Mind Tricks for Presence

Daniel Sjölie
University West
Trollhättan, Sweden
daniel.sjolie@hv.se

Sebastian Badylak
Falmouth University
Falmouth, United Kingdom
sebastian.badylak@gmail.com

ABSTRACT
As virtual and mixed reality (VR/MR) technology moves steadily towards general availability accessible descriptions of the surrounding theory is desirable. An initial focus on high-level concepts can provide common language for diverse teams, including artists, designers and engineers, helping them to quickly get a sense of basic principles and gain a familiarity with related research for further study.

The concepts of synchronized reality and grounded simulation are introduced as helpful starting points for thinking about the design and development of mixed reality systems with optimal presence. This paper provides case studies where recent commercial VR applications are analyzed with the proposed principles in mind, in an attempt to illustrate to developers how to think about design of mixed reality games for optimal presence.

Author Keywords
Virtual reality; the sense of presence; HCI theory, concepts and models; brain function; activity theory; grounded cognition; mental simulation.

ACM Classification Keywords
H.1.2 Models and Principles: User/Machine Systems — human factors; H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
In the interdisciplinary field of game development no one can be an expert at everything and although a familiarity with theory concerning Human-Computer Interaction (HCI) and human perception/cognition can certainly be valuable it can be a challenge to keep such theories in mind in the daily work [24]. Thus, any tricks to stay aware of basic principles can be helpful, providing starting points for deeper dives into theory as circumstances allow. The potential value of basic theoretical principles is particularly notable with novel technologies where more detailed and explicit guidelines have yet to be developed. Currently, the accelerated development of virtual and mixed reality (VR/MR) is an example of an area where new ground is constantly broken and a steady flow of new people are trying to find their bearings in connection to existing theory.

A general understanding of the human brain is particularly important to interaction paradigms that target realistic human interaction, such as virtual or mixed reality [15]. For example, the general principles of brain function described here can be taken as a basis for discussions relating to the sense of presence, often discussed in relation to VR [26].

THE SIMULATING BRAIN
To inform design, basic principles should provide 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. One may take as basic principles 1) everything humans learn must fit into existing mental structures that are ultimately grounded in reality (as one knows it from personal experience), and 2) predicting and simulating potential futures is the basis for cognition. Strongly condensed, these principles may be formulated as grounding and simulation, or together as grounded simulations. Human cognition and the brain can be considered to be all about grounded simulations. 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” [4,135].

The concept of free energy may be difficult to take to heart unless you have a background in mathematics, but the core idea can be explained by considering the importance of surprises [4,5]. 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 in the brain and several theories of global brain function build on this mathematical approach to avoiding surprises [4,13]. 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. Note that these predictions are not conscious expectations but primarily correspond to subconscious low-level predictions in the brain. Prediction errors correspond to the surprises that need to be avoided in order to interact competently with the world and thrive. For example, we want to avoid systems where pressing the same button multiple times produce surprising results.

Throughout their evolution, surviving organisms come to expect and depend on aspects of the environment that facilitate their survival. This corresponds to the evolution of
Presence
Three common definitions of presence are:
1. The sense of being there in a real place
2. The ability to act there as if in a familiar situation
3. The perceptual illusion of non-mediation

In this paper we focus on the 2nd definition. Arguably, this is the most functional one.

Grounded Simulation
Briefly, two basic principles that permeate this perspective are: that new skills and knowledge necessarily build upon and fit into what is already there (grounding) and that predictions and prediction errors are play a key role (simulation).

Synchronized Reality
Efficient interaction with our environment can be said to rely on a synchronization between our subjective mentally simulated reality and the external reality in question, whether the external reality is physical or virtual. To facilitate such synchronization, phenomena in the external reality must be possible to “slot into” our subjective mental reality.

the needs that drive all action according to activity theory [12,18,19]. (For readers unfamiliar with activity theory, the references to activity theory in this paper can be taken as an example of how the simplified principles presented here provide entry points for further digging into comprehensive and established theory). Deviations from these expected states should be avoided: the prediction error, the difference of the current state from the predicted, optimal state, should be minimized. 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 description of the free-energy principle serves to illustrate the connection between the principles discussed and core aspects of life and humanity. Human needs are a constant backdrop to everything we do, and to all interaction with tools or systems. Mindfulness of the connections between such basic statements and descriptions of how our minds are populated by dynamic simulations can support design thinking aiming to develop the computerized environment of humans today.

