Target Article

How and Why the Brain Lays the Foundations for a Conscious Self

Martin V. Butz ◇ University of Würzburg <butz@psychologie.uni-wuerzburg.de>

► **Purpose:** Constructivism postulates that the perceived reality is a complex construct formed during development. Depending on the particular school, these inner constructs take on different forms and structures and affect cognition in different ways. The purpose of this article is to address the questions of how and, even more importantly, why we form such inner constructs. ► **Approach:** This article proposes that brain development is controlled by an inherent anticipatory drive, which biases learning towards the formation of forward predictive structures and inverse goal-oriented control structures. This drive, in combination with increasingly complex environmental interactions during cognitive development, enforces the structuring of our conscious self, which is embedded in a constructed inner reality. Essentially, the following questions are addressed: Which basic mechanisms lead us to the construction of inner realities? How are these emergent inner realities structured? How is the self represented within the inner realities? And consequently, which cognitive structures constitute the media for conscious thought and self-consciousness? ► **Findings:** Due to the anticipatory drive, representations in the brain shape themselves predominantly purposefully or intentionally. Taking a developmental, evolutionary perspective, we show how the brain is forced to develop progressively complex and abstract representations of the self embedded in the constructed inner realities. These self representations can evoke different stages of self-consciousness.
► **Implications:** The anticipatory drive shapes brain structures and cognition during the development of progressively more complex, competent, and flexible goal-oriented body-environment interactions. Self-consciousness develops because increasingly abstract, individualizing self representations are necessary to realize these progressively more challenging environmental interactions. ► **Key words:** Anticipatory drive, self consciousness, mirror neurons, sensorimotor bodyspaces, language, social cognition.

Introduction

1 Our perception of reality continuously develops, adapts, and structures itself throughout our lifetime. However, the fundamental cognitive capacities form during fetal development, infancy, and early childhood. While genetics lays out the general structure of the brain – constraining the flow of information and thus basic neural structures to certain locations in the brain – development and learning shape the actual implementations of inner representations and control structures. Thus, behavior and cognition – including the construction of the self – are products of both individual genetic predispositions and development. If we want to understand the construction of our inner selves, it is consequently essential to study the developmental and learning aspects of cognition in detail.

2 While cognitive, anticipatory approaches to learning and behavior reach back to the 19th century (Herbart 1825; James 1950), behaviorism dominated the field for at least half of the last century. Watson (1913) perceived psychology as a completely determinable subject that was controllable by simple and measurable experiments, restricting himself to reinforcement-based experiments. Skinner (1971) questioned our own capabilities to make actual intelligent decisions in the absence of perceivable reinforcement. Although the insights gained during the behaviorist age were certainly useful, their influence in psychology research prevented many from having a more open mind on the subject. However, some researchers, including the Würzburg School of Psychology (Ach 1905; Stock & Stock 2004) early in the 20th century and Tolman (1932), propagated cognitive approaches to psychological research that acknowledged the existence of inner mental states that guide and control cognition and learning. Only over the last decades, though, have researchers in psychology begun to explicitly acknowledge that behavior is predominantly controlled purposefully by an anticipatory image of the effects, rather than by mere reactions to a given situation or stimulus (J. Hoffmann 1993; H. Hoffmann & Möller 2003; Hommel et al. 2001; von Hofsten 2003, 2004). These insights led to the current belief that anticipatory processes lie at the heart of cognition and learning (Butz & Hoffmann 2002; Butz, Sigaud & Gérard 2003b; Grush 2004; Hesslow 2002; J. Hoffmann 1993; J. Hoffmann et al. 2007; O'Regan & Noë 2001).

3 According to the insights gained, this article proposes that cognition, individuality, and self-consciousness develop on the basis of the principles of anticipation (Rosen 1985, 1991; J. Hoffmann 1993; Butz & Hoffmann 2002). We propose an *anticipatory drive*, which concurrently biases and guides brain development, decision making, and control. The anticipatory drive has two fundamental effects on brain structuring. First, brain structures are generally predictive, that is, neural structures develop in order to predict the consequences of own behavior and of the external dynamics in the environment. Second, the

predictions do not develop for the sake of predicting, but rather for the sake of anticipatory behavior, that is, for the sake of anticipatory processing of sensory information as well as for the sake of anticipatory decision making and behavioral control. Thus, we essentially propose that the brain is an anticipatory device that (1) continuously forms expectations about the future (in various modules of the brain, depending on the respective representations) and (2) uses those expectations for the generation of effective behavior, the development of further behaviorally-effective representations of the environment, and the continuous integration of multiple sensory and motor sources of information.

4 The anticipatory drive leads to the development of highly interactive brain structures and also to the generation of self representations, which constitute the basis of self-consciousness. Since the drive causes the construction of (more or less detailed) predictive structures of how behavior can influence and change the environment, inner structures emerge that situate the self in the environment but also, in advanced stages, that explicitly differentiate the self from others (objects and beings) in the environment. Equally, it enforces the continuous search for cause and effect relations. In consequence, the drive causes the formation of inner realities of the environment and the self in the environment. Given sufficiently abstracted representations of the self, the anticipatory drive allows the detachment of the self from the present and thus enables the imaginary involvement in (possibly impossible) scenarios.

5 Inversely, the anticipatory drive enables us to execute flexible, goal-directed behavior. That is, our knowledge of possible interactions enables us to inversely generate particular changes to achieve current desirable and achievable goals. In general, the benefits of anticipatory capabilities are manifold; they include the effective, context-based action initiation, faster and smoother action execution, improved information seeking, flexible anticipatory decision making, and predictive attention (Butz & Pezzulo 2008). Thus, the anticipatory drive is not a simple freak of nature but it is useful in itself. Essentially, it causes the development of highly flexible control architectures that are able to consider alternative futures and choose those alternatives that seemingly best suit current needs.

6 The remainder of this paper is structured as follows. We first discuss some prerequisites that are necessary to enable purposeful interactions with, and successful learning in, an environment. To act purposefully, though, brains need to have the tendency to construct particular structures that are suitable for the realization of purposeful behavior. This fact leads us to the proposition of the anticipatory drive, which biases the brain towards the construction of these structures and consequently allows the realization of flexible, purposeful behavior. Taking a developmental perspective, we then show how the anticipatory drive stresses the formation of increasingly abstract self-representations because of the increasingly complex challenges posed by the environment and the interactions of body and mind with the environment. Given the self representations, we then discuss which ones are relevant for the constitution of (pre-) reflexive and (pre-) reflective stages of self-consciousness. In conclusion, we discuss processes that may integrate the formed modular representations and consequently result in the overall experience of self-consciousness.

Towards anticipatory processing

7 Constructivism focuses on the study of how our perception of reality and the integrated self develops. While there is a plurality of constructivist approaches (Riegler 2005), all of them presume certain cognitive structures and mechanisms that lead to the construction of inner realities.

Structure, body morphology, and cause and effect

8 To enable the construction of inner realities, the perceived environment needs to conform to some general principles. Maybe the most fundamental property is that of structural conformity and resemblance, as put forward by David Hume (1748: 62–63):

> "We have said, that all Arguments concerning Existence are founded on the Relation of Cause and Effect; that our Knowledge of that Relation is deriv'd entirely from Experience; and that all our experimental Conclusions proceed upon the Supposition, that the future will be conformable to the past. To endeavour, therefore, the Proof of this last Supposition by probable Arguments, or Arguments regarding Existence, must be evidently going in a Circle, and taking that for granted, which is the very Point in Question."

Hume essentially points out that if there was no structural resemblance over time, learning would be impossible and the construction of inner realities could not occur. While we base our cognition on this resemblance supposition, the actual build-up of our inner realities assumes further fundamental structural principles.

9 Immanuel Kant proposed in the "Kritik der reinen Vernunft" (Kant 1974) the existence of a priori, pure esthetics of space and time ("transzendentale Ästhetik") into which inevitably any cognitive thought will be embedded (Kant 1974: A22–A41). Kant suggests that the construction of our realities does not depend only on experiences and observations, but rather also on a priori, pure knowledge ("Erkenntnis"), which allows the occurrence of experiences in the first place.

