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

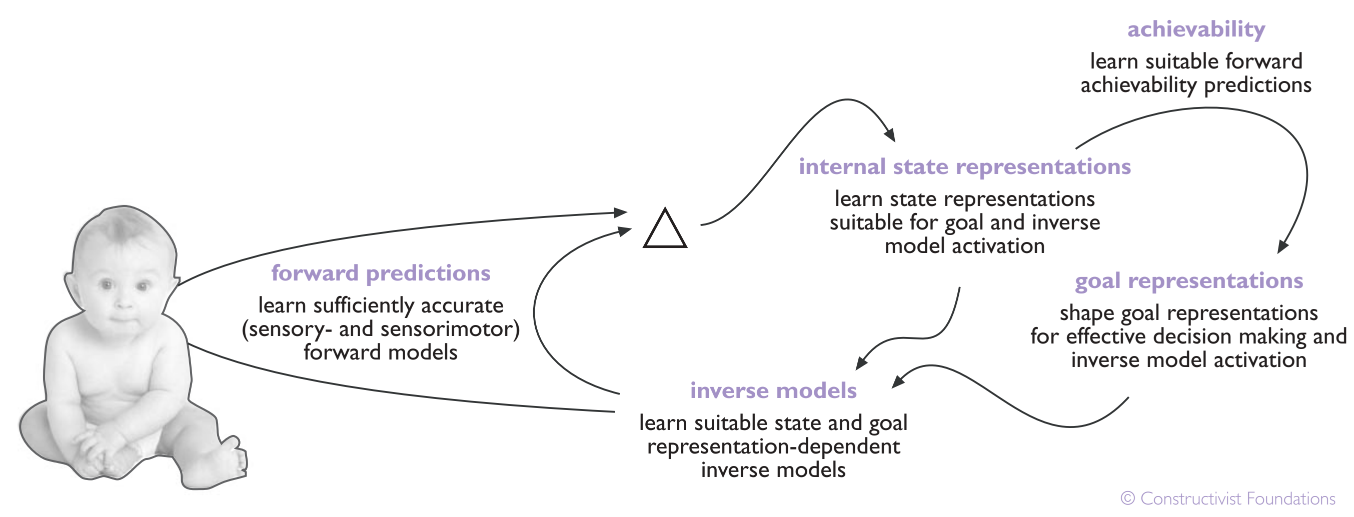

Figure 1: The anticipatory drive results in various learning and structuring biases. When these biases apply through several developmental stages and get involved in progressively more complex environmental interactions, increasingly elaborate self representations can emerge.

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.

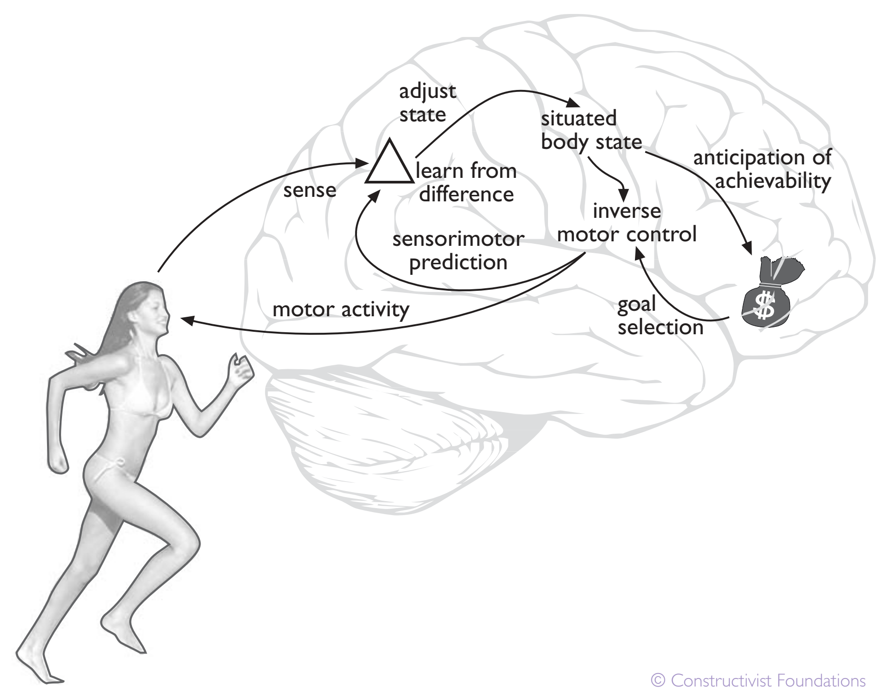

Figure 2: Multiple closed-loop processes result in anticipatory, goal-oriented behavior. Sensorimotor knowledge is used to expect the experienced sensorimotor flow while interacting with the environment. Achievability, knowledge and actual motivational priorities guide goal selection. Finally, goal activities 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.

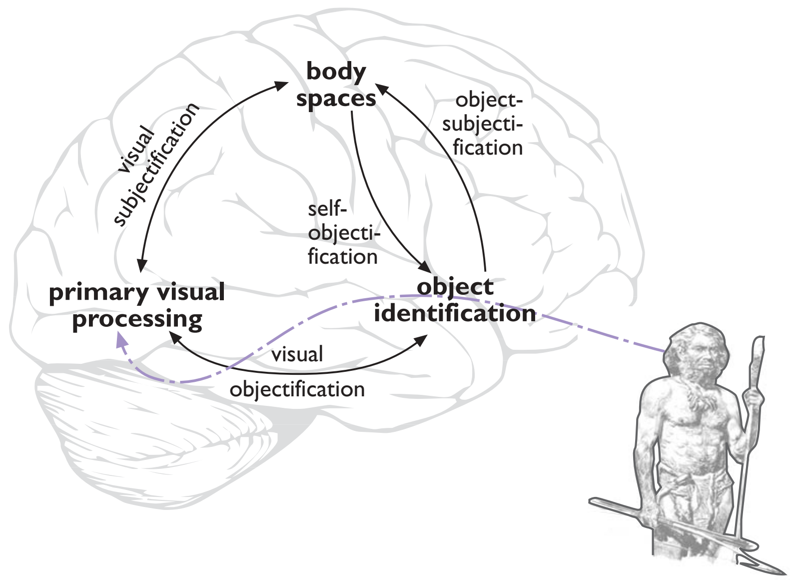

Figure 3: Given perceptual subjectifications in the form of bodyspaces and perceptual objectifications in the form of object identity representations, tool use can lead to the integration of objects into bodyspace representations (object-subjectifications) and, vice versa, to the objectification of bodyspace-originating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

Open Peer Commentaries

Why and How to Avoid Representation

Marco Bettoni
Swiss Distance University of Applied Sciences
<marco.bettoni@weknow.ch>

"La science est une langue bien faite"
– Condillac (1780)

1 At first sight Butz's target article attracted me, principally for three reasons: an ambitious purpose (in the abstract), the extensive mentioning of attention in the conclusion (§§92–99), and a set of potentially powerful ideas, for example the importance given to the anticipatory drive (§3) and the connection between the anticipatory drive and attention (§31).

2 But later, when I began to read the article more carefully, two completely different, more formal issues attracted my attention: first the fact that Butz uses the term "representation" more than 100 times[1] (but avoids "mental representations") and then that he uses the term "brain" 80 times.

3 Since I was trying to read the paper from a radical constructivist point of view, I experienced these facts as surprising and misleading – an obstacle to my understanding that I hoped to be able to overcome by taking a closer look at the concept of representation as it is used in the paper.

4 Why should the term "representation" constitute an obstacle to a radical constructivist understanding? Although dealing in detail with this question would require a lot of investigation, the basic idea is very simple and well expressed in the following quote: "To speakers of English [the word 'representation'] implies a reproduction, copy, or other structure that is in some way isomorphic with an original" (Glasersfeld 1995: 94).

5 With such a term it becomes, then, quite difficult, if not impossible, to keep in mind and comply with the radical constructivist principle that claims that "…what we call knowledge can be 'constructed' without reference to anything outside the experiential confines" (Glasersfeld 2005: 11). The main difficulty consists in the tight relation that – at least in our culture, in our language – exists between "representation" and "reference." For example, in Webster's ninth new collegiate dictionary (1984) you can find the following four main entries for the meaning of "to represent":

> "1: to bring clearly before the mind … 2: to serve as a sign or symbol of … 3: to portray or exhibit in art … 4: to serve as the counterpart or image of…"

6 If the brain is an anticipatory device (§3) and the mind has an anticipatory function, then the term "representation" is potentially misleading. It implies and suggests reference to real-world objects: the mind, when it interprets the term "representation" (1) forms expectations of "reference" and (2) uses those expectations for the construction of mental constructs unawarely conceived of as isomorphic (or covariating) with items outside the experiential confines.

7 A first interesting instance of the term "representation" appears in §1 where Butz mentions that in the brain "development and learning shape the actual implementation of inner representations." Here "inner representation" could mean what in German is designated as "Vorstellungen," and "implementation" could refer to neural structures and processes that embody them in the brain. To avoid "inner representations" we could say, "Development and learning shape the actual implementation of mental constructs."

8 In §3 we find the expression, "development of … representations of the environment." This is an example of where "representation" easily misleads because it suggests reference to real-world objects in the mentioned "environment." Ceccato used to criticize this way of thinking as "raddoppio del percepito" (doubling of the perceived): a thing of the environment is transferred to the inside by perception so that we then have two things: one outside, the unknown thing, and one inside, the known thing (Ceccato & Zonta 1980: 41–45). The particle "of," with its plurality of possible meanings (semantic functions, relations), further increases the chances of misunderstandings in this same direction of cognition, (tacitly) conceived of as doubling.

9 At the beginning of §8 Butz mentions the "perceived environment". If we use this formulation to restate the expression from §3 then we could say "development of … mental constructs of the perceived environment," if we want to focus more on the functional level and its elements (conceptual structures); alternatively, if we want to focus on the neural level, i.e., on structures and processes that embody conceptual structures[2] in the brain we could say "development of … implementations of the perceived environment".

