context or situation of the user has been recognized in many developments of HCI theory, for example in *situated actions*. The situated action approach stresses the value of considering the details of the particular situation of the user for possible interactions, rather than designing for some model of how people ought to interact with a system (Suchman, 1987). This fact is illuminated in the grounded simulation perspective by the realization that perception is completely dependent on the current internal mental simulations of the user; simulations that in turn are grounded in interaction with reality and all the details of the actual situation. New information *must* be in relation to users' expectations and these are highly situation sensitive.

In-the-wild approaches focusing on, for example, *embodiment* or *ecological rationality* also lend themselves well to discussion in terms of grounded simulations. The theory of ecological rationality describes how people often use simple heuristics to make decisions quickly (Gigerenzer et al., 2000). Similarly, decisions based on grounded simulations are based on recognition and analogies, triggering reasonable simulations of complex possible futures to select between. According to both ecological rationality and grounded simulation principles humans often rely on only a few important cues, those that carry the most information in relation to current expectations and the corresponding simulations. As for embodiment, grounded cognition is inherently embodied. The brain is part of the body and the grounding for every cognitive function in the brain is in connections to perception and action, as well as bodily state. As a related example, Hurtienne (2009) showed that interfaces that match existing experience-based image schemas support better performance. This fits well with interaction that matches existing (internalized through experience) mental simulations. The unity of perception and action is also a commonality. Both perception and action are based on matching internal grounded simulations to the external world. In the case of perception, we adapt our simulations to match the external. In the case of action, we act on the external world to adapt it to our mental simulation of how things should be.

Activity theory

In this paper activity theory is primarily used to illustrate the connection between basic principles of grounded simulation and useful HCI theory. However, it may also be possible to use this connection to clarify and formalize some of the more elusive concepts from activity theory. Note that this is not about suggesting changes to activity theory per se, but rather about providing new ways of communicating about activity theory, and possibly new ways to enter into and understand activity theory. In an interdisciplinary field such as HCI such multiple entry points and corner stones for communication can be very helpful.

There are aspects of activity theory that limit its application (Diaper and Lindgaard, 2008), some of which become apparent with the advent of new HCI paradigms and applications. In several instances some of the advantages of activity theory come with associated disadvantages. One of the main advantages of activity theory has been the ability to extend a theoretical framework for HCI to include social and cultural context (Rogers, 2005). This is undoubtedly important, but it has (in an HCI context) pulled some of the focus away from theory about individual cognition and brain function. This is unfortunate in light of the increasing use and maturation of brain-computer interfaces (BCI) (Tan and Nijholt, 2010), in particular in relation to applications with a direct link to brain function such as cognitive training and neurorehabilitation (Daly and Wolpaw, 2008).

A description of activity theory in terms of mental simulations, predictions, etc, has the advantage that it provides some separation between the concepts of the theory and the words used to talk about the theory. Without such separation the discussion of the basis for activity theory and the relation to other frameworks is hampered since the everyday meanings of words such as action, goal and activity cannot be used to describe common ground without risking misunderstanding. For example, the fact that activity, action and operation are all concepts with defined meanings in activity theory complicates the discussion about processes and activity in the brain.

If internalization is largely related to the increasing automatization, improving the prediction performance of mental simulations to give less prediction errors and thus run without bothering higher levels, then *externalization* is about bringing the prediction errors back. A natural explanation of externalization in terms of grounded simulations is that the precision of predictions is increased through attention. Predictions made with a higher precision are more sensitive to deviations in the input; that is, input that would have matched the less precise prediction no longer falls within the predicted range and results in a prediction error. In this manner information from automatized mental simulations that one is normally unaware of can be made to rise to higher levels by paying attention to it, making it available for verbal reporting, etc. Alternatively, externalization can come into play when something goes wrong with an automatized operation. This corresponds more directly to a bottom-up prediction error as the results of an operation do not match the predictions.