10 Modern artificial intelligence (AI) embeds the idea of a priori knowledge into well-designed body morphologies – referring to the structure of a body including the location and type of motor and sensory modules. The term *morphological intelligence* in the embodied AI community refers to the fact that many useful behavioral patterns can be realized by a cleverly designed, purely mechanical, closed-loop coupling of the body's morphology, its sensors and actuators, and the environment it is situated in – without the need for any complex control programs (Pfeifer & Bongard 2006). Thus it is the body morphology that forms the basis of the developmental process in systems that develop further behavioral and cognitive competencies over time. Embedded in the physical constraints of space and time, world universalities can only be detected by means of the pre-programmed (that is, genetically programmed) morphology of bodies, their consequently constrained closed-loop interactions with the environment, and the brain that monitors and coordinates these interactions.

11 While a conformable environment and morphologically intelligent body structures are thus important prerequisites to being able to construct an intelligently behaving system – one that is able to gather enough resources,

reproduce successfully, and thus survive via its descendants – for the construction of an elaborate conscious reality there is certainly more at stake. David Hume already hinted at this third important aspect when he stated "… that all Arguments concerning Existence are founded on the Relation of Cause and Effect" (Hume 1748: 62). This is because there appears to be a continuous flow of interactions present in our environment and these interactions result in various kinds of cause and effect relations. Quantum particles, atoms, fluids, solids, objects, plants, animals, etc. form different types of cause and effect relations when interacting with each other. Thus, due to time, locality in space, and the material concentrations involved, somewhat hierarchically or modularly structured interactions occur. To learn about these interactions, to be able to anticipate them, and, consequently, to act upon them in one's favor, a driving force is necessary that structures the brain to detect relevant interactions and construct explanations of observed interactions by the underlying cause and effect relations.

From sensory-motor couplings to anticipations

12 In living systems, many (inter-)actions often appear somewhat purposeful. A plant "wants" to grow to receive maximum sunlight, a fish swims in a school because it prefers the "protection," etc. However, psychological and biological research, as well as artificial intelligence research, has shown that these interpretations do not necessarily hold true. The problem of the observer and, in particular, our tendency to interpret reality as purposeful often leads us into interpretational traps – assuming an elaborate, purposeful intelligence where there is no explicit one (Pfeifer & Bongard 2006; Rosenblueth, Wiener & Bigelow 1943).

13 In psychology, the awareness of this problem led somewhat to the formation of behaviorism, its strong belief in reinforcement as the only behavior manipulation system, and the disregard of any purposeful interpretations, even of human behavior (Watson 1913). Much later in AI, a similar movement was observable when it was realized that simple sensory-motor couplings, such as subsumption architectures, can lead to very sophisticated behavioral patterns and seemingly purposeful behavior (Brooks 1990, 1991). This was most ingeniously shown in the Braitenberg vehicles experiments (Braitenberg 1984). The behavior of the created robots showed that several aspects of behavioral intelligence may be achieved by simple, cleverly engineered, interactive, closed-loop structures without any complex control mechanism or sophisticated computer program.

14 However, AI also realized that these approaches have their profound limitations, especially in flexibility and adaptability. While reactive, morphologically well-designed control structures can exhibit aspects of purposeful, intelligent behavior, they are not sufficient to realize the behavioral complexity observable in many animals and humans. Tasks that involve memory, context-based decision making and adaptation, and generally more complex, flexible interactions with the environment require more elaborate decision making and control mechanisms. Thus, while an intelligent morphology and clever sensory-motor couplings are essential prerequisites to generating more sophisticated cognitive control mechanisms, they are certainly not sufficient in themselves. To realize actual goal-directed, purposeful behavior, it is necessary for goals to be chosen and activated before the consequently purposeful behavior is initiated. Thus, it needs to be possible to activate an expected future scenario – including a potential goal – meaningfully; that is, properly embedded in the current context.

15 Various disciplines have realized that there is more to behavior than mere sensory-motor couplings. Tolman's "Purposive Behavior in Animals and Men" (Tolman 1932) strongly suggests that behavior is predominantly guided by purpose. Behavior selections in particular were shown to depend on additional environmental knowledge, since they often cannot be explained purely by behaviorism-based stimulus-response learning theories (Tolman 1949; Seward 1949). Thus, a latent learning capability was proposed in which animals associate environmental structures without any immediate benefit or reward.

16 Other researchers in psychology focused more on the question of how behavioral competence, that is, body control, can be learned. A very early account of learning behavior control is now termed the ideomotor principle, which posits that initial random movements lead to bidirectionally linked sensory-motor-effect structures that allow for inverse body control (Herbart 1825; James 1950). As James (1950: 501) put it:

> "An anticipatory image, then, of the sensorial consequences of a movement, plus (on certain occasions) the fiat that these consequences shall become actual, is the only psychic state which introspection lets us discern as the forerunner of our voluntary acts."

17 Sensory-motor-effect couplings (also termed schemata) thus form the basis of control (Drescher 1991; Piaget 1991). The proposition that such behaviors start from rather random, reflex-like behaviors was confirmed in developmental studies with infants. For reaching movements, for example, it has been shown that infants explore their environment in a progressively goal-directed fashion, starting with near reflex-like behavioral synergies (Konczak & Dichgans 1997; von Hofsten 2004). Thus, purpose appears to be at the root of goal-directed motor control and thus also intentional, end-oriented cognition.

18 Ernst von Glasersfeld (2003: v) has summarized the principles of purposeful behavior in the following way:

> "Purposive or goal-directed action could be circumscribed as action carried out to attain something desirable. In each case, the particular action is chosen because, in the past, it has more or less reliably led to the desired end. The only way the future is involved in this procedure is through the belief that the experiential world manifests some regularity and allows the living organism to anticipate that what has worked in the past will continue to work in the future."

19 Similarly, and more recently, Buckner and Carroll (2007) suggest that "we remember the past to envision the future" (Buckner & Carroll 2007: 55), meaning that memory structures do not form for the sake of representing or remembering, but rather for acting upon the environment more effectively when a similar situation occurs in the future (Buckner & Carroll 2007; Schacter, Addis & Buckner 2007).

20 In general, anticipations refer to processes that take advantage of knowledge about potential futures to optimize their current behavior. Anticipatory behavior was thus defined as

> "A process, or behavior, that does not only depend on past and present but also on

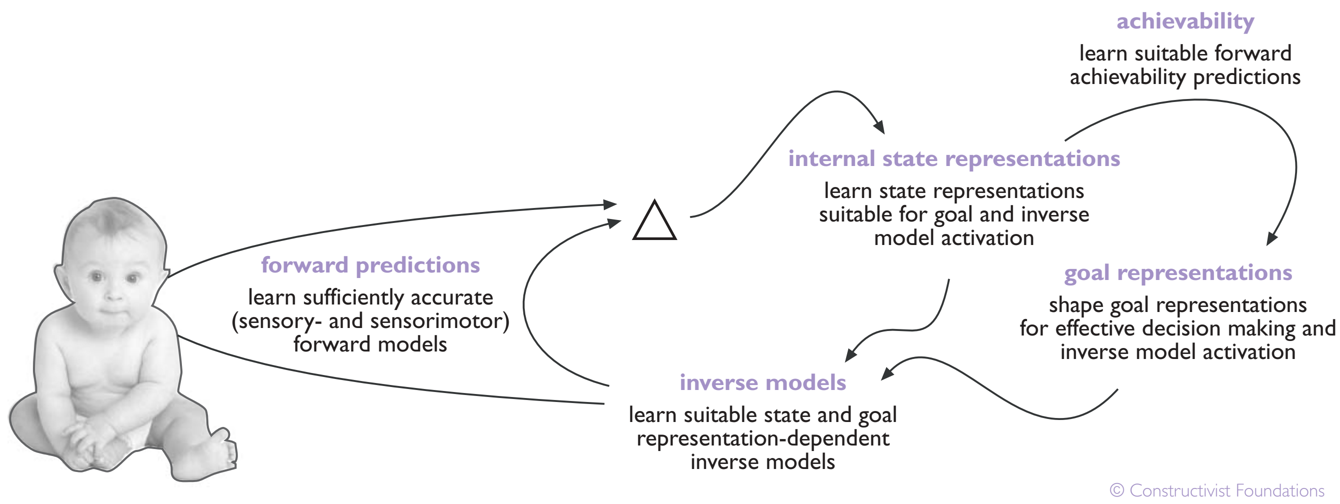

Figure 1: The anticipatory drive results in various learning and structuring biases. When these biases apply through several developmental stages and get involved in progressively more complex environmental interactions, increasingly elaborate self representations can emerge.

predictions, expectations, or beliefs about the future" (Butz, Sigaud & Gérard 2003a: 3)"

In other words, anticipatory behavior refers to predictive knowledge that influences cognition and behavior.[1] How such anticipatory behavior mechanisms are implemented in the brain and what effects they have on brain structuring and cognition are addressed in the following sections.