10 From a radical constructivist point of view, this distinction between a functional level and its enabling neural (device) level, as we have seen in the previous examples, is very important in order to avoid misleading expectations. Unfortunately, the term "representation," as used by Butz, does not support such a distinction and thus potentially misleads to confusion of the two levels.

11 Another distinction that can contribute to reducing uncertainty of interpretation in the description of complex systems such as the mind and the brain is that expressed by the terms "operation", "operator" and "operand," which are typical of cybernetic thinking. The term "representation", as a "nomen actionis" (like many other words ending in English with "-ation" and in German with "-ung"), does not support this distinction because it combines two meanings in the same word: the action of representing (operation, in fieri) and the result of representing (operand, in facto).

12 Among the 111 occurrences of the term "representation" in Butz's article, about 70 are in the form "representation of X" (or the equivalent "X representation"), where X is mostly one of the following 7 concepts: self, goal, state, bodyspace, body, object, entity.

13 In all these cases, the mentioned distinction between the functional and the device levels could be well expressed by using either "mental construct" or "mental operation" for the functional level and "implementation" for the device level (brain). Table 1 presents some examples of original sentences and, for each, two possible reformulations.

"The anticipatory drive leads to … the generation of self representations" (§4)
▪ …the generation of mental constructs of the self
▪ …the generation of implementations of the self
"…the suitability of the inverse structures strongly depends on state and goal representations" (§24)
▪ …strongly depends on the mental operations of states and goals
▪ …strongly depends on implementations of (the operations of) states and goals
"In the motor cortex, body representations are usually posture-encoded" (§40)
▪ …mental operations of the body are usually posture-encoded
▪ …implementations of the body are usually posture-encoded
"For more elaborate object representations, more complex interactions with the object …" (§51)
▪ For more elaborate mental operations of the object …
▪ For more elaborate implementations of the object …
"…brain modules and mechanisms that include various forms of self-representations" (§81)
▪ …mechanisms that include various forms of mental operations of the self
▪ …mechanisms that include various forms of implementations of the self

Table 1: Possible reformulations.

14 To determine which of the two reformulations would be more appropriate in these examples and in the whole 111 instances of "representation" in the target article would require either a deep and complex analysis of this and related texts by Butz or, even better, a collaboration with Butz himself: I would be happy to participate in such a work, if Martin Butz would be interested.

15 Avoiding the term "representation" would make the article much more consistent with a radical constructivist way of thinking. It would also open up unexpected opportunities for realizing the potential of some of its most interesting ideas, such as the connection between anticipatory drive and attention (§31).

Notes

1. The author uses the term "representation" 111 times (5 times in the abstract) and its root "represent-" 140 times, with the forms: represent-ation, represent-ations, to represent, represent-s, represent-ed, represent-ing, represent-able, self-representations, representational (6 in the references). As a comparison, the key term "anticipatory" appears 89 times and its root "anticipat-" (like in "anticipatory", "anticipation", etc.) is used 114 times.
2. Conceptual structures can involve both figurative and operative elements (Glasersfeld 1995: 98): figurative elements are abstracted from sensorimotor experience; operative elements (for example conceptual relations) are constituted of attentional operations.

Treacherous Terms

Ernst von Glasersfeld
University of Massachusetts
<evonglas@hughes.net>

1 The author argues that the ability of organisms to anticipate outcomes of actions is crucial for the construction of experiential reality and, ultimately, the generation of consciousness. His paper is a most interesting exposition of the hypothesis that the most evolved organisms have acquired an "anticipation drive" and is a detailed account of this drive's proposed role in the development of the higher functions of the human brain. The author wisely states at the end that "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."

2 I am not sufficiently versed in the neurophysiology of the brain to evaluate the physical plausibility of this hypotheses but there are a number of questions that a student of cognition and language can raise that may be of help in tightening the author's proposal.

3 Those of us who have tried to expound and explain constructivism in the past have been incessantly hampered by the traditional implications inherent in the use of a natural language that was formed and developed in a climate of naïve realism. It is difficult to remain aware of the fact that if someone says, "There is a squirrel," he is actually saying that he *seeing* (i.e., isolating in his visual field) an item that he categorizes as "a squirrel." Although it sounds like it, he is not talking (and cannot talk) about squirrels as though they were independent of his perceptual activity.

4 A somewhat analogous ambivalence is created for Butz by the term "code" in the variety of combinations that the author uses ("perceptual" §23, "anticipatory" §28, §32, "sensorimotor" §37, "neural" §40, "interaction" §52, §54, "behavioral" §66, "language" §76) and other terms such as "encoding" (§§39, 40, 41, and more). Let me stress that I am not bringing this up as a criticism, but as something that, in my view, requires clarification.

5 In ordinary English, "code" means an item or list of items that are *semiotically* linked to something else, something to which they are not otherwise related. The semiotic link is the result of a convention and its intension cannot be inferred from either of the two linked items. Genetics and computer science have borrowed the word "code" and given it a different meaning. An item of the genetic code transfers its "information" by a physical copying process in the generation of other molecules. In our computers, the transfer of "information" is achieved by the interaction of electrical charges. Hence such transfer is a causal affair and in neither case is there the need for a reflecting agent

that is aware of a conventional meaning. This is an important difference from the semiotic domain, where knowledge of the particular convention is the *only* way of "reading" a code.

6 In §44, the author mentions the reliability of information and the incorporation of prior information by means of a Bayesian-like "information integration processing mechanism..." and he concludes:

> "Unlike sensory information sources, motor information activates predictive sensorimotor codes, which predict changes in body perception that are dependent on the executed motor commands."

7 I find it difficult to unravel the paragraph, because I cannot make up my mind whether "information" has to be interpreted semiotically as the instruction to select a specific item from a pre-established code, or biologically as the causal trigger to an action. The first interpretation, it seems to me, would invalidate the hypothesis as a model of the arising of consciousness because it entails an agent who is aware of coded meanings. The second interpretation would, I think, require further explanation that avoids ambivalent terms.

8 Similarly, in §72 the author states: "In turn, these complex interactions," (i.e., between different individuals) "mediated by simple commands, must have started to lead to increasingly advanced symbolizations." – If a command, as for instance in the military, is nothing but the trigger for a particular action, it does not function as a symbol, but its sound-image has, for the receiver, become the physical cause of an action. As such, it may well lead to more complex causal connections; but in order to lead to "more advanced symbolizations," a reflective agent has to be posited, and this, it seems to me, seriously interferes with the intention of explaining the genesis of consciousness in terms of neurophysiological mechanisms.

9 In short, I feel that the use of terms such as "code," "information," and "symbol" for neural constellations that are not further described defeats the intention of the analysis presented because unless these terms are explicitly given specific neurobiological definitions they inevitably suggest the presence of a consciously reflecting agent.

The Mind Is Not In the Brain

John Stewart
Université de Technologie de Compiègne
<johnmacgregorstewart@gmail.com>

1 This is a fine article, which makes many excellent points, in particular about the importance of anticipation; and how nice it is to see Kant so aptly quoted in a scientific text, what's more in the original German! I do, however, have one serious criticism: it concerns the "internalist" stance adopted by the author. The article opens with the statement "perceived reality is a complex construct"; clearly, no constructivist could disagree with that! However, in the very next sentence Butz simply assumes, without argument, that we are dealing with an "*inner*" construct; he goes on, throughout the article, to speak of "*inner* realities." I would like to explain (a) why I think this is a mistake; and (b) why it is a serious mistake … *especially* for constructivists.

2 Much of the problem stems from the way the article rather glibly talks as though the *brain* were itself a cognitive subject. This is apparent right away in the title, where "the brain" is attributed the status of an agent; later on "… the brain has to develop an additional representation of self…" (§66), and so on. A related problem concerns the way the author talks of "mirror neurons" (§§63–69), again as if neurons could actually do things such as imitate, empathize, and so on. The neurophysiological observations concerning "mirror neurons" are certainly striking, and it is understandable enough that they have caught the popular scientific imagination. But it is vital to understand that "mirror neurons" are merely correlates, phenomena that are *to be explained*; they are *not* themselves a proper explanation of any cognitive behavior and function. Talking about neurons and brains as though they were themselves *bona fide* subjects is a category mistake: it is *people* that perceive, think and so on, not brains. I suppose that for most people nowadays, in the West anyway, it simply seems "obvious" that consciousness, and mental activity generally, are processes that take place "in the brain"; indeed, that psychic activity *just is* – neither more nor less than – brain activity. Sometimes, however, it can be salutary to question the obvious.

3 Much confusion arises from the fact that the brain is so enormously complicated, and its functioning so mysterious, that it is easy to believe almost anything and indeed to attribute it with quasi-magical powers. Artificial neural networks comprising a mere dozen or so "neurons" can give rise to very intricate behavior in real or simulated robots; and the dynamics involved are already so complex that even in these relatively simple situations, where it is possible to have complete knowledge of the system, it is far from trivial to understand exactly what is going on. What are we to make, then, of the human brain with its 10^9 neurons and 10^{11} synapses? What is done in current neuroscience is to correlate *differences* in mental activity with *differences* in brain activity, giving rise to the colored brain-images that have become so familiar. Precisely because we do not really understand what is happening, the temptation is strong to believe that we are actually *seeing* mental activity going on. However, correlation is not cause; and it is important to resist that temptation. To explain why, I will adopt a ploy proposed by Mikael Karlsson (1996) and compare the relation between brain activity and cognizing with a far simpler case that we *can* understand properly: to wit, the relation between leg activity and walking.

4 Clearly, we could not walk if we did not have legs – just as I fully admit that we could not cognize if we did not have brains. But it does not follow, *at all*, that walking "*is*" neither more nor less than leg activity. For example, an astronaut floating in weightless conditions could move her legs all she likes – that would not be walking. For leg movements to be involved in actual walking, a whole set of contextual conditions are necessary. The legs must be attached to an upright body, in a gravitational field and on a reasonably flat, more or less horizontal surface; there must be adequate friction between the feet and the ground; and so on.