One aspect of mental simulations is the strong neurological and psychological similarity between these mental reenactments and actually being in the corresponding situation. Imagining a desirable object motivates someone in much the same way as actually perceiving the object. This corresponds to making a prediction. When a prediction is driven by a need it is pushed down through the hierarchy of related simulations, activating them, essentially searching for ways to resolve the prediction error. The interpretation of needs as innate or developed priors can be described as an increase in the precision of the most important predictions. That is, need-driven predictions correspond to predictions of how things *should* be, according to the developed preferences of the organism. A need to quench your thirst may trigger a simulation of yourself at the water cooler and in light of this prediction the observation that you are nowhere near the water cooler is a prediction error. The brain notes that things are not as they should be. This in turn triggers a simulation of getting out of the chair, prompting the motor system to act in order to eliminate the error between the current sitting and the predicted activity, including the sensations related to rising from that chair.

This description of how a mental simulation of an external (potential) object is given the power to drive action by connecting to a basic need is very similar to how *motives* are described in activity theory, suggesting a possible neural basis for the concept of a motive in activity theory. Investigating the details of this similarity is outside of the scope of this paper, but the foundation in basic principles such as the free-energy principle described in this paper offers a way to resolve current debates and compare different interpretations.

5. Practical implications

Grounded simulation principles are well posed to consider both details of actual brain function and the role of realistic complex contexts. This section further develops practical implications for interactions that focus on the brain or on “reality” and wraps up with some more general remarks.

Interacting with the brain

Increased understanding of the brain functions underlying the processes of internalization and externalization is helpful to understand and interpret brain measurements during interaction. An ability to relate measured brain activity to the dynamics of internalization and externalization would be particularly valuable for training applications and adaptive computer applications. Recall that internalization is described at the level of the free-energy principle as the continuous updating and adaptation of a hierarchical model, implemented in the hierarchical anatomy of the brain’s cortex. Improving internalization should correspond to a reduction in prediction errors and related brain activity in the parts of the hierarchy that implements the corresponding mental simulation. Externalization, on the other hand, should correspond to an increase in prediction errors and a corresponding increase in related brain activity.

The combination of activity theory and brain function is of particular interest when designing applications for training in rich contexts; even more so when the training in question targets cognitive functions. Cognitive training applications have the potential to be important tools in the increasingly urgent fight against dementia and cognitive decline (Jones et al., 2006) but recent results cast doubt on the potential for achieving general cognitive improvements (Owen et al., 2010). This motivates an increased focus on activities that a particular user wants to improve on, and on training in realistic interaction environments that allow for specific improvements to transfer to everyday life.

The connections between human activities and brain function presented in this paper facilitate reasoning about how brain measurements change over the course of the training of an activity. Specifically, brain measurements can be related to expected prediction errors that in turn depend on the user’s familiarity with presented phenomena and the fundamental predictability of the phenomena. Within the perspective presented in this paper learning is essentially the same thing as adaptation and adjustment of the mental simulations and associated generative models. Such adaptation requires a certain amount of prediction errors, at the right level in the hierarchy. By considering how the hierarchical composition of a realistic activity relates to such models it is possible to adjust aspects of a computer application to deliberately manipulate brain function. For example, the movements of animated characters in a VR environment can be made more random, forcing the brain to deal with and pay attention to the resulting prediction errors. If an adaptive BCI is used to feed brain measurements back to the application it becomes possible to monitor the response to different attempts to provoke prediction errors, and adapt in real time in order to optimize the conditions for training.

These rapidly growing areas of HCI could benefit greatly from the kind of perspective and framework currently provided by activity theory to HCI in general. This is particularly clear when considering reality-based brain-computer interaction (RBBCI) (Sjölie, 2011), where reality-based interaction (Jacob et al., 2008), such as virtual reality, is combined with adaptive BCI to construct interaction systems where the reality presented to the user is adapted based on interpretations of brain measurements. The environment of the user, with associated objects and activities, needs to be related to brain function to create a basis for allowing the computer to interpret the response of the brain to generated events and changes in the computer-generated reality.

Interacting with "reality"

Several emerging areas of human-computer interaction put an emphasis on designing human interaction with "reality", where the reality in question is either enhanced or replaced using computers. Ubiquitous computing and virtual reality (VR) are two established areas that include such ambitions and the framework of reality-based interaction (Jacob et al., 2008) attempts to capture the general trend towards realism in human-computer interaction, such as tactile interfaces and physics-based games. Within the VR community the basis in human cognition and brain function for the general value of realistic interaction has often been related to the sense of *presence*.