Anticipatory drive

21 To realize anticipatory behavior, we propose that brain development is predominantly controlled by an anticipatory drive, that is, a learning bias that enforces the formation of bidirectional, anticipatory brain structures. The anticipatory drive is considered the dominant force in the brain that causes the (modular) construction of predictive representations, which enable the activation of goal representations and eventually the construction of our complex inner realities and our conscious selves. The anticipatory drive influences various aspects of the development of brain structures and representations.

22 We now first propose several influences on brain structuring and also discuss consequences for brain activity. The next section then provides various evidences from the literature in psychology, neuroscience, biology, and computational modeling for the existence of the anticipatory drive and its implications for cognitive development and structuring. Essentially, we then plot a pathway that leads to the construction of our conscious selves.

Structuring Influences

23 The most obvious influence of the anticipatory drive is that it biases the brain to learn forward predictions. That is, sensory and sensorimotor structures will be learned that allow the prediction of sensory changes in the environment. To be able to learn such predictions, the brain needs to continuously compare predicted with actually occurring sensations and adjust the predictive model accordingly. Thus a fundamental learning principle is the formation of associative relations over time, which are often additionally conditioned on actual motor control. During actual motor activity, the relations form a closed-loop interactive process of motor-dependent percept associations, which are verified and adjusted by the actually sensed perceptual codes.

24 The anticipatory drive also shapes inverse motor control structures in a goal-oriented way. That is, since only the very first behavioral patterns of an organism can be assumed to be purely reactive, motor control relies on inverse structures that translate desires to motivations and goals, and goals to actual context-dependent motor commands. In consequence, such inverse structures should be shaped to optimize the resulting goal-oriented control. Moreover, the suitability of the inverse structures strongly depends on state and goal representations.

25 Vice versa, state and goal structures must be pro-active so that they are easily translatable into executable motor commands. However, since state and goal representations originate on the perceptual side, perceptions must also be structured not for the sake of perception itself but rather for the sake of motor control. That is, current state representations, goal representations, and the difference between these two all need to be easily transferable into those motor commands that are believed to minimize these differences, thus approaching the represented goals.

26 Additionally, goal representations need to have a structure that is suitable for decision making. That is, goal representations need to differentiate between different motivational drives (such a hunger and thirst) so that current motivations can activate those goal representations that usually satisfy these motivations (such as eating or drinking – or food or water sources). Thus, goal representations need to be structured in a way that is motivation-suitable – in anticipation of their satisfaction upon respective goal representation activation.

27 Besides motivation-dependent goal activation, goal activation also needs to be depen-

dent on the current state. After all, we usually do not formulate goals that are unachievable and we usually do not activate goal pursuance behavior to absolutely unachievable goals. The anticipatory drive thus must generate predictive structures that allow the determination of the achievability of potential goals, since successful goal-directed behavior requires the activation of goals that are not only perceivable but also achievable.

28 Finally, brain modules that are not directly connected to sensory input or motor output will process inherently anticipatory codes. Processed information will typically not only encode the present state of affairs but be continuously and locally suggestive about potential future affairs. Moreover, information processing will not only represent an internal estimate of currently relevant static state properties but also the dynamic sensory and sensorimotor flow. Thus, inputs to some models in the brain will not only consist of actual sensory information or static state information, but also of dynamic information about change in state over time.

29 Figure 1 illustrates the interaction of the discussed brain structuring biases caused by the anticipatory drive. The figure also illustrates that ultimately all internal representations – including state representations, and forward and inverse models – are grounded on the actual sensorimotor interactions of the organism with the environment.

Processing Influences

30 Besides the structuring for prediction and inverse control, the anticipatory drive is proposed as the dominant mechanism that controls brain activity over time. We distinguish the following processing influences.

31 A first fundamental influence is the one on attention. From the bottom up, significantly unexpected stimuli can draw attention. Due to the availability of a forward model, the degree to which a stimulus is unexpected will depend on the forward model so that strong changes in perception do not necessarily need to draw attention if they are expected. In this sense, attention depends on the currently active predictive filters, which are realized by forward model activities. Meanwhile, top-down, goal-oriented attention can yield task-dependent, preparatory increases in particular processing capacities, which result in the capability to analyze particular environmental aspects in more detail but also, potentially, to neglect others – which, for example, leads to the effect of inattentional blindness (Simons & Chabris 1999). In retrospect, the attentional mechanisms also shape further learning so that the anticipatory drive – due to its influence on attention – controls brain structuring in yet another way.

32 Due to continuously active forward predictions and anticipatory codes, the anticipatory drive is expected to result in modular brain activities that not only represent the present state but also, concurrently, potential subsequent futures. The consequence is that the state of the mind is never solely situated in the present but also somewhat "one step" (represented in multiple and various abstract, diverse steps) in the future. In this way, our inner reality is a diverse construct that continuously prepares to process and interact with subsequent stimuli. Behavior decision making and control is thus inherently anticipatory – always ready to act according to the expected future.

33 Interactions between the expected potential futures and (also somewhat expected) current priorities lead to goal selections and the appropriate invocation of the associated inverse and forward models. These co-activations consequentially guide our cognitive apparatus with the invoked attention and behavioral control on a preferably stable pathway through an anticipatory landscape, which is represented by the potential and desired futures embedded into the current contextual state. Decision making chooses amongst alternative future options. Control compensates for undesired disruptions. Even abstract, symbolic, and language-based thought is destined to consider only potential future alternatives on syntactically and semantically constrained pathways – possibly even confabulating stories that do not necessarily match up with the heard truth or even the currently executed, own actual behavior (Riegler 2007).

Construction of inner realities

34 Neuroscientific, psychological, biological, and artificial intelligence research provide evidence that the anticipatory drive controls the development of our inner realities, including our conscious selves. This section lists several important aspects of development that seem essential for the successful construction of these inner realities, conscious thought, and self-consciousness, and relates them to the available scientific evidence. While we consider the discussed aspects highly important for the development of self-consciousness, we do not want to claim that the listed ingredients are exhaustive.

Body control

35 The construction of an individual's reality starts with the capability to control one's own body. As suggested by the ideomotor principle (Herbart 1825; James 1950), body control may start with random, reflex-like behavior but soon starts to shape inverse control structures. The development of the fundamental control capabilities is guided by bodily constraints, which are often referred to as morphological intelligence (Pfeifer & Bongard 2006). To be able to predict the usual sensory effects caused by our own body movements – and thus not to be continuously surprised when we move – a forward model of our own body is necessary. Such purposeful, viable, and goal-directed body control capabilities, as well as the forward projection of behavioral consequences, must be learned unsupervised. The anticipatory drive is ready to induce self-exploration and to consequently learn behavioral self-control and also behavior-dependent self-knowledge. This is also the tenet of the cognitive learning theory of anticipatory behavioral control (J. Hoffmann 1993; J. Hoffmann et al. 2007), which proposes the comparison of predicted and actual action effects as one of the fundamental learning principles.

36 The capability of processing anticipatory motor-activity-dependent information is also in accordance with the reafference principle of perception (Holst & Mittelstaedt 1950), which states that movement execution generates a concurrent signal of the anticipated reafference, that is, the expected sensory effects of the invoked action. Since successive perceptions (even continuously changing ones due to, say, the haptic exploration of an object with closed eyes) directly depend on concurrent motor activities, sensory information is correlated action-dependently. In this way, spatial representations and distance representations are self-constructed and motor-dependent (Butz, Her-

bort & Hoffmann 2007; Butz, Reif & Herbort 2008; Wolff 1985).

37 Such codes may be called *sensorimotor codes* since they correlate sensory codes motor-dependently. Sensorimotor codes have been recently associated with various types of cognitive processes including visual consciousness and imagery (Grush 2004; Hesslow 2002; O'Regan & Noë 2001). Cognitive psychological experiments have shown the intimate correlation between sensorimotor knowledge and its effect on behavioral control. For example, anticipated stimuli can affect action selection and initiation speed (J. Hoffmann 1993; Kunde, Koch & Hoffmann 2004). Moreover, sensorimotor forward projections are used for the substitution of delayed and missing sensory feedback (Desmurget & Grafton 2000; Mehta & Schaal 2002).

38 In sum, in very early developmental stages, bidirectional forward-inverse sensorimotor structures are developed and progressively used to control one's own body efficiently and flexibly. Thus, sensorimotor control structures lie at the heart of self-perception and self-control.