5 An interesting point arises if we ask *where* the walking is taking place. Actually, it is rather difficult to give a precise spatial location to the walking. *Parts* of the walking have a precise location: I can say that this morning my walking started from my apartment when I left it, and stopped when I got to the café where I sat down to have a drink. But the

nature of the "contextual conditions" is such that overall, the walking has a *nebulous* location, without any clear-cut boundaries. Moving outwards from the centre, the location includes more than just the bits of ground where I put my feet; it arguably includes the whole street, and, in a way, (but less definitely) the buildings and trees and parks that I pass by. Moving inwards from the cosmos as a whole, we can say that my walking is definitely happening on the planet Earth, rather than the solar system (because of the composition of the atmosphere); in France (because of the linguistic context), and in Paris (because of the weather, and the general atmosphere…). Thus, there is no clear-cut boundary between what is contextually relevant and what is not.

6 But even if the location is "nebulous" in this way, and difficult to pin down, there is one thing that we *can* say for sure: the walking is not happening "in the legs." I claim that exactly analogous considerations hold for the relationship between the brain and the mind. The brain is only involved in perceiving, thinking, imagining, feeling, being conscious and so on – what I have called "cognizing" for want of a better general term – *because* it is contextually situated in the body of a living organism, itself engaged in actions in an environment.

7 Butz himself implicitly recognizes this to a very considerable extent, by the importance he (quite correctly) attaches to the "grounding" of cognizing in embodied sensory-motor dynamics. However, the "brain-centered" talk remains a niggling worry; and as I indicated at the start of this commentary, it has what I consider to be a particularly damaging consequence.

8 "Perceived reality is a construct"; yes, we can all agree with that. But why should we assume that it is an "*inner*" construct? If the mind is not in the brain, as I have argued, the "perceived reality" is not "in" the head either. Since the brain is only involved in actual cognizing to the extent that it is situated *in a body that is interacting with its environment*, the reality that is "brought forth" (Maturana & Varela 1987) or "enacted" (Varela et al 1991) is *co-constructed* in the interaction between organism and environment. This co-construction is constrained (and made possible) by the particular features of *both* the organism and the environment. The position I am arguing for is neither "internalist," nor "externalist," but rather seeks to go beyond the opposition between them; the construction happens in the interaction, and insofar as it is "located" anywhere, it has a "nebulous" location, rather like the walking in my simple example.

9 Why does this matter for constructivists? Well, I think it is important because even at the best of times constructivism is already widely accused of idealism, solipsism and/or relativism, so the last thing we need is to give free ammunition to our opponents! We need to be able to say that if reality is indeed constructed, the construction in question is *constrained* by a "reality principle"; constrained, that is, not just by the particular features of the organism, but equally by the environment, and above all by the interactions that occur between the two.

Anticipation and Self-consciousness

Are these Functions of the Brain?

Humberto Maturana Romesín
Matriztic Institute, Santiago (Chile)
<info@matriztica.org>

1 My reflections will be first, about how the brain operates in the generation of the adequate behavior of an organism in a changing medium, and second, about how self-consciousness appears in the course of the history of humanness.

2 The first question arises from our daily experience of seeing an organism behaving in a way that seems to anticipate some desirable result, or from observing a developmental process as if it were guided by a drive to reach some particular form. These observations have given rise to the suggestion that the brain and the organism operate under the action of some *anticipatory drive*. The second question arises when we hold the view that language is an instrument that we human beings use to refer to entities that are external to us, and we find ourselves asking, *how do we distinguish ourselves* if we are not objects external to ourselves?

My claims

3 In these reflections I shall claim that the operation of the nervous system is not *anticipatory*, and that as a structure-determined system it cannot be anticipatory, even if for an observer it may seem to be so as he or she sees an organism behaving adequately in its changing niche. I shall also claim that self-consciousness is not the result of some particular neuronal process in the nervous system, but that it is a manner of living that has arisen in our human history as the consequence of our living as languaging beings in a flow of recursive coordinations of coordinations of consensual doings in which we are objects of our coordinations of doings.

4 Indeed, I shall claim that there are no anticipatory processes in the cosmos and that the result of a process is not and cannot be part of its occurrence, and that self-consciousness is the particular inner feeling that we feel when we see that we are doing what we are doing.

Structural coupling: My fundaments for the answer to the first question

5 We living systems are molecular systems. As molecular systems we are structure-determined systems, and as structure-determined systems we are systems such that nothing external to us can specify what happens in us. Something external to us impinging upon us can only trigger in us structural changes determined in our structural dynamics. Therefore, we human beings as molecular living systems are structure-determined systems, and all that applies to living systems as structure-determined systems applies to us. Structural determinism is not an assumption; it is our condition of existence.

6 A living system arises in the operations of distinction of an observer as existing in three non-intersecting operational domains: the domain of the realization of its molecular autopoiesis, the domain in which it operates as a totality, that is, as an organism, and the domain in which it realizes its relational living in operational dynamic congruence with its niche as this arises continuously in the actual realization of its manner of living as an organism of a particular kind.

7 When the observer distinguishes an organism, he or she brings forth in the same act the operational-relational environment in which he or she sees it, as well as the opera-

tional-relational medium in which he or she imagines it to exist, and in which the organism realizes an individual niche that the observer cannot see unless the organism itself shows it to him or her with its behavior.

8 The observer calls "environment" all that he or she sees surrounding the observed organism, and calls "medium" all that he or she imagines as the great container in which an organism realizes its living. The niche arises as that part of the medium that the organism "touches, sees, hears … or more generally encounters" in the many dimensions of the realization of its dynamic manner of living, and that the observer cannot see unless the organism reveals it through its living.

9 An organism exists in a continuous process of structural changes as a result of its own internal structural dynamics modulated by the structural changes triggered in it by its interactions in its niche. The niche in which an organism realizes its living occurs in a continuous structural change arising in the interplay of the dynamics of the medium of which it is part, and the structural changes triggered in it by its encounters with the organism.

10 The organism and its niche constitute an operational unity or totality in which both, the organism and its niche, change together congruently, in a process that follows a path that is continuously arising anew in the flow of their interactions. The path followed by the congruent changes of the organism and its niche arises in the encounter of otherwise independent processes, and an observer cannot predict its course even though he or she can expect one if the encounter occurs as part of a recurrent conservative dynamic that he or she has seen before as the environment of the organism, or has imagined as the medium in which it exists.

11 What an observer distinguishes as the behavior of an organism is not something that the organism does by itself, but a changing relational dynamics that arises in the recursive encounter of the organism with its niche. As a changing relational dynamics, a behavior involves at the same time the organism in the realization of its autopoiesis, and the niche in the realization of its participation in the structural dynamics of the medium at that instant.

12 In these circumstances, what an observer sees as the behavior of an organism is its tangent encounter with its niche in a structural dynamics of coherent structural changes that is the result of the history of recursive interactions between the organism and its niche. I have called such dynamics of coherent structural changes in which the organism conserves its autopoiesis, *structural coupling*.

13 When an observer sees an organism conserving its living (autopoiesis) in its domain of structural coupling in its niche, he or she sees it as an organism generating adequate behavior in its niche, whatever this may be. For the observer, the adequate behavior of the organism may appear as anticipatory, that is, as foreseeing what "the organism needed." Yet the organism was only operating in the coherences of its structural coupling in its niche, in the present of a history of recurrent dynamic structural coherences that constituted a matrix of operational-relational coherences in which the organism can conserve its living precisely because the organism and its niche change together around the conservation of the manner of living of the organism.

14 Systemic laws are abstractions of the spontaneous operational coherences of systems in any part of the cosmos that the observer brings about in its living. Three are the most fundamental ones, and I present them below.[1]

15 *Systemic law of the observer and observing*: "Everything said is said by an observer (a human being making distinctions in languaging) to another observer that could be him or herself."

16 *Systemic law of conservation and change*: "Whenever in a collection of elements a configuration of relations begins to be conserved, a space is opened for everything else to change around the configuration of relations being conserved."

17 *Systemic law of the course of history:* "The course that follows the history of the structural change of organisms in general, and of human beings in particular, arises at every instant of the living of the organisms or of the human beings defined by the preferences and desires of the organism or the human being, and not by what an observer may think are opportunities or possibilities for the organisms or human beings involved. Something is an opportunity or a possibility only if it is desired."

18 *In synthesis:* What an observer sees as *adequate behavior* in an organism, is its operation in the present in dynamic structural coherence with the medium in its niche that is the result of the conservation of the operational structural coherence of the organism and its niche in a history structural drift in which the organism and its niche have changed together congruently. Such a process occurs spontaneously without the participation of any guiding orientation towards an end as a result of the operation of the organism as a structure-determined system. Structural determinism is a constitutive basic feature of the cosmos that we human beings bring forth with our operation as molecular systems. Structural determinism does not imply predictability. Structural determinism is the basic condition that creates the possibility of understanding and explaining of all processes in the cosmos, even probabilistic ones.

19 There is no possibility of the operation of any process that could be legitimately called *anticipatory* or that could be legitimately considered to occur under an anticipatory drive. If an observer sees an ordered process giving rise to some result that is surprising or admirable to him or her, and if he or she does not understand structural determinism, he or she will not understand the dynamic architecture that gave origin to that result, and will invent some semantic notion to connect the different instances of the process in a way that he or she can accept. The notion or idea of an *anticipatory drive* is such a semantic notion under the form of an *a priori* explanatory principle (cf. Ximena & Maturana 2008).

20 Everything in the cosmos occurs as it occurs as a continuously changing present in which complexity arises in the encounter of processes that happen to be locally dynamically independent, even though they may be part of a larger systemic one. This is expressed in the following systemic law: "The result of a process does not and cannot operate in the process that gives origin to it" (Ximena & Maturana 2008).

Languaging and objects: My fundaments for the answer to the second question

21 If we attend to what we do in language, we will realize that language occurs as a flow of living together in coordinations of coordinations of consensual doings. That is, we will realize that language occurs as languaging, in the flow of our living together in recursive consensual coordinations of doings. Language has the concreteness of the doings in

the domain of doings in which we coordinate our doings.

22 Objects, entities, notions, ideas, concepts etc., arise as coordinations of coordinations of doings, and do not exist otherwise. The meaning of the words, sentences, signs and symbols is not in the words, but in the flow of coordinations of doings that they coordinate. And a word can have as many different meanings as there are different flows of recursive coordinations of doings in which it participates.