Presence is commonly described as "the sense of being there", as if it were a real place, even if it is, for example, a virtual environment. The primary question suggested by grounded simulation principles when designing for different degrees and forms of presence is: How to match designed experiences in a virtual environment to grounded mental simulations? This question can also be formulated as: How to control the *synchronization b*etween a particular designed phenomenon and the user's subjective mental reality (Sjölie, 2012)? This leads to the question of how the mental simulation of this phenomenon may be positioned in the hierarchy of the user's set of mental simulations. For example, if the phenomenon in question is a virtual remote control, the mental simulation of the remote will depend on the mental simulations of context and purpose, for example, playing a movie, at higher levels, as well as on mental simulation for basic interaction, for example, pressing keys, at lower levels. Once one has a grasp on these dependencies one may consider how the familiarity and predictability of these surrounding phenomena affect the continuous synchronization of the phenomenon under investigation. If everything is familiar and predictable, after the initial exposure, further synchronization will be hampered by a lack of prediction errors, that is, information. This is desirable for any interfaces or equipment that should be transparent, for which the perceptual illusion of non-mediation is wanted. Such interfaces should be as familiar and predictable as the designer can make them. Phenomena that should be synchronized into the user's mental reality, however, should not be entirely familiar and predictable. Instead, the primary design challenge becomes to add just enough unfamiliarity and/or unpredictability to support synchronization, while avoiding large prediction errors that may invalidate the high-level context and trigger breaks in presence.

A typical example of a bad environment for presence according to these principles is a static, sparsely decorated, virtual room with no novel task. The addition of a novel task is probably the simplest way to insert unfamiliarity into the environment. An unfamiliar context immediately introduces uncertainty in all lower level predictions, leading to more prediction errors and greater synchronization of the environment with the mental reality. Adding unfamiliarity or unpredictability to lower-level phenomena, for example with rich and dynamic decoration of a room, also promotes synchronization. In this case the synchronization starts with lower-level mental simulation and progresses upwards as needed, depending on the amount of prediction errors generated or explained away in higher-level simulations.

Tool acquisition

In terms of activity theory, learning to use a new interface can be described as tool acquisition. Considering tool acquisition as the internalization of its possible uses into a mental simulation may provide some insight into how to optimally support interaction with a new interface. In essence, internalization of a new phenomenon is supported by two factors: the basic predictability of the phenomenon, and how well it fits into the pre-existing hierarchy of simulations and predictions. In other terms, how well can it be simulated in principle, and what possibilities are there for grounding? The fit into the hierarchy depends on the relation to both higher and lower level predictions. Higher level predictions correspond to the context of the tool, for example, the purpose and use of the tool in an activity. Lower level predictions correspond to an existing familiarity with the skills needed to interact with the tool, such as grasping or the use of a drop-down menu. New tools should build on existing skills if quick and easy acquisition of the tool is important. This factor is increasingly appreciated in recent developments within HCI, for example, as the utilization of real-world skills in reality-based interaction (RBI) (Jacob et al., 2008). Finally, in order to quickly construct a simulation of the tool that can be used to predict the effects of its use these effects must be predictable. The effect of a hammer in the real world may be quite reliable but the same thing is not always true of the computer applications and electronic devices that are prevalent in the human environment today. While the implications of these considerations may be largely unsurprising the connection to brain function and mathematical models provide a solid foundation that may be helpful when working to resolve conflicts or facilitate tool use in the context of an imperfect reality.

Focus on user expectations

It is worth noting that one does not need to know the details about how a user's mental model is setup in order to take advantage of basic principles when designing interaction systems. Identifying what is familiar and what is predictable is valuable at each level and even if it is only done for parts.

One possible conclusion based on the principles described in this paper is that the question to ask before all others is: *what does the user expect?* The principles presented in this paper, and the related research, gives support for the critical importance of this question, as well as suggestions on how one may begin to answer it. What expectations must, can, or should be violated, and how does this happen? If one understands the expectations of the user, in context, it is possible guide to the user deliberately by introducing information in relation to these expectations. Expectations are almost entirely based on previous experience, and what information reaches the user is based on these expectations.