The theories of brain function described here are still a matter of discussion. The account below aims to convince the reader that the presented perspective is promising and useful. We cannot, however, expect to convince skeptics in this brief format. For a more extensive introduction to the discussion surrounding this theoretical perspective we recommend a target paper by Andy Clark [3].

MENTAL SIMULATIONS
‘Mental simulation’ and ‘mental model’ are expressions that have been used in many theories. At this point, however, mental simulation should be understood as a general concept for describing how the human mind works, as suggested within the framework of grounded cognition [1]. Grounded cognition has been suggested as a high-level description of recurring trends in modern theories of human cognition. 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 [12,18,19]. 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 systems for perception, action and introspection [1,2]. Mental representations are stored in areas of the brain directly related to perception and previous experience of interaction with associated phenomena. 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. 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 concepts such as mental imagery or mental models 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 interact with this simulated keyboard and rely on the resulting predictions of how a real keyboard would respond. 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.

In activity theory such internalizations are created through a developmental process whereby existing physical objects and activities can become internal, giving rise to mental objects representing potential outcomes in the real world. It is towards these potential outcomes, this vision of an object existing in a certain form, that activities are directed [21]. For example, imagining oneself the owner of a new house or the author of an accepted paper can be such driving objectives.

These mental objects can be thought of as corresponding to mental simulations, and their importance is described with the principle of object-orientedness in activity theory. Both mental simulations and mental objects are grounded in interaction with reality and it is a fundamental tenet of activity theory that the human mind is dependent on human activity in the real world [17]. You are what you do, and without action, without interaction with the environment, there can be no mind, no consciousness and no cognition. This reinforces the idea that we should look at real-world activities to understand the structure of human cognition and brain function, and vice versa.

PRESENCE IN MIXED REALITY
The potential for creating a sense of presence, a subjective experience of being “there”, is one of the primary selling points of VR. In particular, the acceptance of a (partly) computer generated environment as real and familiar enables efficient interaction by allowing the user to use established skills like turning an object in your hand to investigate it from all angles [16,25,28]. This focus on the ability to act in an environment is particularly relevant to MR systems. It can be difficult to give a full description of the experience of a MR environment, but it should be clear that the ability to interact naturally in familiar manner with virtual objects in the environment is valuable. This functional description of the sense of presence can be related to the concept of synchronized reality [27]. Synchronized reality builds on the idea that being present in an environment is functionally dependent on a synchronization between the external environment in question (physical or virtual) and your subjective mentally simulated reality [26]. If the mentally simulated interaction possibilities occupying your mind and attention (e.g., what can I do with this virtual brush) match the interaction possibilities in the environment our brain is working in this context in a very concrete manner.

CONNECTIONS
There are many parallels between the principles discussed in this paper and established HCI theory. The examples given below are intended to illustrate the potential of condensed principles to serve as an entry point to and a bridge between established theories. Note that we do not suggest that condensed principles should replace more comprehensive theory.

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 [7,8] 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 [22]. 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 being perceived or actual [23]. Perceived affordances are interactions readily imagined as possible while actual affordances are all interactions that are actually possible (corresponding to Gibson’s original concept of affordance [6]). It is possible to throw a chair, but it is generally not the first imagined use. Conversely, we have noticed how it is increasingly common to see (young) people trying to follow links by touching a non-touch-sensitive screen. This view fits well with a conception of mental simulations as continually simulating and predicting what could happen. Such predictions often correspond to perceived affordances and one may consider how grounded simulations develop, to further understand how such perceived affordances arise. The probability of each simulation depends on the context provided by higher levels and the most likely possibilities will result in expectations to check through actions and sensory 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 [29]. 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 [9].
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 a related example, Hurtienne [14] 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. As for embodiment,
grounded cognition is inherently embodied. Every cognitive
function in the brain is grounded in connections to bodily
state as well as perception and action, and both what we
perceive and what we can do is tightly connected to our body.

**PRACTICAL IMPLICATIONS**

To reiterate, grounded simulations principles focuses on 1)
human experiences as dependent on existing mental
structures that are ultimately grounded in reality, and 2)
prediction and simulation of potential futures as the basis for
cognition. Thus, grounded simulation principles are well
posed to provide guidance for the design and development of
speculative and reality-based interaction systems,
considering both the scope of human imagination and the
reality-based nature of human competence. Current
developments within mixed reality fit well into this
description.