23 When a child learns to name an object he or she does not learn to name a preexisting entity, but learns a flow of recursive coordinations of doings with the languaging persons with which he or she may be living. So a baby that learns the ball, learns balling, and when he or she learns the doll, learns dolling. Thus, the baby learns in the same way, eyes, feet, mouse, lips, … self, thinking … as flows of recursive coordinations of consensual doings with other human beings, as manners of living together in consensual coordinations of doings … and emotions as manners of relating in coordinations of consensual coordinations of relational doings.

24 As an object arises as a flow of consensual coordinations of doings; the domain in which the arising object arises and has presence also appears as a domain of recursive consensual coordinations of doings in which the arising object participates in the recursive coordinations of consensual coordinations of doings that constitute its meaning. As the objects do not pre-exist the flow of consensual coordinations of doings that they are, the domains of recursive consensual coordinations of doing in which they exist as coordinations of doings arise anew with them, and new objects constitute new domains of existence as new domains of recursive consensual coordinations of doings.

25 The *self* arises in the same way that any other object or entity arises, namely, in the recursive coordinations of consensual doings, first in the coordinations of doings in relation to doings with the body, and then, in the recursions of the coordinations of coordinations of doings with doings with the body in relation to other coordinations of doings. When we participate in this recursive dynamics of coordinations of doings, there arises in us the special configurations of inner feelings that we now distinguish in the flow of our languaging as *self* and as *our self.*

26 The configuration of recursive coordinations of consensual doings that constitute an object in our coexistence with other human beings is what I call "operations of distinction". So, when someone says that he or she is distinguishing an object or entity of any kind in his or her living as a languaging being, he or she is bringing forth a domain of consensual doings and recursive consensual doings in which that which has been distinguished has presence in a flow of recursive coordinations of consensual doings. And that flow of coordinations of recursive consensual doings constitutively implies an operational-relational matrix of coordinations of doings as a domain of human living in which the participating human beings distinguish entities that could be themselves.

27 *In synthesis:* The *self* is not an entity; it is a particular feeling in a manner of operating in a flow of recursive coordinations of consensual coordinations of doings that involve the distinction of the doer of the doings as the observer of the doings being done. Furthermore, when in the recursions of the distinction of the observer, occur recursive coordinations of the observer doing its doing, the special feeling of *self-consciousness* arises as the feeling of feeling the coordinations of doing that the feeling of observing entails. In other words, self-consciousness occurs as an inner feeling felt by an observer that is seen by another observer (that could be him or herself) in the circumstances of distinguishing him or herself distinguishing him or herself.

28 Whenever a recursion takes place an intrinsically new domain of doings arises, and at the same time a new domain of feelings is lived, which we may live as a completely new domain of meanings in our doings. For example, science, philosophy, theories … technology have arisen like this. Once a new domain of recursive coordinations of doings, and hence, a new domain of reflections in doings, has arisen, our human living changes and we live the arising of new surprising happenings that we do not know immediately how to explain, and we feel that we are in front of a mystery. What we should never forget, however, is that structural determinism is the fundamental constitutive condition of our existence, and new operational domains arise in our living whenever our living becomes associated with recursive process in our doings, and … our thinking and reflecting, and that these are dimensions of our understanding that we cannot forsake if we want to understand our living as human beings.

Final remarks

29 We all know that the result of a process does not participate in the process that produces it as a result. But as we live a culture in which we are accustomed to think in finalistic terms, that is, in a process designed with the purpose of obtaining the desired result, we frequently confuse our description of what we see in the appearance of what happens, with what may be happening that gives rise to such appearance. Thus we frequently treat a process in which we see a purpose as if there were a purpose in the operation of that process. This is what we do in biology when we use teleological considerations to understand the function of some unknown structure in an organism. That way of thinking may be useful for a while to find out how that structure operates in the relational space of the organism that has it, but does not tell us how that structure does what it does.

Note

1. The systemic laws here presented were taken from the essay on systemic and meta-systemic laws published in Ximena & Maturana (2008).

Two Basic Agreements and Two Doubts

Giovanni Pezzulo & Cristiano Castelfranchi
Institute of Cognitive Sciences and Technologies, Rome (Italy)
<giovanni.pezzulo@istc.cnr.it>
<cristiano.castelfranchi@istc.cnr.it>

1 Martin Butz's target article draws a picture of how anticipation has shaped the mind, from simple forms of sensorimotor engagement to higher level cognitive abilities and consciousness. We agree with the author that the "cognitive mind" is the product of a construction of an "inner reality": an internal, endogenous representation of the word that is autonomous from sensory input and the external word. We are also very sympathetic to the author's effort to spell out this con-

structive process in terms of its basic mechanisms. In particular, we like the view that the need to anticipate future events in order to coordinate with them, or to imagine what is not already there, what does not (yet) exist, is a primary pressure, advantage, and mechanism for such an "emancipation" of the mind from perception and reality: a precognition for a truly cognitive mind (Pezzulo 2008; Pezzulo & Castelfranchi 2007). This view is well expressed and argued in the target article.

2 The second very important point on which we agree (and which should be appreciated more in the cognitive sciences) is that this process is also responsible for the development of self-consciousness, and has the potential to explain some of its crucial features. That is, self-representation and self-consciousness may have roots in the cognitive agent's need to locate itself in the anticipated (imagined/simulated) scenario, or in having a third-person-like representation of itself as an agent in the world. We think that the author's account of consciousness phenomena is especially remarkable, for two reasons. First, it is conducted in continuity with an analysis of simpler adaptive and cognitive processes, and this is essential to firmly ground consciousness in nature. Second, it is a rare attempt to scale up ideas and models circulating in motor control, which have the potential to explain the cognitive mind in all its facets.

3 And now, the two doubts we have. One intriguing concept that the author introduces and uses throughout the paper is the idea of an "anticipatory drive," which is described as explaining the systematic tendency to develop anticipatory capabilities that ultimately support goal-oriented action. Although the idea of a common mechanism that explains a multitude of capabilities can be appreciated, it is unclear if the author uses the term "drive" in a literal or metaphorical sense. Should the evolutionary advantage, the selective "pressure" for anticipatory capacities and representation be considered a drive in the sense of Hull (1943) like, for example, the sex drive, or hunger? If this is true, it should be related to some bodily control structure like the other mentioned drives. Another possibility is that the anticipatory drive is an intrinsic motivation (Berlyne 1960), like curiosity, which "pushes" towards certain situations, such as those that are more predictable. If this is true, it should be explained what is the exact nature of such intrinsic motivation. It is also possible that the author uses the term "drive" in a more metaphorical sense. For example, for indicating that evolutionary pressures have selected various specific mechanisms for developing anticipatory representations, and that this advantage also drives (but it is not a "drive") our developmental steps, and characterizes many cognitive and behavioral functions.

4 Overall, we find that the idea of "anticipatory drive" is potentially very interesting, but its precise meaning and (possibly) its "operationalization" are not completely clear to us. This is especially true since throughout the paper the anticipatory drive is assigned diverse functions, such as to "enable us to execute flexible goal-directed behavior," to "control attention and decision making," to "continuously strive[s] to improve predictive capabilities," or to "force[s]" individualization. As discussed before, we fully agree with the author's analysis of the adaptive function of anticipation, which ultimately influences mind design. However, if the author intends to propose "anticipatory drive" as a novel theoretical construct that enters the vocabulary of cognitive sciences (like, for example, the cybernetic notion of "goal" or the control-theoretic notion of the internal model) he should disambiguate its meaning and possibly distinguish its mechanism from its functions and adaptive advantages.

5 Our second perplexity is about the primacy of anticipation over prediction. At the beginning of the paper, the author claims that prediction arose for the sake of producing anticipatory, future-oriented behavior. However, it could be the other way around. One hypothesis that we have put forward (Pezzulo & Castelfranchi 2007) is that anticipatory (and successively goal-directed) capabilities could have been an exaptation of simple predictive mechanisms that were originally required for action control. According to an influential theory of motor control (Wolpert & Ghahramani 2004), the brain makes use of internal (inverse and forward) models for selecting and guiding action. In such a control-theoretic perspective, internal models generate and use sensory predictions for several reasons, such as compensating delays in and filtering of sensory feedback. Our hypothesis is that such predictive capabilities could have been exapted for increasingly sophisticated anticipatory and cognitive uses, since they opened up the evolutionary possibility to anticipate future events and take goal-directed action. Admittedly, neither our hypothesis nor the author's is currently supported by any data and thus should be considered as speculative. In any case, we argue that the author's picture of the development of increasingly complex forms of cognition from anticipatory capabilities is not hindered at all by our hypothesis, except, perhaps, the idea that an anticipatory drive is necessary when (maybe) an exaptation could be sufficient.

Cause and Effect

The Anticipatory Drive and the Principle of Least Time

Samarth Swarup
Virginia Bioinformatics Institute
<swarup@vbi.vt.edu>

Introduction

1 Is there a single unifying principle that explains all brain structure and function? It is a tantalizing prospect, and there have been many suggestions, such as neural Darwinism (Edelman 1987) and hierarchical Bayesian inference (Lee & Mumford 2003), to name just two. Butz proposes an *anticipatory drive* that is postulated to be responsible for brain function and the development of brain structure. It is especially interesting because Butz suggests that the anticipatory drive guides brain *development*, in addition to function. This is an ambitious and provocative proposal, and bears close examination. I focus on just one aspect here: in the spirit of constructivism, I ask, *where is it?*

Locating the anticipatory drive

2 Some human drives, at least, are relatively well understood. Hunger, for instance, is known to be triggered through a signal transmitted from the stomach and liver to the hypothalamus. When blood sugar levels start to drop, this signal causes the hypothalamus to activate hunger-related behaviors, such as food-seeking; and when we have eaten and food starts to move from the stomach to the intestines, another signal causes the hypothalamus to suppress the hunger drive and related behaviors.