6. Conclusion

What makes the addition of new theory to HCI helpful? It may be helpful in presenting a multidisciplinary field with additional perspectives that may suit some practitioners better, but in order for this to work well in a wider context it is important that new perspectives can be related to existing theories. If a new theory makes grand claims about being better than previous theory by presenting a truly novel and different perspective this places a heavy burden of evidence on the new theory. However, new perspectives that are easily related to existing theory, and that focus on extending and clarifying general themes, have a value simply by being a new way of looking at, and communicating about, old things. The fact that the basic principles discussed in this paper are not new is a strength, not a weakness.

In many cases it may be a good idea to continue using the theoretical framework that one is used to and knows best, but value can be added by interpreting it in the light of grounded simulation principles, not least by facilitating communication with practitioners not familiar with a particular theoretical framework. An expanded grounding of HCI theory is particularly valuable since it is an interdisciplinary field, where different explanations may support understanding and applications for researchers and practitioners with diverse background. Focusing on relatively simple basic principles also helps to make it practically feasible for a wide range of practitioner to actually take the time needed to get acquainted with them. Complexity has otherwise been a major stumbling block for HCI theory, often preventing theoretical frameworks from gaining traction (Rogers, 2012).

The connections between activity theory and brain function described in this paper serves as an example of how basic principles may support the continued development of general activity theory. Such development facilitates the use of activity theory to provide a theoretical framework for design of systems employing emerging interaction technologies, such as, adaptive brain-computer interfaces in realistic scenarios. The explicit connections between internalized activities, brain measurements, and factors that can be manipulated by design in computer applications, such as the predictability of a computer generated phenomenon, exemplifies how theory can guide application development. Similar connections and potential benefits are possible for theoretical frameworks other than activity theory.

Acknowledgments

This work was funded by Swedish Brainpower, Umeå University, and the Kempe foundations (JCK-2613). Thanks to Johan Eriksson for comments and feedback on the paper.

References

Barsalou, L.W., 2008. Grounded Cognition. Annu. Rev. Psychol. 59, 617–645.

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, 84–91.

Barton, J.J.S., Radcliffe, N., Cherkasova, M.V., Edelman, J., Intriligator, J.M., 2006. Information processing during face recognition: the effects of familiarity, inversion, and morphing on scanning fixations. Perception 35, 1089–1105.

Daly, J.J., Wolpaw, J.R., 2008. Brain-computer interfaces in neurological rehabilitation. The Lancet Neurology 7, 1032–1043.

Den Ouden, H.E.M., Friston, K.J., Daw, N.D., McIntosh, A.R., Stephan, K.E., 2008. A Dual Role for Prediction Error in Associative Learning. Cereb. Cortex bhn161.

Diaper, D., Lindgaard, G., 2008. West meets East: Adapting Activity Theory for HCI & CSCW applications? Interacting with Computers 20, 240–246.

Ericsson, K.A., Charness, N., 1994. Expert Performance: Its Structure and Acquisition. American Psychologist 49, 725–747.

Fadiga, L., Fogassi, L., Pavesi, G., Rizzolatti, G., 1995. Motor facilitation during action observation: a magnetic stimulation study. Journal of neurophysiology 73, 2608.

Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G., 2005. Parietal lobe: from action organization to intention understanding. Science 308, 662.

Friston, K., 2005. A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360, 815–36.

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat Rev Neurosci 11, 127–138.

Friston, K., Stephan, K., 2007. Free-energy and the brain. Synthese 159, 417–458.

Gallese, V., Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. Trends in Cognitive Sciences 2, 493–501.

George, D., Hawkins, J., 2009. Towards a Mathematical Theory of Cortical Micro-circuits. PLoS Comput Biol 5, e1000532.

Gibson, J.J., 1983. The Senses Considered as Perceptual Systems, 1st ed. Praeger.

Gibson, J.J., 1986. The Ecological Approach To Visual Perception, New Ed. ed. Psychology Press.

Gigerenzer, G., Todd, P.M., Group, A.R., 2000. Simple Heuristics That Make Us Smart, 1st ed. Oxford University Press, USA.

Halverson, C.A., 2002. Activity theory and distributed cognition: Or what does CSCW need to DO with theories? Computer Supported Cooperative Work (CSCW) 11, 243–267.