When developing novel interactions, it may be helpful to
consider how interaction can be expected to internalize into
a mental simulation. 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. How well can it be
simulated in principle and how may it find familiar
grounding, in a human mind?

Finding familiar grounding can be directly related to finding
a place in the hierarchy of existing mental simulations where
the new phenomenon can fit in. This fit into the hierarchy
depends on both higher and lower levels. Higher levels
correspond to the context of the tool, for example, the
purpose and use of the tool in an activity. Lower levels
correspond to an existing familiarity with the skills needed
to interact with the tool, such as grasping a tracked hand
controller or using it as a virtual laser pointer. 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,
the utilization of real-world skills in reality-based interaction
(RBI) [15]. Finally, in order to quickly construct a simulation
of the tool it must be predictable. Any potential randomness
needs to be carefully considered and guarded against.

By being mindful of hierarchical relationships one can start
to consider how the familiarity and predictability of
surrounding phenomena affect the continuous
synchronization of a considered tool. If everything is familiar
and predictable further synchronization will be hampered by
a lack of prediction errors, that is, information. This is
desirable for any interaction that should be transparent, for
which the perceptual illusion of non-mediation is desired.
Such interfaces should be as familiar and predictable as the
designer can make them. However, phenomena that should
be synchronized into the user’s mental reality should not be
entirely familiar and/or predictable. Instead, one important
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.

**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 set up 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 conclusion based on grounded simulation principles 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 can, should, or must be
violated, and how does this happen? If one understands the
expectations of the user, in context, it is possible to guide 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.

**CASE STUDIES OF COMMERCIAL VR APPLICATIONS**

This section attempts to illustrate how the principles of
grounded simulations can be applied to achieve the second
definition of presence, as stated above. Two commercial VR
applications have been selected for this purpose: Tilt Brush

In Tilt Brush and PaintLab the user is allowed to paint freely
in ‘mid-air’ by the use of systems such as the HTC Vive [31],
that features sub-millimeter, 6-degrees-of-freedom tracking
of controllers and head-mounted display. Both applications
offer the user a familiar context (free-form painting/drawing)
as well as a familiar set of tools (brush and palette). Thus,
theoretically leading the user to mentally simulate a familiar
situation. Furthermore, the HTC Vive controllers’ virtual
representations strongly resemble their physical counterparts
in a meaningful way - the shape of the controller shaft and
button layout are represented virtually, providing congruent
visuo-haptic cues. With this in mind, we can consider what
the user expects to be able to do with the tools at hand in the
designed context and how those expectations relate to
potential presence in the VR application.
The actual drawing is performed much in the same way as waving a sparkler in the dark\[10\], leaving a similar trail of light suspended in the virtual ‘air’, albeit less fleeting than the real world analogy. In fact, the default brush in Tilt Brush is aptly called ‘light’. The palette, whose virtual representation is more abstract than a traditional palette, is attached to the hand not wielding the brush controller and serves as the ‘toolbar’ common to many traditional painting applications. I.e., the palette mainly lets the user select different brushes and colors to paint with.

The described interface components offer a basis for grounded simulations by using a highly congruent representation of the controller in the context of painting, which is arguably a familiar activity. By synchronizing certain elements in this manner, the developers can then attempt to expand on the concept of a palette by portraying it as 1) Tilt Brush: an abstract cube where each face contains a set of familiar toolbar icons; 2) PaintLab: a circle of abstract beads suspended just above the controller, each bead showing a sample of the paint it represents. In doing so, the application designers could potentially add just enough uncertainty to support synchronization, without overwhelming the user. In other words, the novelty of specific interaction solutions draw attention, and if they are just a little bit novel this attention is converted to understanding as the user interacts with the tool. By adhering to the principles of grounded simulation in this way, developers can provide an entry-point for users that may be completely new to the medium, which is arguably an effective approach to interface design in the advent of consumer VR.