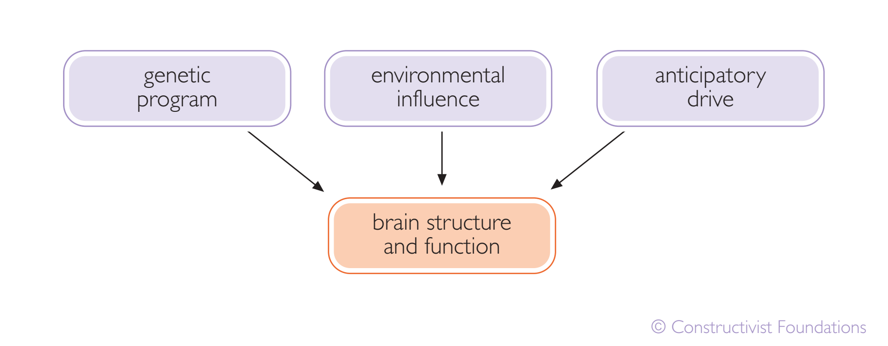

Figure 1: The anticipatory drive is independent of the genetic program and of environmental influence.

3 Can we similarly locate the anticipatory drive? Since the anticipatory drive is supposed to influence brain development as well as function, we are essentially asking about the causal structure that relates four abstract components: the genetic program, environmental influence, the anticipatory drive, and brain structure and function. The simplest assumption is that the first three are independent of each other and together determine the fourth. This is illustrated in Figure 1.

4 This is an unsatisfactory solution, however, because it does not provide a physical substrate for the anticipatory drive. In fact, it seems clear that Butz is not arguing for this view, since he states that a structured environment and a structured body ("morphological intelligence") are pre-requisites for anticipation and the anticipatory drive (§11). This suggests, then, the modification shown in Figure 2 where the anticipatory drive is created through the combined influence of genes and environment, and then, in turn, shapes the development of brain structure and function.

5 In this view, the anticipatory drive would have to be embodied in a structure that develops prior to the brain. If we assume, and the general discussion in the article supports this view I think, that by "brain" Butz is actually referring to the cerebral cortex, and not the brainstem or spinal column, then perhaps the anticipatory drive could be located in these areas. These areas are known to develop very early, mostly during the first and second trimesters of gestation, and govern most autonomous functions such as the heartbeat and breathing, and reflex actions such as grasping, feeding, simple eye movements, etc.

6 The problem with this view is more subtle. It is, at least in some cases, that the developmental influence that leads to anticipatory behavior goes the other way. For example, the deep superior colliculus, which sits at the top of the brainstem, is thought to be responsible for generating eye saccades towards anticipated target locations in the visual field (Anastasio, Patton & Belkacem-Boussaid 2000). This anticipatory behavior depends on being able to correctly integrate multi-modal information in the calculation of target probabilities, which is manifested in a phenomenon known as multi-sensory enhancement. This refers to the fact that multi-modal neurons in the superior colliculus respond much more strongly when they receive input from two modalities simultaneously (auditory and visual, say), than they do to either modality alone. This enhanced response reflects the enhanced probability of a target due to information from multiple channels. Crucially, this enhancement appears only when descending projections from higher areas of cortex, such as the anterior ectosylvian sulcus (AES) and the rostro-lateral suprasylvian sulcus (rLS), reach the superior colliculus, which happens a few months after birth (Jiang et al. 2001). Clearly, in this case, causality is directed from the cortex to the brainstem in the emergence of anticipatory behavior.

7 Thus, it seems that we must make a further modification to our schematic diagram. At least some cortical structures must emerge before the anticipatory drive. This revised view is presented in Figure 3. I believe, however, that this view is also unsatisfactory because it seems to be heading towards a "god of the gaps" argument. I am afraid that the more closely we examine the development of various brain structures, the more structures we will end up putting in the upper "brain structure and function" box in Figure 3, and the fewer in the lower one.

8 There is no doubt that the brain anticipates. The view of the brain as an anticipatory device represents a deep insight, in my opinion – of the sort that could form the basis for

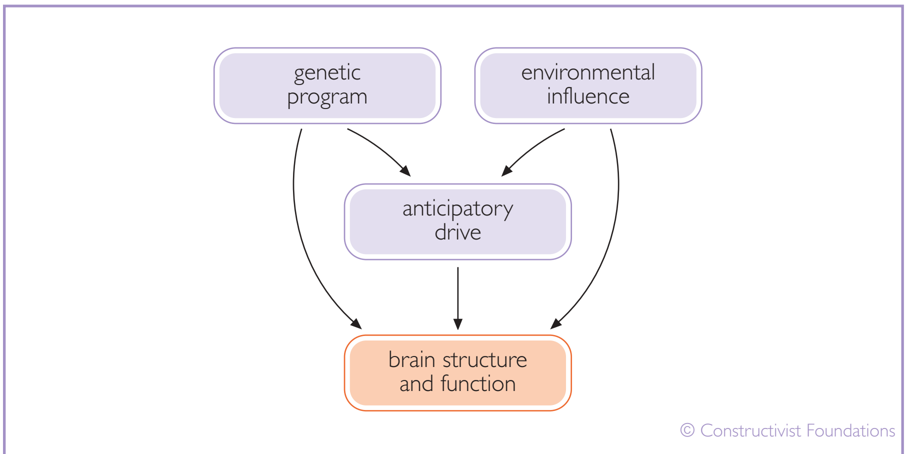

Figure 2: The anticipatory drive is generated by the genetic program and environmental influence, and in turn generates brain structure and function.

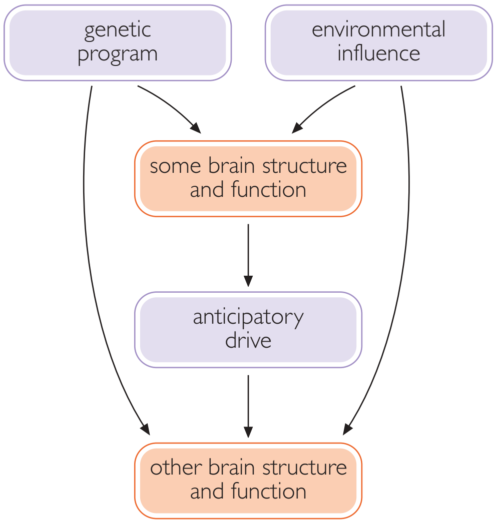

Figure 3: Some brain structures are created by the genetic program and environmental influence. These produce the anticipatory drive, which creates the remaining brain structures and functions in combination with the genetic program and environmental influence.

a unified theory of the brain. Despite this, the evidence for an anticipatory *drive* generating brain structure and function is somewhat tenuous. However, I believe there is a fourth view, besides the three I have presented above. *Assuming* the anticipatory drive exists may turn out to be an excellent guiding principle for inferring structure and function in various parts of the brain. In this view, the principle of the anticipatory drive is true, but acausal. Let me explain this statement through an analogy.

The principle of least time

9 Pierre de Fermat, in 1662, proposed the Principle of Least Time, which can be paraphrased as, *light travels between two points along the path that takes the least time.* This has been confirmed repeatedly (in a slightly revised form) through experiment. It explains, for example, why the surface of a road appears wet in the distance on a hot day. It happens because the air close to the surface of the road gets heated up and becomes less dense. Light travels faster through a medium of lesser density, and therefore light from straight ahead curves downward as it comes towards us (the observer), and makes the road look reflective, or wet. Fermat's principle can also be used to derive the laws of reflection and refraction, among other optical phenomena. In fact, the principle is so well-accepted that it has long been taken as the definition for a ray of light (Schuster 1904).

10 A moment's thought, however, reveals Fermat's principle to be acausal. It determines the path that light will take, based on where it will end up. In other words, when we hear that light takes the path of least time, it makes us ask, *how does it know?* How does it know where it is going, and how does it calculate the appropriate path? The answer, of course, is that it doesn't know. Fermat's principle is more appropriately viewed as an *effect*, i.e., a consequence of a deeper theory (Salmon 1998: 169). In fact, it has been shown to emerge from Huygens' wave theory of light in the classical framework, and from the main principle of quantum electrodynamics in the quantum framework (Feynman 1988). However, it remains a widely used principle in optics for deriving the paths of light rays in many practical problems.

11 I believe that the anticipatory drive may turn out to be like the principle of least time, i.e., acausal, but very handy.

The Role of Sensations in the Anticipating Self

Mathias Osvath
Lund University Cognitive Science
<mathias.osvath@lucs.lu.se>

1 In his target article, Martin Butz introduces a new learning mechanism: an anticipatory drive, enforcing the formation of bidirectional anticipatory brain structures – a thought-provoking idea.

2 Such a pre-disposition for incorporating anticipatory relationships would indeed make sense in the light of evolution. Reliable forecasts about future states are most certainly rewarding in an array of niches. A large number of chemical, behavioral and cognitive mechanisms endorsing proaction have been naturally selected throughout the history of life. It is not unreasonable, it is perhaps even probable, that a learning mechanism such as an anticipatory drive has been selected. The hypothesis of an anticipatory drive is well worth pursuing in specifically designed studies.

3 Having said this, I would like to highlight a mechanism that is vital for the anticipating self: the formation of sensations detached from the current environment. Without a sensing self it would be impossible to project oneself into possible futures of a certain kind – futures that are simulated in the kind of inner world that provides the subject with affective experiences.

4 The author of the target article identifies the above mechanism to a degree, but chooses not to deal with it. In §97 it is argued that the so-called binding mechanism must be involved in subjective qualitative conscious experience, but that the qualia-debate is out of the article's scope. The author is right in discarding this philosophical squabble from the account (in passing, I agree that there is no evidence for conscious experience being immaterial).

5 However, there is a risk of throwing out the baby with the bath water if one reduces the role of subjective conscious experience to the topic of the qualia quarrel. This experience might actually be central to the concept of the anticipating self. The "feeliness" of the self, seeing red and tasting chocolate, seems to be essential to having a feeling of a self, in *being* a self.

6 This is elegantly portrayed by Humphrey (1997, 2006), who argues that the perception and the sensation systems might be distinct and not necessarily dependent on each other. By using examples from blindsight patients (and other rare disorders) he makes a case for the fact that it is possible to have perceptions without sensations (and the opposite). The important point is that if one has perceptions that are decoupled from sensations then no experience of a self in relation to those perceptions exists. When a blindsight subject is forced to report on visual perceptions it is experienced as a complete guess or as if taken out of the blue, something quite unrelated to *oneself*.