Bodyspaces

39 Depending on which sensory and motor information sources are correlated, sensorimotor knowledge leads to distinct bodyspace encodings. Bodyspaces represent body postures and situate the body in space. They come in various forms and are found in various brain areas, including the pre-motor and motor cortex as well as parietal areas (Butz 2008; Graziano 2006; Holmes & Spence 2004; Maravita, Spence & Driver 2003; Rizzolatti et al. 1997). The body representations are typically population-encoded; that is, by a population of neuronal receptive fields that cover a certain perceptual and motor space. Moreover, they integrate various sources of information, including auditory, various visual, somatosensory (skin perception), proprioceptive (posture perception), as well as current motor control signal information.

40 In the motor cortex, body representations are usually posture-encoded (Gentner & Classen 2006; Graziano 2006). Here, a neural code typically represents a certain body posture and its activation leads to direct movements to the encoded posture. Interactions between pre-motor and motor cortex have been shown to translate visually-dependent codes into proprioceptive, posture-dependent codes in the motor cortex, given that the focus lies on the task-dependent visual position encoding. Moreover, motor-dependent connectivity appears to invoke anticipations of the sensory effects of self-movement (Schwartz, Moran & Reina 2004). Thus, bodyspaces encode sensorimotor correlations so that closeness in a bodyspace is not sensory but rather motor-dependent. In this way, bodyspaces also indirectly encode how effortful it is to translate one sensory state into another.

41 In the parietal cortex, these representations encode how the body is situated in the space surrounding it. Peripersonal spaces encode the space in the immediate vicinity of particular body surface parts. The parts are encoded dependently on the current body posture but independently of the current point of visual focus (Rizzolatti et al. 1997). Peripersonal spaces exist for arms, hands, face, and other body parts (Holmes & Spence 2004; Làdavas, Zeloni & Farnè 1998). Typically, anticipatory closeness is also encoded in that a stimulus that is distant but that moves towards a certain body part may activate neurons that represent that body part – but not if the same stimulus moves in a different direction. Moreover, it is shown that highly unexpected stimuli that are very close to a particular body part – such as an unexpected strong puff of air or the respective stimulation of body-part-representing neurons – can lead to immediate defensive behavior that protects the stimulated region (Graziano & Cooke 2006). Andersen, Snyder, Bradley, and Xing (1997) relate posterior parietal encodings not only to bodyspace encodings for interactions with the body, but also show that the encodings are dependent on current intentions. Thus, peripersonal spaces encode reachability and allow intentional priority-dependent modulations of the encodings. Parietal bodyspace encodings consequently do not serve the purpose of self-perception per se, but rather exist for the purpose of efficient behavior decision making and control – including self-control, self-manipulation, and self-protection.

42 In sum, sensorimotor bodyspaces allow the prediction of action-dependent sensory consequences as well as the invocation of goal-directed behavior, realized by an inversion of the sensorimotor representations. Besides the direct body control representations in motor cortex, parietal areas represent the body in a more abstracted, sensory-integrating, pro-motor manner that situates the body in the environmental context.

Body state maintenance and control

43 Perception and motor control are multi-layered processes in which bottom-up sensory-based inputs are compared with and filtered by top-down anticipations (Herbort, Butz & Hoffmann 2005; Poggio & Bizzi 2004; Tani 2007). Moreover, lateral activation propagations and inhibitions yield diffuse predictions of continuations in time and space. The interactions of these different sources of information result in constructive rather than passively perceptive representations and motor control.

44 Bodyspace perceptions are maintained in a closed-loop process that integrates visual, auditory, proprioceptive, and motor information clues. If one of the sources of information becomes less reliable, its influence on the update process is lowered, while highly reliable information has a larger influence. Also prior information is incorporated, suggesting Bayesian-like information integration processing mechanisms (Deneve & Pouget 2004; Körding, Ku & Wolpert 2004; Rao 2005). Unlike sensory information sources, motor information activates predictive sensorimotor codes, which predict changes in body perception that are dependent on the executed motor commands.

45 Even if nearly all sources of information are unavailable, the inner image is still maintained. This can be typically experienced when walking with closed eyes or in a dark room, whereby the surrounding objects and walls are perceived with increasing (location) uncertainty. Using our hands and feet, we then start probing the space around us to verify its emptiness as well as supposed obstacle locations. However, when there is no sensory feedback available at all, the inner body image cannot be maintained. This is the case for patients that suffer from a very rare disease that destroys their proprioceptive feedback – they can learn, for example, to maintain their body posture and even walk by means of visual control; however, if the light is switched off so that there is no visual information available, they inevitably collapse (Cole 1995).

46 Thus, internal states are maintained by continuous update processes that integrate various sources of information. Forward and top-down anticipations lead to the expectation of future states, which are then verified and distinguished through a regression process by bottom-up, sensory evidence. As Tani (2007: 2) puts it with respect to his computational model of cognitive behavior:

> "… the internal parameters […] are determined through dynamic interactions between the top-down anticipation from the higher level and the bottom-up regression from the lower level."

47 While during body state maintenance body posture is maintained by appropriate stabilizing motor commands, during movement control sensorimotor knowledge enables anticipatory body control. A desired body state triggers those movement commands that can lead to that state, given the current body state and possibly further constraints. For example, the anticipated movement path can be used to invoke predictive control commands, basically inducing the maintenance of a moving stability point (Butz et al. 2007; Tani 2007; Toussaint & Goerick 2007). In this way, small state disruptions can be compensated automatically since they have already been considered in the active representation. Strong disruptions, on the other hand, can induce further processing, attentional focus, and thus further anticipatory learning, since the anticipatory drive stresses the identification of the sources of disruptions.

48 In general, brain modules communicate by means of top-down, bottom-up, and lateral interactions. In bodyspace representations, anticipated activities are verified and disambiguated by the perceived sensory information, which leads to the perception of complete states by the integration of the available sources of information. In motor control, behavior activity leads to the prediction of sensory effects (the most immediate being proprioceptively perceived body posture changes), which are compared with actual effects. Figure 2 shows a very crude illustration of the brain mechanisms involved during behavioral decision making and execution. Given the current body state represented in various bodyspaces, the achievability of future states can be determined by anticipatory knowledge, goals can be selected based on current priorities, and behavior can be controlled inversely starting from the activated goal representations. During behavioral control, sensory effects are predicted and compared to the actual effects (1) to adapt the internal state representation of the body within its environment, (2) to adjust the forward sensorimotor model in the case of small prediction errors, and (3) to detect unexpectedly large disruptions and learn from these disruptions. This last aspect leads to the possibility of forming representations of external entities.

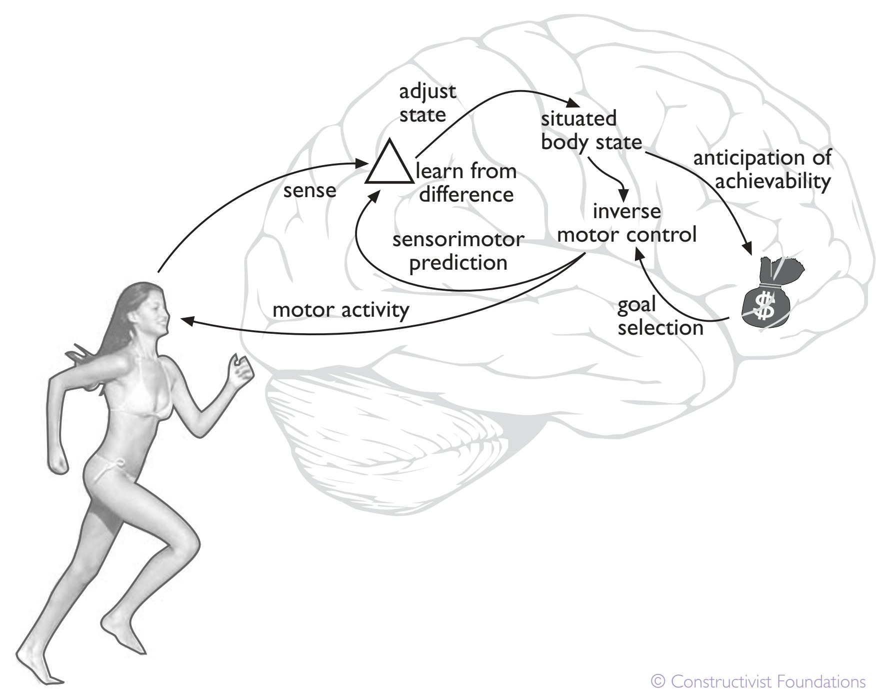

Figure 2: Multiple closed-loop processes result in anticipatory, goal-oriented behavior. Sensorimotor knowledge is used to expect the experienced sensorimotor flow while interacting with the environment. Achievability, knowledge and actual motivational priorities guide goal selection. Finally, goal activities and the current internal body state representation invoke inverse motor control activations.[2]