Hawkins, J., 2005. On Intelligence. Owl Books.

Holzman, L., 2006. What Kind of Theory is Activity Theory? Theory & Psychology 16, 5 – 11.

Huang, G., 2008. Is this a unified theory of the brain. New Scientist 2658, 30–33.

Hurtienne, J., 2009. Cognition in HCI: An ongoing story.

Jacob, R.J.K., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Solovey, E.T., Zigelbaum, J., 2008. Reality-based interaction: a framework for post-WIMP interfaces, in: Proc. CHI 2008. ACM Press.

Jones, S., Nyberg, L., Sandblom, J., Stigsdotter Neely, A., Ingvar, M., Magnus Petersson, K., Bäckman, L., 2006. Cognitive and neural plasticity in aging: General and task-specific limitations. Neuroscience & Biobehavioral Reviews 30, 864–871.

Kaptelinin, V., 1996. Activity theory: implications for human-computer interaction. Context and consciousness: Activity theory and human-computer interaction 103–116.

Kaptelinin, V., Kuutti, K., Bannon, L., 1995. Activity theory: Basic concepts and applications. Human-Computer Interaction 189–201.

Kaptelinin, V., Nardi, B.A., 2006. Acting with Technology: Activity Theory and Interaction Design. MIT Press.

Kilner, J.M., Friston, K.J., Frith, C.D., 2007. Predictive coding: an account of the mirror neuron system. Cogn Process 8, 159–66.

Kuutti, K., 1996. Activity Theory as a potential framework for human-computer interaction research, in: Context and Consciousness: Activity Theory and Human-computer Interaction.

Leiser, B., 1992. The presence phenomenon and other problems of applying mental models to user interface design and evaluation, in: Models in the Mind: Theory, Perspective and Application. pp. 267–281.
Leontiev, A.N., 1978. Activity, consciousness, and personality. Prentice-Hall Englewood Cliffs, NJ.
Nardi, B.A., 1996. Context and consciousness: activity theory and human-computer interaction. MIT Press.
Norman, D.A., 1988. The Psychology Of Everyday Things. Basic Books.
Norman, D.A., 1999. Affordance, conventions, and design. interactions 6, 38–43.
O'Malley, C., Draper, S., 1992. Representation and interaction: Are mental models all in the mind, in: Models in the Mind: Theory, Perspective and Application.
Owen, A.M., Hampshire, A., Grahn, J.A., Stenton, R., Dajani, S., Burns, A.S., Howard, R.J., Ballard, C.G., 2010. Putting brain training to the test. Nature 465, 775–778.
Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., 1992. Understanding motor events: a neurophysiological study. Experimental brain research 91, 176–180.
Rao, R.P.., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. nature neuroscience 2, 79–87.
Richardson, D., 2010. Tracking the eye and body: methods, theories and applications.
Rizzolatti, G., Fabbri-Destro, M., 2008. The mirror system and its role in social cognition. Current Opinion in Neurobiology 18, 179–184.
Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L., 1996. Premotor cortex and the recognition of motor actions. Cognitive Brain Research 3, 131–141.
Rogers, Y., 2005. New theoretical approaches for human-computer interaction. Ann. Rev. Info. Sci. Tech. 38, 87–143.
Rogers, Y., 2012. HCI Theory: Classical, Modern, and Contemporary. Synthesis Lectures on Human-Centered Informatics 5, 1–129.
Shneiderman, B., 2002. Leonardo's Laptop: Human Needs and the New Computing Technologies. MIT Press, Cambridge, MA, USA.
Sjölie, D., 2011. Reality-based brain-computer interaction, Report / UMINF. Umeå University, Department of Computing Science.
Sjölie, D., 2012. Presence and general principles of brain function. Interacting with Computers.
Suchman, L.A., 1987. Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge University Press.
Tan, D.S., Nijholt, A., 2010. Brain-Computer Interfaces and Human-Computer Interaction, in: Brain-Computer Interfaces - Applying Our Minds to Human-Computer Interaction. Springer, pp. 3–19.
Wilson, T.D., 2009. Activity theory and information seeking. Ann. Rev. Info. Sci. Tech. 42, 119–161.