7 From this, it could be questioned whether it is actually valid to speak about self representations that are not experienced as *self* representations. To put it in other words, does a non-feely self-representation represent the same self that is felt, or would it perhaps be less confusing to use other terminology? What if the sensations very much constitute the self (as is indicated by Humphrey)? What then is all the rest?

8 This is somewhat off topic and also implied by the author in the discussion of Legrand's (2007b) distinctions. However it is still worth a moment's thought as it might have implications for how the suggested anticipatory drive could be said to be involved in the construction of the *self*. The author is admittedly vague on how the non-feely representations of the self merge together into a subjective conscious experience, into the feely self – they might in fact be distinct non-overlapping systems.

9 Regardless of whether the sensational system is the result of an anticipatory drive or not, it does seem to have point-blank anticipatory value when it is mentally detached from current environmental stimuli. It enables sensations stemming from potential future environments and it produces pre-feelings that can be compared to the current situation and serve as a decision device. Scientists and philosophers from various fields and epochs have considered the idea of an inner mental world that is similar to the real world in that it evokes sensations in the subject. In contemporary science one can identify at least three major directions that are based on empirical research and have started to cross-fertilize one another.

10 One direction is the field of mental time travel, which mainly deals with episodic memories and prospections. This field was founded by Tulving (1972), who also introduced the concept of "autonoetic consciousness" (providing first person perspective on mental episodes). Then there is the neurological approach that, among other things, has invested interest in the so-called wakeful rest state of the brain, a state that is highly associated with the sensed inner world. A seminal work in this line of science is the paper by Ingvar (1979). He also coined the idea of "a memory of the future." A third field is that of affective forecasting, which investigates abilities to forecast future mental states in different scenarios. In the front line of this research stand Gilbert and Wilson (e.g., 2005).

11 The above three directions have somewhat different approaches and address partly different questions. However, they reach some common conclusions. One of these is, perhaps ironically, that humans are inaccurate in matching the construction of an inner world to the past real world or the future real world. Despite this lack of truthfulness, the construction of a sensed inner world appears to be highly adaptive. Another important consensus that could be derived from these research directions is that the inner world probably serves its best function in relation to potential futures, and not to the present or the past.

12 Arriving at the core of the argument: the subjective experience of the self, and perhaps the self itself (whatever it is), is necessary to unlock the inner world of potential futures that impact current decision making in radical ways. If you do not feel it is *yourself* in that future, you will have no reason to act according to the prospection.

13 Furthermore, you will naturally never be able to visit the future with the sensing self if deprived of sensations because *you* would simply not exist in that future. This leads to the intriguing question: is the sensational self mainly an adaptation for anticipation? It might be that an immediate situation does not require the strong sense of a self that is needed for projecting it into a mental future. However, one should bear in mind that even if the sensed inner world is indeed mainly an adaptation for anticipation, it does not follow logically that the sensing self should be such an adaptation.

14 Nevertheless, from an evolutionary perspective, it certainly raises the possibility. The target article hints at the fact that the sensational self indeed is an anticipatory adaptation; if this could be comprehensibly explained by the anticipatory drive, then this concept would have proved its worth to me (my sensing self).

Maladaptive Anticipations

Martina Rieger
MPI for Human Cognitive and Brain Sciences, Leipzig <rieger@cbs.mpg.de>

1 The author highlights the benefits of an anticipatory drive for a variety of abilities and intelligent behaviour, including motor control, cognition (e.g., decision making, information seeking), and social capabilities (§5, §77, §80). However, there are circumstances when anticipation can be maladaptive. In the following paragraphs, the occurrence of maladaptive anticipation will be illustrated in reference to psychological disorders (depression, generalised anxiety disorder, social phobia). It will be shown that anticipation does not always lead to improved control of oneself and the environment and that anticipation is not always beneficial. Finally, the question is raised of whether it is the strength of the anticipatory drive or the content of the anticipations that is the important factor in the development and construction of the self.

2 "Learned helplessness" refers to situations in which a person has learned to act or behave as if helpless, even when they have the power to change unpleasant or harmful circumstances. The concept of learned helplessness was originally invented on the basis of learning studies in animals. In the experiments (Seligman & Maier 1967; Seligman 1975) animals were exposed to unavoidable shocks. Subsequently they were brought into a situation in which they were able to escape the shocks. They were not able to learn this task (control animals that had not been exposed to unavoidable shocks did learn it). This was attributed to the fact that the animals had previously learned that they could not escape the shocks. This behaviour may generalize to situations other than the learned ones. Learned

helplessness has subsequently been investigated in humans and is implied in some forms of depression (a psychological disorder, characterized by feelings of sadness, anger and frustration). It is assumed that a perceived absence of control over the outcome of a situation contributes to depression. The anticipation of not being in control is contrary to what the anticipatory drive stands for: the ability to control oneself and the environment. Thus, anticipations may not always promote enhanced control of oneself and the environment.

3 Everyone has probably at some point in their lives experienced something akin to "worry" – a lasting preoccupation with future bad events, which may or may not occur. Though worry has some beneficial effects, because it enables someone to prepare for negative events (e.g., one starts to search for a new job when one is worried about losing the current one), this is not always the case. Chronic and exaggerated worry (without a substantial cause justifying the degree of worry) is the key characteristic of generalized anxiety disorder. Persons with generalized anxiety disorder may worry excessively about health, money, family, or work, and continually anticipate disaster. The capacity to anticipate aversive events is therefore, on the one hand, important for successful adaptation, and on the other hand also plays a role in the abnormalities that contribute to excessive worry and anxiety. Thus, anticipation may not always be beneficial.

4 Memories of emotional events are enhanced compared to memories of other types of events. Nitschke et al. (2006) showed that brain activation during the anticipation of seeing aversive pictures predicted memory of those pictures after they had been viewed. Anticipation of aversion recruits brain regions that are associated with memory for emotional events, thereby potentially enhancing the responses to aversive events. The act of anticipation may play an important role in how fresh the memory of a negative event remains. Thus, anticipation of emotional events plays a key role in the enhancement of emotional memory, particularly with negative emotions. This mechanism seems to be an important aspect in social phobia, the fear of being evaluated negatively in social situations (for example when giving a presentation). The expectation that something bad is going to happen may enhance the memory of it if indeed a social interaction does not work out as smoothly as one might have wished for. This leads to a vicious cycle, increasing anxiety before and during the next social situation even further. Again, this example illustrates that anticipation may not always be beneficial.

5 Deficits in processes related to anticipation have been proposed for a variety of other disorders, e.g., schizophrenia (Frith, Blakemore & Wolpert 2000), autism (Williams et al. 2001), and alien hand syndrome (Spence 2002). On the one hand, this and the above examples of maladaptive anticipations strengthen the importance of the concept of an anticipatory drive due to its explanatory power. On the other hand, one should be aware that anticipation per se is not necessarily only adaptive, but that anticipations can be maladaptive.

6 One open question with respect to adaptive and maladaptive functioning is whether the construction of the self is related to the strength of the anticipatory drive or to the content of anticipations. It could be that the strength of the anticipatory drive itself is important – excessive worry may be due to a too strong anticipatory drive. The strength of the anticipatory drive could also be irrelevant; rather, it is the content of anticipations that may shape the development and the construction of the self.

Predicting Events Without Miracle Neurons

Towards a Sober Consideration of Brain Data

Ricarda I. Schubotz
Max Planck Institute for Neurological Research (Germany) <schubotz@nf.mpg.de>

1 Butz's proposal is heir to the challenging tradition of conceiving of the brain (and mind) as an anticipatory device. He outlines anticipatory mechanisms referring, inter alia, to the external environment as containing objects with which we learn to interact, and as containing other selves whose actions we learn to understand. Surprisingly, the paper entirely neglects the issue of the dynamic properties of our environment. Focusing on (static, inanimate) objects only, it fails to acknowledge that anticipation becomes especially relevant when things around us change without being under our control: this is when we are forced to adapt quickly to new circumstances. To estimate as precisely as necessary *what* will *when* be *where* is of vital meaning. Surprisingly though, this issue is not addressed at all. Although it is acknowledged that the environment contains dynamics that are to be predicted somehow (§3) and that have to be represented in the brain somehow (§28), concepts about and empirical data on the prediction of external events are not dealt with further.

2 There is a large body of research that uses imaging methods in humans and patient studies to investigate anticipation of environmental change and its relation to anticipation of change induced by ourselves (for overviews, see Schubotz & Cramon 2003; Schubotz 2004, 2007). It shows that change induced in our environment, no matter whether generated by animate or inanimate entities, calls for our particular *premotor* attention. Using abstract stimulus material, it was found that the prediction of external events, even those that are not reminiscent of actions or agents, relies on our "motor system." The pattern of activations suggest that predictive algorithms differ with regard to their neuroanatomical location within premotor sites, showing that spatial, object-based, rhythmical and pitch-based predictions engage distinguishable dorsal-to-ventral premotor fields. Against the background of these data, I have recently outlined the idea that a predictive account of the motor system can be generalized from action to events (HAPEM framework, Schubotz 2007). Accordingly, prediction of events is achieved with the aid of sensorimotor-driven forward models that are housed by the premotor cortex and that are neuroanatomically ordered according to the styles of transformations they describe. It was suggested that they can be applied to actions as well as to any kind of event that happens in the several seconds range.

3 The HAPEM framework suggests a close relationship between anticipatory attention and our ability to control our movements. This seems in line with Butz's account that

stresses that prediction has not evolved for the sake of prediction but for the sake of anticipatory behavior and action (§3). However, this may be a trivial statement as all of our abilities, including perception, have, in the end, evolved for adapted behavior and adapted actions. More importantly though, pointing teleologically to action a priori biases the appraisal of empirical findings and particularly brain imaging findings in a special way, as recently reflected by the great elation for *embodiment*. It has to be kept in mind that it is not at all clear what kind of algorithms we see at work when the "motor system" is engaged in a task: hundreds of thousands of neurons have their place in a single measured voxel and may serve a mosaic of phenomenologically moderately related behavioral functions. Thus we call the motor system "motor system" simply because motor control is one of its most prominent functions – until we see how motor control and further functions are more appropriately subsumed under a new umbrella term – for instance "prediction system" instead of "motor system."