External environment and objects

49 Given bodyspace representations and sufficient sensorimotor control knowledge, external entities in the environment can be detected when these entities disrupt the usual sensorimotor information flow. As proposed elsewhere (Porr & Wörgötter 2005), disturbances during behavior control can provide information to an organism to allow it to distinguish between the inside of the organism (its body representation and the learned sensorimotor flow) and the outside, which potentially disrupts the usual sensorimotor flow. Recent evidence from neuroscience actually suggests that ventral midbrain dopaminergic neurons may be involved in the latent learning of action-effect correlations, rather than in reward prediction learning, as had been hypothesized previously (Redgrave & Gurney 2006; Wörgötter & Porr 2005). It is shown that these neurons fire in the case of an unexpected event and the timing of the firing suggests that the activity is highly useful to form correlations between context, behavior, and effect – leading to the detection of the particular context and behavior combination that yields the unexpected effect.

50 These mechanisms are in accordance with the postulate of an anticipatory drive that continuously strives to improve predictive capabilities and, inversely, interactive control capabilities. Once an organism is able to control its own body sufficiently well, it essentially also has a sufficiently accurate forward

model available that predicts how sensory information changes in the usual case while motor actions are executed. Given sensory perceptions that are sufficiently different from the predicted flow of perceptions, the anticipatory drive enforces the detection of the causes of these differences. In the simplest case, this leads to the development of external obstacle representations – at least of those obstacles that prevented the execution of undisrupted movements – since only the representation of obstacles allows the prediction of interference. Once interference prediction is possible, again, inversely, behavior adjustments become possible, such as obstacle avoidance behavior but also controlled obstacle interaction. At this point, Piaget's developmental stage two may be reached, after which the child masters a primary sensorimotor loop of behavioral control and object interaction (Langer et al. 2003; Piaget 1975, 1991).

51 For more elaborate object representations, more complex interactions with the object will be necessary. Particularly, more elaborate interaction capabilities (such as hands) will be necessary to generate sufficiently distinct interaction patterns and consequently generate sufficiently distinct object-dependent sensorimotor codes. Equally importantly, more elaborate sensorimotor models (such as hand-eye coordination) will be necessary to distinguish between different object-dependent interactions because only once the sensorimotor model can filter out the usual sensorimotor flow sufficiently accurately, are object-particular differences in sensorimotor flow detectable and consequently representable. Thus, while object distinctions start with the distinction of different sensorimotor dynamics for different objects, ultimately these interactions lead to objectifications of object-dependent causalities; that is, the detachment of objects from behavioral sensorimotor causalities to distinct object-dependent causalities.

52 At this point, the child has reached stage four of Piaget's developmental theory of cognition, in which an elementary externalization and objectification of object-dependent causalities is achieved. Continuous further practice and object interactions then lead to complete externalizations and further refinements of object-dependent sensorimotor interaction codes and the causalities involved (Piaget 1975). The development of knowledge about sensorimotor causalities and, initially to a lesser extent, about perceptual causalities thus leads to the ontogeny of objectifications and distinct object perceptions, as has been verified in various developmental psychology studies (Langer et al. 2003).

53 Once sufficiently distinct and externalized object representations exist – similar to internally represented body states – even incomplete perceptual and sensorimotor clues about objects can lead to the perception of whole objects because the most likely hypothesis corroborates enough information (1) to generate the whole representation internally and (2) to project that whole onto the (not) perceived substructures. Despite this interactive process, it comes as a surprise that we are able to integrate these patterns into a coherent three-dimensionally perceived representation. After all, what we actually sense with our eyes is a highly distorted retinal image, with marginally accurate vision only in the very center of our current point of focus. Thus, successive points of focus must be correlated and integrated into a complete representation.

54 The only invariant information that may connect successive points of focus is the executed motor activity, such as an eye saccade command. Thus, successive sensory information must be correlated motor-dependently and spatial representations are inherently motor-dependently encoded, leading back to sensorimotor codes. With respect to eye saccades, for example, it has been confirmed that the consequences of a saccade are predicted and stabilized by reafference copies stemming from the superior colliculus projected through the thalamus (Sommer & Wurtz 2006; Vaziri, Diedrichsen & Shadmehr 2006). Along the same lines, computational models have been proposed that model the learning of eye saccade control (Mel 1991; Schenck & Möller 2007).

55 Given a particular coherent whole object representation (given current perceptions and possibly also sensorimotor interactions), different object properties will be activated concurrently, including typical perceptual and spatial properties as well as dynamic, behaviorally relevant properties. These latter properties typically have an inherent affordance character, as suggested by Gibson (1979), meaning that the object perception inherently affords appropriate object interactions. Ultimately, action-dependent codes facilitate object interactions and open up the possibility of using objects as tools. In this case, the object representation needs to be integrated into the body representation, since the body is initially the only tool that allows the manipulation of the external environment (Smitsman & Bongers 2003).

Tool use: Linking object and body representations

56 So far we have discussed how the brain may learn to control the body, how it may represent the body dependent on the developing control capabilities in sensorimotor bodyspaces, how it may maintain body state representations and realize body control in interaction with such representations, and how it may develop representations of other entities in relation to the body (e.g., close to a bodyspace) and as external objects, with distinct perceptual and sensorimotor properties. Object representations and body state representations have been analyzed in the most detail in relation to the visual cortex. Two pathways are generally distinguished in the visual cortex and there are indications that similar (soft) pathway splits can be found in the somatosensory processing stream as well as the auditory processing stream (Fiehler et al. 2008). Particularly for vision, the ventral path of visual perception is often referred to as the "what" path of visual perception (Riesenhuber & Poggio 1999, 2000). It processes and integrates object features in a modular, hierarchical, progressively abstract fashion and realizes object identification. Equally, the dorsal path of visual perception – often referred to as the "where" or "how" path of visual perception – is responsible for processing movement and body location in space, closely correlating visual inputs with the aforementioned bodyspaces (Giese & Poggio 2003; Grill-Spector & Malach 2004).

57 Thus, sensory processing distinguishes between *perceptual objectification* and *perceptual subjectification*, that is, perception-based object identification and self-subject identification. Again, both representations are shaped by the anticipatory drive. While perceptual objectification is closely coupled with its significance for behavior, including affordance and motivational characteristics, perceptual subjectifications closely tie percep-

tions with bodily interactions, such as if a stimulus or stimulus object is reachable, manipulatable, or even dangerous.

58 A next step in the construction of the self is to interlink these perceptual objectifications and subjectifications. The development of higher levels of self-consciousness is often associated with the capability to perceive the self as an object (Legrand 2007b; Taylor 2002). To be able to activate such self-as-object perspectives, it is necessary to correlate bodyspace representations (perceptual subjectifications) with object representations (perceptual objectifications) – essentially opening up the possibility to objectify oneself.

59 Recent neuroscientific evidence suggests that self-objectification may be realized by means of cortical interactions between dorsal and ventral processing streams, and in particular, parietal and temporal areas that encode bodyspaces and object identities, respectively. It has been observed that connections between the respective brain areas were much more pronounced in monkeys that were raised in the laboratory and were accustomed to use diverse tools from a very early developmental stage on (Iriki 2006). Here, the anticipatory drive focuses on the mastery of tool usage. Since each tool has particular interaction properties, each tool has distinct sensorimotor interaction patterns. Thus, different tool object identification codes need to project distinct patterns onto the bodyspace encodings in order to integrate the tool into the body perception successfully, consequently enabling effective tool use.

60 Psychological and neuroscientific investigations have shown that tools are in fact integrated into the bodyspace whereby, for example, neurons that represent the hand in a peripersonal space extend their receptive fields onto the tool. The tip of the tool becomes a part of the body in that a neuron that encodes index fingertip locations is now also activated when the tool tip is manipulated or certain stimuli are presented at the tool tip, which previously invoked responses solely close to the fingertip. Similarly, behavior is influenced in that a stimulus at the tool tip has a somewhat similar effect to that of a previous stimulus on the fingertip (Holmes & Spence 2004; Maravita et al. 2003). Thus, it can be said that one learns to subjectify tools in order to use them for manipulation purposes to maximum efficacy.