However, anecdotal accounts indicate that some users still have difficulties understanding certain mechanics of the abstract palette in Tilt Brush - such as rotating it around its vertical axis by swiping left or right on the controller’s touchpad. The rotation allows for more convenient access to all of the ‘palette cube’s’ faces. However, turning the physical controller in your hand is entirely possible due to the precise positional tracking of the system, although potentially less convenient. Another conflicting concept which may cause some confusion is that the user needs to point at an icon on the palette with the tip of the brush controller and then press the trigger button in order to make a selection. In the real-life analogous situation, paint selection is normally performed by the brush’s physical contact with a dab of paint on the palette. However, pressing a button conforms with the conventions of traditional painting applications, which, given the computer context, may provide some priming to the mental simulation. In this respect, PaintLab offers a closer visual likeness to the real world palette with its suspended beads of paint, and contrary to Tilt Brush, the selection of paint is performed without pressing the trigger, but by touching the paint bead with the ‘pointer’. Considering grounded simulations the task then becomes to establish which of these groundings in previous experience, real world dabbing of pencils or button clicking to select colors in computer applications, are a better fit both to the expected users and to the desired actions. Which selection interaction provides a good foundation for simulations of continued painting (in space)?
One particular feature of the ‘pointer’ on the brush controller in PaintLab is that it’s located on the side of the controller’s head, pointing perpendicularly to the controller’s orientation in the hand. This feature does provide a highly precise means of selection, because the tip of the pointer can be placed in full view with less repositioning of the hands. Nevertheless, the sideways oriented pointer can likely be considered unconventional and may therefore cause some confusion. In both applications the trigger on the brush controller also needs to be pressed in order to paint. The key difference between the two applications’ input methods for painting is that Tilt Brush’s trigger is binary, while PaintLab’s is analog – the more pressure on the trigger the broader the brush stroke. Neither method may be obvious to new users, but the latter may also be hard to make practical use of. In Tilt Brush the stroke width is set via swiping on the brush controller’s touchpad, which severs the direct haptic connection to the painting but, in turn, makes painting with a constant stroke width more attainable.

To reduce uncertainty and promote synchronization, the trigger interaction in these applications could, according to the principles presented herein, be replaced with proximity-based interaction (such as the drums in SoundStage, where the user plays the drums by hitting a virtual drum with a virtual drumstick. In other words - air drums!). In such a case, the user would touch the tip of a brush to a dab of paint on the palette (much like in PaintLab) and start painting by running the brush across a surface. The closer the brush - the broader the stroke. One apparent limitation of this interaction model is that painting needs to happen on a clearly defined surface and not in mid-air, in order to prevent unwanted strokes. However, unwanted strokes could be mitigated by taking a reductive approach to part of the painting process – one where the user would select an eraser (or paint remover) to trim any excess strokes. Conceivably an inconvenient method for situations where strokes are small and densely located, but nevertheless more grounded.

Further potential alternative solutions to the present ‘palette cube’ and ‘brush controller’ paradigm could be to; 1) offload brush selection from the palette and represent the various brushes as actual virtual brush objects that attach to the tip of the virtual controller when picked up (such as the pens in Tvori [30]); and 2) reframe the meta functions (save/open file, clear scene, change backdrop etc.) to fit in the virtual environment cast as various virtual objects (such as the desk in Tvori, ‘save-room’ in Fantastic Contraption or the components in SoundStage). These changes would arguably clutter the virtual space and perhaps reduce efficiency for expert users, but would in return refine the brush and palette to more analogous representations - i.e. a palette with only a ‘color picker’ and a brush whose representation is more congruent with real life brushes. Thus, perhaps increasing the accessibility for new users. Clutter of the workspace could be mitigated with a virtual assistant that keeps all the tools organized and always positions itself at a convenient angle and distance (like Neko in Fantastic Contraption). Another benefit of having virtual brush objects that users need to pick up is that there is opportunity to allow the users to reposition or realign the brush objects in their grasp, making it easier for them to find a comfortable painting angle. When applying grounded simulation principles to the design of virtual brush objects one should establish that 1) the virtual brushes are recognizable and presented in a familiar arrangement, and 2) that they can be predictably grabbed, avoiding, e.g., accidently grabbing the wrong brush or getting the brush in an unintended position in the hand.

CONCLUSION

The primary motivation for this paper is the value of additional theoretical entry points. In many cases, the value and usefulness of a theoretical framework depends on the background of the reader, as it must be integrated into each person’s pre-existing understanding of the world before it can be used efficiently to support thinking. This paper does not suggest a new objectively superior theoretical framework, but we are convinced that the concepts of grounded simulation and synchronized reality provide valuable entry points to game designers with varied backgrounds.
REFERENCES