4 Two casual claims of the paper clearly have to be rejected on the basis of neglected imaging data: firstly, that "unlike sensory information sources, motor information activates predictive sensorimotor codes" (§44); and secondly, that "brain modules that are not directly connected to sensory input or motor output will process inherently anticipatory codes" (§28). Provided that the variable meanings of terms such as "code" or "brain modules" are properly understood here, findings suggest that perception suffices to activate "predictive sensorimotor codes" (namely in premotor-parietal loops), and that the "processing of anticipatory codes" does happen inter alia in brain areas that are indeed directly connected to sensory input and motor output (namely in the motor system) (for details, see Schubotz 2007).

5 Strikingly, although the paper refers in large portions to issues and models of motor control, it neglects a function of mirror neurons in motor control that has been put forward as their genuine one. As pointed out by Keysers and Perrett (2004), the network that embeds mirror neurons plays a role in predicting change produced by the animal itself and in the distinction of this kind of change from change induced by another animal (explaining why, for instance, the animal is not frightened at the appearance of his own forelimb approaching a target in front of it). This, in fact, may be an interesting point at which to start speculating, if one wants to, about mirror neurons' contribution to perceiving our own bodies and becoming aware of our "embodied self." As Keysers's and Perrett's paper stresses, however, it seems reasonable to step back and try to recover a more realistic sense of proportion: the interpretative burden on mirror neurons now seems to overwhelm a thin and sober data basis.

6 However, Butz, like many others these days, alludes to these veritable *miracle* neurons as underlying our ability for empathy (§69). It is noteworthy that no one has yet demonstrated any empirical evidence in favor of this claim. Furthermore, *no one has found any direct evidence for the existence of mirror neurons in the human brain*. The only evidence available is of higher metabolism in an area that is suggested to be homologous to the macaque area F5 . Thus, we have an idea, but no data. Mirror neurons belong to a big family of sensorimotor neurons housed by the premotor cortex. They are tuned to our own and others' actions, just as canonical neurons are tuned to objects (Rizzolatti & Fadiga 1998) and other premotor neurons (lacking a catchy name) are tuned to space (e.g., Graziano & Gross 1998). Macaque studies from the last two decades strongly suggest that premotor neurons are generally relevant for all kinds of interaction with our environment, including "other individuals" (§69), be it in the context of action planning or in the context of merely paying attention to our environment (cf. premotor theory of attention, Rizzolatti et al. 1987).

7 Merely as an aside note, the target article's way of using the notion of an anticipatory drive appears at many points in the form of the *breath of life*. To pick out only one of many examples, "a self-representation … allows the anticipatory drive to distinguish self from other" (§63). Similar metaphorical use of neurons and the brain as agents doing this and that (e.g., in §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," or in §65 "mirror neurons distinguish between different behavioral intentions") should definitely be avoided, particularly when we aim to bridge gaps between philosophical, psychological and neurocognitive accounts. When loop-shaped internal model accounts of motor control are discussed and brain studies are cited, it does not seem tenable to speak in a naïve manner about systems in the (brain) system *controlling*, *representing*, *deciding*, or the like; otherwise, we face the homunculus problem and step into an infinite regress. Brains (or neurons) are not persons, nor is the anticipatory drive. Personalization is suspect since it may generate pseudo-solutions when trying to elucidate the function of complex systems.

A Computational Linguistics Perspective on the Anticipatory Drive

Günter Neumann
German Research Center for Artificial Intelligence (DFKI) <neumann@dfki.de>

1 In this commentary to Martin V. Butz's target article I am especially concerned with his remarks about language (§33, §§71–79, §91) and modularity (§32, §41, §48, §81, §§94–98). In that context, I would like to bring into discussion my own work on computational models of self-monitoring (cf. Neumann 1998, 2004). In this work I explore the idea of an anticipatory drive as a substantial control device for modelling high-level complex language processes such as self-monitoring and adaptive language use. My work is grounded in computational linguistics and, as such, uses a mathematical and computational methodology. Nevertheless, it might provide some interesting aspects and perspectives for constructivism in general, and the model proposed in Butz's article, in particular.

2 The understanding and production of natural language is often interleaved in many situations of language usage. For example, humans *monitor* what they are saying and how they are saying it. They already plan and *revise* what they are going to say before they actually spell it out, e.g. in order to reduce the risk of misunderstandings (of course, depending on the degree of attention). Or they try to *control* the generation of un-

ambiguous utterances (presupposing that the underlying message is as clear as possible). They can *adapt* themselves to the language use of others (by mutually synchronizing the individual activation of each interlocutor). For example, in the case where new ideas have to be expressed for which no mutually known linguistic terms exist (e.g., in situations of information exchange between experts and novices), the speaker's adaptability to the hearer's use of language is necessary in order to make it possible for the hearer to understand the new information. Humans are also good at *completing* the production of an utterance that was started by the interlocutor ("Oh, I know what you are going to say!"), and they are quite good in filling gaps in utterances (as presented in some psycholinguistic experiments, or similar language games).

3 It is also a wide-spread assumption that understanding and production share a grammatical database ("the language we speak is the same as the language we understand," as remarked by Pinker 1994). The idea of representing grammatical knowledge only once and using it for performing both tasks seems to be quite plausible, and there are many arguments based on practical and psychological considerations for adopting such a view, e.g., Ristad (1993), Kuhn (2000), Evans et al. (2007). Furthermore, developments in constraint-based grammar theories – due to their declarative and formal status – demonstrate that grammar reversibility is computationally feasible.

4 In my computational model of language processing, I propose and realize (through a concrete computer program) a consequent approach to grammar reversibility through a model for interleaving parsing and generation on the basis of a uniform grammatical processing model. This model uses a reversible mechanism for interleaving parsing and generation in order to model interactions in which both understanding and production take place, e.g., monitoring, revision, and anticipation feedback loops.

5 If we distinguish two principle ways of interleaving, namely where generation is used in support of parsing, and where parsing is used in support of generation, then interleaved parsing and generation means a) the use of one mode of operation for monitoring and controlling the other, and b) the use of structures resulting from one direction directly in the other direction. For example, during parsing of an utterance, generation can already take place for the just-parsed parts, by taking into account the parsing results at a very early stage of processing. Self-control of the parsing process through interleaved generation is also important for handling under-specified or ill-formed input where generation is used to "guess" the missing parts or to perform some sort of repair work (e.g., to "guess" what the ill-formed utterance probably means). During natural language production, interleaved parsing is important to obtain hearer-adaptable production of utterances. The basic task of monitoring is to gain information about processing that is not necessarily obvious, i.e., a device is called up that can make this information available to the speaker or the hearer. Clearly, additional knowledge or preference-based mechanisms are needed for the realization of its full functionality, so that interleaved parsing and generation is only one step in that direction – but it is, however, a substantial one.

6 Note that this idea of interleaved interactions comes close to the idea of the anticipatory drive (see §5 in the article of Butz). For example, if we are in the production mode, then generation follows a goal-directed behaviour by computing possible target utterances from some semantic representations, which are interpreted bottom-up by the parser in order to analyse "how the utterances might be understood or interpreted by the expected audience" (cf. also §14). Note that I am assuming that the interaction actually takes place on the level of the input/output (i.e., on the level of semantics and phonology). I consider this to be a consequence of the assumed modular status of the grammatical system, as discussed below.

7 A major innovation of the interleaved approach is the notion of *item-sharing*, which permits partial results computed in one direction to be re-used in the other direction. This possibility allows an *incremental* self-monitoring process in which partially generated expressions are parsed to identify ambiguities and cause the generator to consider other, possibly less ambiguous, paraphrases without redundant re-computations. Modelling such an interleaved approach on the basis of non-uniform processes is problematic – if not impossible. For example, if two different grammars and processes are in use, additional translation operations are necessary for parsing and generation in order to exchange partial results. Since this is a complex process in itself, not only maintaining two specific grammars but also two different processes, it will be a handicap for an interleaved approach. The item-sharing approach has also been extended with Machine Learning and statistical-based approaches in order to model domain and language adaptation, cf. Neumann (2004) and the references there.

8 Considering linguistic objects (i.e., words and phrases) as utterance-meaning associations (cf. e.g., Chomsky 1995) is widely accepted. Thus viewed, a grammar is a formal statement of the relation between utterances of a natural language and representations of their meanings in some logical or other artificial language, where such representations are usually called *logical forms* (Shieber 1993). Thus a reversible grammar defines a common interface for parsing and generation on the level of strings (I consider a string to be an underspecified normalized representation of an utterance) and logical forms. If the assumption that human communication has access to an *infinite* set of meaningful utterances is true, the grammatical search space can only be defined *implicitly* by finite means. This is why "un-compressing" the search space for some input need to be done *on-line* by a computational and compositional process in order to master the combinatorial power of a grammar.

9 Usually (at least in computational linguistics) the grammar has a modular status. Grammar modularity means that the grammar is not distributed across or shared by different components of the natural language system, but rather is located in a designated area of the natural language system – the grammar module. Other components or processes do not have any detailed grammatical knowledge, and communicate with the grammar module only by its interface levels (usually abstract phonological and semantic representations). Note that grammar reversibility and modularity can also be viewed from the point of view of declarative and procedural knowledge sources, such that grammar reversibility requires a specification of grammatical knowledge that is independent from its actual use either during parsing or generation. Note also that consequently

parsing and generation (considered from a competence view) are non-deterministic processes, i.e., without any further (non-grammatical) information both processes have inherent degrees of freedom; cf. also Shieber (1993). Furthermore, the mentioned properties (reversibility, implicit search space, and modularity) are also important in the context of the "recursion-only hypothesis" discussed by Hauser, Chomsky & Fitch (2002), who claim that recursion (i.e., providing the capacity to generate an infinite range of expressions from a finite set of elements) is the only uniquely human component of the faculty of language.