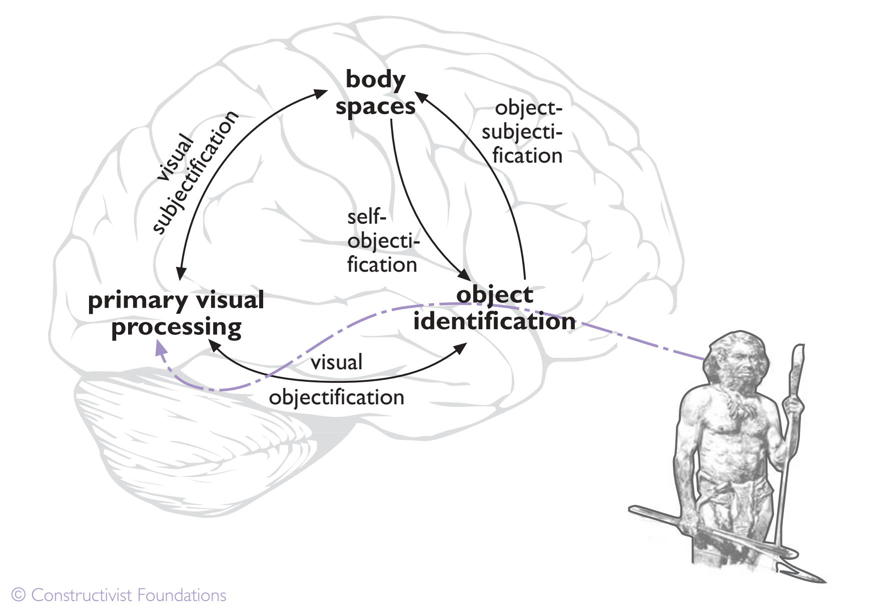

Figure 3: Given perceptual subjectifications in the form of bodyspaces and perceptual objectifications in the form of object identity representations, tool use can lead to the integration of objects into bodyspace representations (object-subjectifications) and, vice versa, to the objectification of bodyspace-originating self-representations (self-objectifications).[2]

61 Vice versa, because brain structures are typically bidirectional, this established pathway due to tool usage also enables the objectification of the bodyspace-encoded self, that is, the *self-objectification* of the available perceptual subjectifications. This hypothesis is not only put forward based on neuroscientific evidence (Iriki 2006), but also from a philosophical perspective – bodyspace representations are proposed to yield pre-reflexive stages of conscious experiences of oneself-in-the-world (Legrand et al. 2007). Through tool-use, these stages may be extended to realize objectifications of the available pre-reflexive self representations. The anticipatory drive to efficiently interact with objects, and the object-as-tool correlation, may thereby induce the reversal; that is, the perspective that a particular body part (such as the hand) represents a particular (and very flexibly adjustable) tool.

62 In sum, tool-use opens up an additional dimension of self perception because object subjectifications result in the possibility to objectify the situated "self-in-the-world" bodyspace-based representation. Figure 3 illustrates the proposed formation of self-objectifying pathways.

Mirror neurons

63 So far, agency and perception have only been discussed in terms of a sole self, that is, body and mind in interaction with the perceived world. However, to force the formation of a distinct self, self-perception and even the capability and utility of objectifying the self (for flexible body-as-tool use) do not seem to be enough. The anticipatory drive does not care about self-perception for its own sake. Rather, a distinct self can only form if the self-perception capabilities also serve another purpose. In this case the self-perception facility needs to be able to distinguish self-perception actually caused by oneself from self-perception caused by other events. To be able to do so, a self-representation needs to be formed that allows the anticipatory drive to distinguish self from other.

64 The detection of mirror neurons in the brain, which are located in bodyspace-near parts in the parietal cortex as well as in the premotor cortex (Rizzolatti et al. 1996; Rizzolatti & Craighero 2004), shows that the brain uses self-perception and self-control facilities to represent others, too. Mirror neurons encode particular goal-directed actions,

such as object-oriented grasping movements. They are active when such an action is executed by oneself but also when a similar action is executed by another person (or monkey) while one is passively watching the action unfold. Interestingly, the existence of mirror neurons also confirms Immanuel Kant's hypothesis that we put ourselves in the place of others, according to von Glasersfeld (2008).

65 For mirror neuron activity to take place, the intention or goal of the action needs to be deducible. It has been shown that if there is no conceivable goal, mirror neuron activity is not detected. Moreover, mirror neurons distinguish between different behavioral intentions although seemingly identical actions are monitored (Umiltà et al. 2001; Rizzolatti & Craighero 2004). Thus, mirror neurons realize the anticipatory drive of understanding others' intentions when observing their behavior. In this context, it has also been shown that the subjective time at which actions and effects are perceived is approximately the same when the action is executed by oneself or by another person as long as the intention of the action can be deduced – adding further evidence that intentions are attributed to others in the same way as they are to oneself (Wohlschläger et al. 2003).

66 Because the brain recruits its own behavior control system to represent the behavior of others, it needs to be able to distinguish self from other behavioral codes. Thus, the brain has to develop an additional representation of self (or recruit available representations) that allows the distinction of self-originated mirror neuron activities from those that are other-originated.

67 One of the clearest distinguishing clues of self and other lies in the much stronger sensorimotor correlations in self-induced motor actions. The reafference principle (Holst & Mittelstaedt 1950) discussed above proposes the linkage of action codes with self-generated sensory changes, so that the encoding that predicts self-induced, motor-dependent sensory changes can serve very well as a self-indicator (cf. also Legrand 2007a). Thus, the necessary integration code that links expected reafferences into the discussed sensorimotor bodyspace representations may be the origin of such self representations.

68 In sum, mirror neurons show that bodyspace-based self representations are recruited to represent the behavior and intentions of others in the environment. Thus, to be able to distinguish the behavior and intentions of others from own behaviors and intentions, brains need to develop additional (or to particularize) distinct self representations.

Imitation, language, and symbols

69 More recent publications on mirror neurons focus on two different consequences of mirror neuron capabilities: (1) mirror neurons are a prerequisite to learning by imitation and to learning a language (Arbib 2001, 2002); and (2) mirror neurons are a prerequisite to experiencing and showing empathy (Gallese & Goldman 1998; Gallese 2001, 2003). Both advances suggest that mirror neurons not only enable more efficient interaction with other individuals but also the development of more complex interactions and further cognitive abstractions. Given appropriate (mainly social) motivations for imitation, the anticipatory drive appears to enforce further improvements and developments of mirror system structures.

70 A major social motivation can be deduced from a game-theoretic perspective. The prisoner's dilemma is a game-theoretic concept (and large research area in itself) that shows that only a society of individuals that has an incentive for common benefit can develop strategies that are not mutually defective (Kuhn 2008). When it is possible for the individuals to remember past interactions with other individuals and, even more importantly, when the individuals are able to distinguish interactions with different other individuals, mutually beneficial behaviors readily emerge (Ridley 1996). The individualization of other individuals essentially allows a better anticipation of the behavior of the other individuals while interacting with them. Thus, the anticipatory drive in social beings forces further individualizations. Since these individualizations are co-represented using self representations (the mirror neurons), the brain needs to establish representations that allow the proper distinction of self from other, which leads to further particularizations of the self.

71 Based on an incentive to interact and distinguish other individuals, Arbib (2005) proposes several successive stages in language evolution that may have led to the complex and diverse language structures we find in our world today. First, beginning with mirror capabilities, simple and complex imitation stages need to be reached. In these stages, the mind learns to imitate observed, increasingly complex goal-directed behaviors. In doing so, it is very difficult to directly map observed actual movement but comparably easy to map goal-orientedness; that is, the actual effects of the environmental interaction, such as object manipulations. This observation again confirms the intentional characteristic of mirror neuron activities: the anticipatory representation is mirrored, not the movement itself.

72 Once a sufficiently complex imitation stage is reached, movement coordination comes into play that often requires the common usage of language commands and instructions (Knoblich et al. 2005). In turn, these complex interactions, mediated by simple commands, must have started to lead to increasingly advanced symbolizations. Commands are a symbolized activity on their own, since a command usually implies a certain goal-oriented activity. Complex imitations and, most likely, coordinated actions must have then led to further stages of symbolizations (Arbib 2005; Deacon 1997; Sebanz, Bekkering & Knoblich 2006). Sophisticated hunting strategies and even elaborate warfare show that effectively symbolized communication and thus efficient group coordination can give significant survival advantages.

73 Given the capability to imitate complex behaviors and the need to communicate by initially simple signs (proto-signs) during collaborating activities, Arbib (2005) proposes the further development into protospeech and finally language. Thus, given the capability to deduce the intentions of other individuals, to imitate them, and the drive to collaborate with them, the capability arises to mirror the potential meaning of perceived words onto the own current understanding of the world. Swarup & Gasser (2007) argue that language evolution presupposes mirror capabilities, sufficient memory capabilities, and adaptive value for advanced social interactions, among other factors. Due to the rise in adaptive value of increasingly (but boundedly) complex forms of communication and society, language and social structures have co-evolved: increased language capabilities enable larger and more

socially complex interactions, which, in return, result in the incentive to improve language capabilities even further (Arbib 2002, 2005).

74 Besides the advantage due to elaborate social interactions and collaborations, it is also acknowledged that most of these stages are accompanied by cultural developments. That is, beginning with simple imitation capabilities, which can also be found in other mammals and birds, cultural knowledge can be passed on to children in an increasingly effective manner. However, since knowledge that is passed on in this way also stresses the establishment of further brain structures (as seen in the tool use example), these developmental and cultural evolutionary stages forced the establishment and differentiation of brain structures that would not develop if the cultural influence was not available. Thus, language and culture must have progressively coevolved, yielding increasingly higher survival and reproductive advantages (Deacon 1997). Meanwhile, language and cultural co-development must have structured our self-conscious experiences even further.

75 In our sophisticated cultures, the involved symbolization in language is further enforced and structured by various other factors. These include learning how to read and write, counting and mathematics, or learning other languages. The increased abstraction of cultural interactions – the concept of money, admission tickets, supermarkets, etc. – give rise to further complex abstractions, objectifications, and symbolizations of concepts, which would otherwise not exist with such clarity in our minds. Thus, thought projections and projections of the concept of self become more and more diverse, enabling imaginations hardly possible without the existing cultural influence – not to mention the vast amount of literature, cartoons, movies, and various forms of art spanning realism and impressionism to various types of modern conceptualizing art forms, all of which put forward various other perspectives on reality and forms of surreality.

76 In sum, social interaction, coordination, imitation, and language capabilities require yet more sophisticated and symbolized codes that represent the self and distinguish the self from others to satisfy the anticipatory drive. Meanwhile, increasingly complex forms of interaction, communication, collaboration, and coordination are becoming possible, which are embedded in increasingly complex social and cultural human environments. In particular, language and all the even more complex symbolic structures that arise from language force the construction of highly symbolic structures. Although highly symbolized, all these structures are still strongly grounded in the initially constructed bodily sensorimotor codes because the symbolized structures emerge during development, starting from the discussed sensorimotor codes for behavioral control. In fact, researchers are now beginning to model sensorimotor-grounded language codes and are developing parts of a sensorimotor-grounded grammar of behavior (Guerra-Filho & Aloimonos 2006; Guerra-Filho & Aloimonos 2007).

Self embedded in society

77 While the discussed language-based aspects of the self yield rather symbolized forms of self-representations, the social self also comprises more fluid, emotional-based self-representations. As mentioned above, mirror neurons are also considered a prerequisite for the development of empathy, relating mirror capabilities to simulation theories of mind reading and understanding others (Gallese & Goldman 1998; Gallese 2001; Hesslow 2002). Recent neuroscientific evidence also supports the idea that empathy is realized by means of sensorimotor-grounded, simulation-based processes that are mediated by mirror neurons (Banissy & Ward 2007). By simulating the behavior of others via mirror neurons, their current emotional states become perceivable. Coming from the simulation quality of mirror neurons and the consequent social comprehension of others in the environment, Gallese (2003) proposed that social reality is represented by a *shared manifold*, which is grounded in sensorimotor, embodied structures, including mirror neurons. Since others and the self are projected onto this shared manifold, the representation of a common social reality emerges. The construction of the social self thus begins with the representation of self and others in a common, bodily-grounded manifold.

78 The capability for actual conversations – be they light conversation or about abstract concepts – is controlled and guided by mapping one's own knowledge onto the perceived communicative patterns. Perceived sounds are projected onto own language patterns and the underlying syntactic and conceptual structures. However, there is certainly no one-to-one mapping, since we have developed the ability to distinguish different individuals. Embedded into our own cognitive structure, we have theories about other people's minds, which specify their assumed knowledge, their current potential intentions, thoughts, and feelings. All these suppositions may help to make sense of the heard auditory inputs, deducing both (1) the actual words and sentences being uttered and (2) a self-constructed potential meaning of the words. Comprehension consequently depends directly on our current knowledge about the conversed topic as well as on our knowledge about the other individual and, most importantly, on the expectation of what the other individual might currently want to convey.

79 The elaborate language system then enables complex social interactions, and thus the construction of both an increasingly complex social reality and an understanding of the perceived society. The different aspects of social interactions and distinct individual properties may be embedded in the shared manifold (Gallese 2003) of the perceived overall social reality. Thus, humans with different cultural and developmental backgrounds must inevitably perceive society from different anticipatory perspectives. Self-perception and one's role in society are products of learned social constraints, circumstances, peer pressures, etc. that are integrated into prior (genetic) individual developmental differences. Similarly, other individuals are perceived distinctly – such as the suspicions we might have of strangers or the trust we put into our friends – resulting in context- and individual-dependent social interactions and unique individual perceptions of social reality.

80 In sum, the perception of the self in society represents yet another aspect of the conscious self. Since many distinct particular properties are attributed to other individuals (to be able to anticipate their behavior and thoughts), types of properties are also distinctly attributed to the self, enforcing a representation of the social self that is integrated in the constructed inner social reality.

Facets of self-consciousness

81 As plotted in the previous section, the anticipatory drive as the basic learning mechanism that underlies brain structuring has now created brain modules and mechanisms that include various forms of self-representations. We now reflect on the developed representations and their interactions and relate them to successively complex forms of self-consciousness.

82 We distinguish reflexive and reflective stages of self-consciousness, based on Legrand's terminology (Legrand 2007b). The idea of a distinction between reflexive stages, in which the "I" appears as the subject that experiences, and reflective stages, in which the "I" is observed as an object (by the "I" as subject), however, reaches back (if not further) to Immanuel Kant, who pointed out the necessary distinction between "transzendentaler Apperzeption" (transcendental apperception) of the self and the recognition of the self as object:

> "Wie aber das Ich, der ich denke, von dem Ich, das sich selbst anschauet, unterschieden (indem ich mir noch andere Anschauungsart wenigstens als möglich vorstellen kann) und doch mit diesem letzteren als dasselbe Subjekt einerlei sei, wie ich also sagen könne: Ich, als Intelligenz und denkend Subjekt, erkenne mich selbst als gedachtes Objekt, …" (Kant 1974: B155)[3]

83 Thus, Kant asks how the perceiving self can be distinguished from the self-perceived self, that is, how can the "I" as subject recognize the (same) "I" as object? While further elaborations and analyses of Kant's perspective on this matter can be found elsewhere (Brook 2008; Legrand 2007a), we now focus on how reflexive and reflective stages of consciousness can be realized within the developed representations discussed above.

Reflexive stages

84 Reflexive stages of self-consciousness comprise representations of the self-as-subject, that is, the self as the egocentric frame of reference. The following aspects form the basis for such reflexive stages of self-consciousness.

85 First, to learn the skill of flexibly controlling one's own body, body control mechanisms develop that allow forward predictions of self-induced motor-dependent sensory changes as well as inverse, goal-directed body control. Second, bodyspace representations situate one's own body in space, creating the self-related frame of reference and consequently enabling self-protective, self-exploratory, and interactive activities. Third, since body control and self representations in bodyspaces are mediated by top-down anticipatory and bottom-up regressive interactions, the resulting internal representations allow the prediction and inverse control of one's own perceptions. These first three parts of self-representations comprise the embodied self and may be found in diverse (and more or less elaborate) forms in most brain-controlled animal minds. In sum, the self-as-subject perspective constitutes predictive and inverse control capabilities that interact in body-originated frames of references that are represented in sensorimotor bodyspaces.

86 In pre-reflexive self-conscious stages, these representations are used to interact with the environment from a body-centered, egocentric frame of reference. In reflexive stages, attention focuses on the "I" as subject and may adjust the egocentric frame of reference in order to improve environmental interactions (Legrand 2007b). Essentially, the brain activates the self-as-subject perspective during any sensorimotor interaction, which is also the tenet of related sensorimotor perspectives put forward elsewhere (Grush 2004; Hesslow 2002; O'Regan & Noë 2001). In these cases, the readiness of processing subsequent sensorimotor interactions itself is proposed to constitute the current state of conscious awareness.

87 During reflexive stages of self-consciousness, however, the environment, including one's own body, is not necessarily represented from a self-as-object perspective. Such a perspective leads to "higher," reflective stages of self-consciousness.

Reflective stages

88 The interactive form of perception and motor control in terms of top-down anticipatory mechanisms and bottom-up sensory and motor-feedback driven regressions, however, do not only lead to representations of bodily-induced selves. They also cause the representation of other entities –objects, obstacles, substances, plants, animals – in the environment because the resulting internal representations lead to different forms of anticipatory interactions and relevant anticipated entity behaviors. Thus, entity representations lead to a first stage of objectification of the environment.

89 Given objectifications, it becomes possible to objectify the self, but it is far from necessary. When acquiring the skill (which is also strongly culturally mediated) to interact and utilize objects (or entities) in the environment as tools, the brain learns to "subjectify" objects and other entities. Then, vice versa, this subjectification lays out the pathway for an objectification of the available self-representations. Knowledge of bodily capabilities, such as the perspective of our hands-as-tools with high versatility, then lead to the association of body parts with tools – where the one can replace or enhance the capabilities of the other. Self-manipulations start to be comprehended in objectifying forms, and the self-as-tool perspective leads to the possibility of establishing the first pre-reflective forms of consciousness (Legrand 2007b). These allow for, for example, the exploration of one hand with the other hand or with the eyes, perceiving it as the object of interest.

90 Further abstractions of this objectified self are then mediated by various additional social and cultural factors. Social interaction, the mirroring of other individuals onto the self representation, and the consequently necessary distinction of self (and properties of the self) from others leads to a further individualization of the self. Social coordination and interaction, on the other hand, also lead to an integration of the self in the group of individuals and thus a localization of the self in society (and aspects thereof), represented in a shared manifold of social reality.

91 Language provides an entirely additional source that enforces symbolization, objectification, and abstraction. Naming objects and naming the self (the "I") is yet another source of inevitably strong individualization and abstract self-perception. Furthermore, language allows the interchangeable usage of names as subjects and names as objects, further facilitating imaginative subjectifications and objectifications.

Conclusion

92 The anticipatory drive, that is, the tendency of the brain to form predictive structures and inverse control structures, in conjunction with developmental and various environmental influences, leads to the construction of representations of the self in various forms. So far, however, the question of how and when which of the forms is actually active has only been marginally addressed. By itself, the anticipatory drive does not account for the actual choice of currently activated representations, or, to put it another way, it does not account for the currently active interactions between these representations. Thus, the understanding of how we perceive a unified self that appears to be continuously embedded in the individual forms of self-perceptions remains obscured.

93 As discussed above, the anticipatory drive controls attention and decision making based on desired future states. This decision making and top-down attention can be mediated by current priorities, motivations, emotions, and goals. Knowing future alternatives enables choice. And the consequently necessary decisions must be made based on current internal prioritizations (which most likely stem from current motivational and emotional biases), which are projected onto the available alternatives. In this way, choices become prioritized in a goal-directed manner. Attention focuses the mind on those perceptual and representational aspects that are task-related. And as a whole, the mind focuses its mental processing capabilities on those aspects that are relevant in some way.

94 While how this is accomplished by the brain is still under fierce debate, a couple of aspects seem relevant. From an artificial neural network perspective, Velde & Kamps (2006) proposed a model of neural blackboard architectures, which can integrate current thought into a complex network structure. Neural blackboard architectures are essentially a model to solve the *binding problem*, which also underlies the unified perception of consciousness. The question in the binding problem with respect to consciousness is: How can different aspects of self and of current perceptual inputs and motor activities be combined such that the subjective unified self is perceived? Other types of blackboard architectures have been proposed before (Newell 1990), tackling the same problem. Global workspace theory (Baars, Ramsøy & Laureys 2003; Shanahan & Baars 2005) has been proposed as the enactor of the observing self – selecting and binding currently relevant brain activities. Embedded in the developed modular structures discussed above, these approaches thus propose different binding techniques to realize coherent interactions.

95 A somewhat similar binding approach was proposed that correlates attention with consciousness. Here, the mechanisms that control attention are considered to be the same mechanisms that evoke consciousness (Korsten et al. 2006; Taylor 2002). Given that attention is guided by motivational and emotional biases, as suggested above, once attention is applied to the self-representing structures, and particularly once attention uses the different objectification capabilities discussed above, a unified conscious self-perception can emerge.

96 Besides the attentional root of conscious thought, more details on the actual neural mechanisms of interacting cortical structures may be found in neural synchronization mechanisms. It has been shown that neural synchronizations between cortical modules are a strong indicator of neural communication (Ward 2003; Fries 2005; Fries, Nikolic & Singer 2007; Singer 1999; Womelsdorf et al. 2007). Thereby, several different neural cycles prioritize information and extend them in time, whereby the most significant information comes first. Moreover, communication between different brain areas is established through synchronization. Thus, information binding and the involved attentional processes appear to be mediated by neural synchronizations so that conscious states and self-consciousness may also be realized by neural synchronizations.

97 Irrespective of the exact origin and functionality of the binding mechanism, though, the effect of the mechanism must be the invocation of our unified subjective conscious states, including reflective self-consciousness. While this mechanism must also invoke the subjective qualitative conscious experiences – also integrating motivational and emotional aspects – the *qualia* debate of why qualitative self-perceptions feel the way they do (Levin 1999) is out of scope of this paper's intention. From the proposed structures that lie at the root of consciousness, though, it clearly follows that consciousness has nothing to do with a body-detached soul. The proposed emergent self-representations close the mind-body problem by integrating the conscious mind into bodily perceptions, sensorimotor interactions, language, and society, which are constructed based on the brain's modularity and the anticipatory drive that structures the modules and their interactions.

98 In closing, it is open to discussion whether the consciousness arising out of the proposed

THE AUTHOR

Martin V. Butz holds a Ph.D. in computer science from the University of Illinois at Urbana-Champaign (USA). Since October 2004, he has been back at the Department of Psychology III (Cognitive Psychology) at the University of Würzburg. Butz has been studying anticipatory mechanisms with various computational modeling approaches for nearly ten years. He has been collaborating with linguists, neuroscientists, and cognitive psychologist in order to further the understanding and the conceptualization of anticipatory representations and mechanisms. He is the co-founder of the "Anticipatory Behavior in Adaptive Learning Systems (ABiALS)" workshop series and has published and co-edited six books on learning classifier systems and anticipatory systems. Currently, Butz is leading his own junior research group, which is funded by the Emmy Noether Programme of the German Research Foundation, focusing on the development of cognitive control architectures that model actual human behavior and reach new levels of behavioral flexibility and adaptability.

mechanisms is now an epiphenomenon or actually a very useful entity that controls our selves. We agree with Taylor's (2002) perspective, which regards consciousness as an attention-based control process: Given that consciousness arises from attentional processes – and attention is essentially thought and behavior control – our "highest" states of consciousness are also actual control states. Thus, even symbolic language-mediated conscious states have a control character and can therefore be used to control less abstract, bodily thoughts and behavior. However, it remains to be understood when a particular control module can be considered the currently dominating control instance, or rather, how responsibility may be distributed amongst the modules that are part of the overall self-control process.

99 The question of how these mechanisms work together, how they maintain the continuous overall activity balance between the interacting brain areas, and how they ultimately control our individual selves and constitute our selves at the same time will still be under debate and researched for many years to come. Nonetheless, it is hoped that the overall picture drawn in this article will foster this debate and guide it towards further insights into how our brain-body system works and how consciousness self-develops and self-structures, depending on the unfolding interactions of body, mind, environment, and society.

Footnotes

1. It should be noted that there is nothing mysterious about such anticipatory behavior since future representations are brain constructions, which are created due to the brain's knowledge of cause and effect relations and its supposition that the future resembles the past.
2. Brain process localizations are kept general and are certainly neither anatomically precise nor necessarily restricted to one particular area or location in the brain.
3. "… how 'I who think' is distinct from the 'I' that intuits itself (other modes of intuition being cogitable as at least possible), and yet one and the same with this latter as the same subject; how, therefore, I am able to say: 'I, as an intelligence and thinking subject, cognize myself as an object thought' …" (Kant 2003)

Received: 8 August 2008
Accepted: 19 September 2008