10 However, at least from a language usage perspective, concrete language utterances seem to be deterministic, i.e., at some point some decisions are made. What is the nature of these decisions, if they are not grammatical? At least two possibilities can be considered: either the decisions are based on preferences (which are learned through past experience) or through control, i.e., explicit strategies that are used to interpret the results of other processes in order to provide feedback. I consider preferences as "un-intelligent" in the sense that they are merely applied (blindly) and control-strategies as "intelligent" because they are applied purposely. Of course, both aspects are somewhat integrated, i.e., language processing is both preference-directed and controlled. It seems that the interleaved approach and the anticipatory drive at least have important properties that classify them as an explicit control-strategy.

11 As already said at the beginning, the proposed computational model is mainly rooted in mathematics and computational linguistics and does not claim any cognitive "realism". However, the realization of the underlying ideas (i.e., grammar reversibility, uniform parsing and generation, self-monitoring) on such a technical algorithmic level requires fine-grained details. Furthermore, the idea of the interleaved approach of parsing and generation is also strongly motivated by the assumption that complex anticipation feedback loops are necessary for the modelling of highly self-adaptive natural language systems. And as such, it might be of interest for the further outline of the model of the anticipatory drive proposed by Butz, especially concerning the aspects of language and modularity. Clearly, the computational model in its current form is realized on a high symbolic level. Probably, it is too high to integrate it directly on a neuronal level. Seen as such, it could be of scientific interest to explore a) how to integrate sub-symbolic approaches into such computational models as I have outlined, and b) how to integrate such complex symbolic interactions into the model of the anticipatory drive.

Objectifying the Subjective Self

An Account From a Synthetic Robotics Approach

Jun Tani
RIKEN Brain Science Institute (Japan)
<tani@brain.riken.jp>

1 The current essay by Butz brilliantly illustrates a constructivist account for one of the essential problems in psychology and cognitive science: that of how the subjectively perceived self can be objectified. His theory stands on so-called "anticipatory behavior" (§4), which is considered to play an important role in learning behavioral causalities in the environment that forms the inner reality during development. Butz links two distinct pathways in brains – a dorsal one and a ventral one (§59) – in which the former constructs a body space based on proprioceptional encoding of body postures and the latter does so for visual categorizations of objects. His interesting argument is that bidirectional interactions between these two pathways initiate the objectification process of the subjective self, especially during tool usage as described in Iriki's studies (§61). During the use of each familiarized tool, a distinct sensory-motor structure appears in seamless coupling with parietal neuronal activities that in turn subjectify the tool usage within the body space. On the other hand, when the tool is detached from the body, the tool that was once subjectified within the body space, is now objectified via visual categorization within the object-centered coordinate system in the ventral pathway. Finally, he postulates that this process of objectification of formerly-subjectified tool usage might lead to the objectification of the subjective self (§89).

2 Being impressed by Butz's psychological account for the process of objectification of the subjective self, I would like to postulate, from my expertise in synthetic neuro-robotics studies, possible neuro-dynamic mechanisms that account for his psychological theorem.

3 Before considering the actual mechanisms, I would like to start by discussing differences between notions of self and self-consciousness. A particular concern is that the state of self-consciousness in the reflective stage might not occur just by being able to anticipate motor-caused sensory feedback. Instead, the self might become consciously aware only when a prediction goes wrong, generating errors.

4 This interpretation of self-consciousness may be supported by Heidegger's (1962) example of the hammer, which is well-known in phenomenology. For a carpenter, when everything is going smoothly, the carpenter and the hammer function as a single unit. But, when something goes wrong with the carpenter's hammering or with the hammer, then the independent existences of the subject (the carpenter) and the object (the hammer) are noticed by the carpenter. Here, the carpenter becomes self-conscious, in the same ways that he or she becomes conscious of the world becoming problematic when things just do not match expectations.

5 Tani (1998) reconstructed this phenomena in his neuro-robotics experiments with emphasis of the cognitive roles of regression for learning from the past and prediction of the future. In this experiment, a mobile robot with vision learned to predict the next landmarks it would encounter while it explored the environment by utilizing a dynamic neural network as a forward model. After a certain period of exploration, the robot became predictable by achieving coherence between prediction by the internal forward dynamics and the environmental sensory feedback given when such coherence broke down intermittently, generating prediction error. In the incoherence phase, the internal process is stressed, with a search for better internal parameter values in order to reduce prediction error; while in the coherent phase everything goes smoothly and automatically without the need to update the parameters. In analogy to the Heidegger's example of hammering, it can be said that the robot became self-conscious only in the incoherent phase.

In other words, the self can exist in the coherent phase, but it can be consciously aware only in the incoherent phase. It is further argued that self-consciousness, which appeared only intermittently in Tani's robot, may well represent the momentary self described by William James (1950). Gallagher (2000) regards this type of self as a minimal self, which is only a momentary, subjective experience of self, which may correspond to the reflexive state of self-consciousness introduced in the current essay. Gallagher (2000) wrote that the minimal self can be developed to a narrative self that is constituted with a past and a future in the various stories that we tell about ourselves. This self-referential nature of narrative self seems to correspond well to the reflective stage of self consciousness in the current essay by Butz. It is also noted that this development from the reflexive stage to the reflective one can be related to the transition from the pre-empirical level to the objective time level in Husserl's theorem on immanent time (Husserl 1964), as will be illustrated later.

6 Now, I will propose possible neuronal mechanisms for extending the reflexive stage of self-consciousness to the reflective one in Butz's terminologies. Although the apparently difficult part is how to objectify the subjective inner reality of sensory-motor experiences, this can be modeled by taking two different neuronal representation approaches: namely, those of local representation and of distributed representation. In the local representation approach, each distinct sensory-motor structure experienced can be embedded in its corresponding local forward model module through winner-take-all (WTA) type competitions with other modules (Wolpert & Kawato 1998; Tani & Nolfi 1998; see Figure 1a).

7 The competition proceeds with a gating mechanism associated with each module. If a particular forward model module is good at predicting the coming sensory flow while generating less error compared to others, the gate associated with this module tends to open more, while others do so less, in the WTA manner. The winning module is entitled to more learning and generation of more prediction outputs for the current inputs. As a result of the competitive learning, distinct sensory-motor primitives, in terms of forward models, are self-organized into corresponding local modules. After this learning, the original sen-

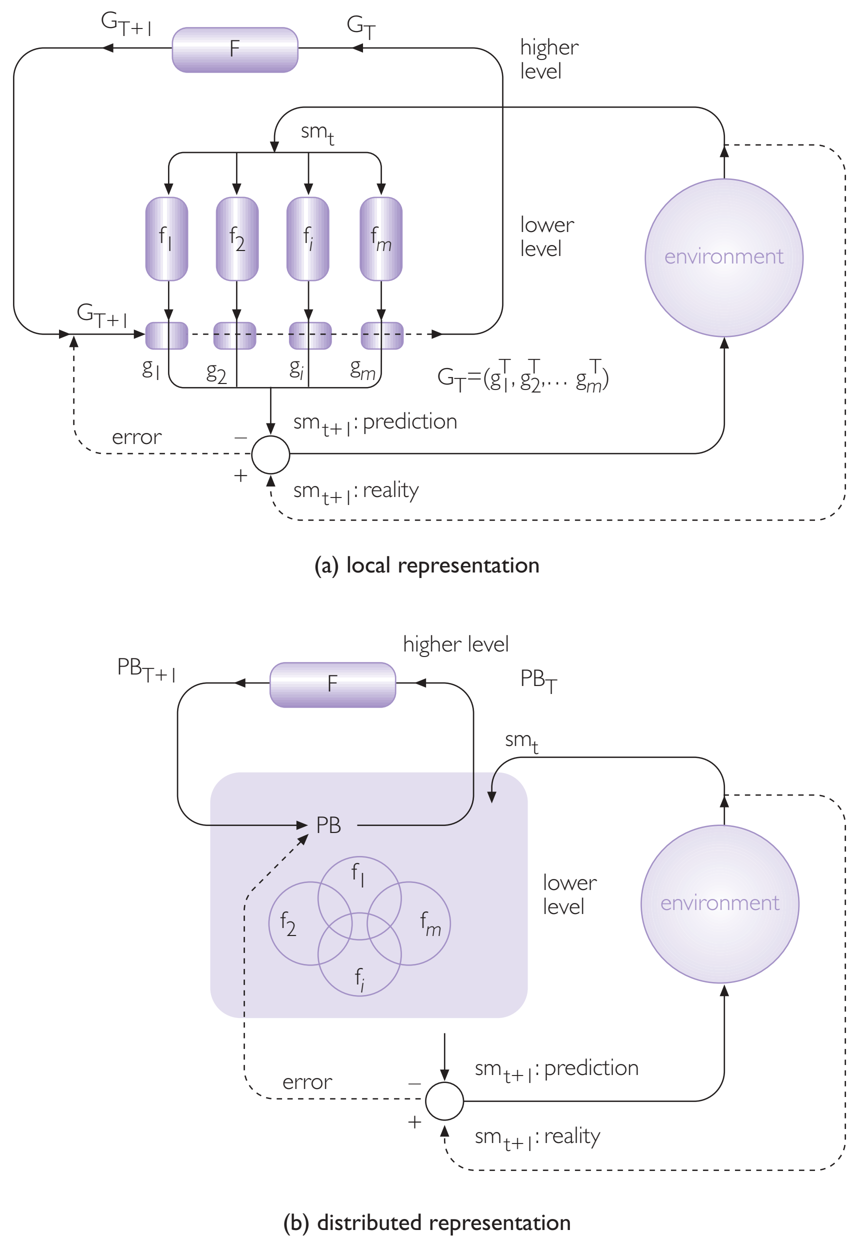

Figure 1: A local representation scheme of the gated modular networks model is shown in (a). In the lower level each *i*-th forward model depicted by f_i competes, using winner-take-all dynamics, to predict the next sensory-motor state sm_{t+1}, while the higher level forward model *F* predicts which gate will open in the next time step in G_{T+1}. (b) shows a distributed representation scheme of the RNNPB, where multiple forward models, f_1, ...*fm*, are distributionally represented in a single RNN with the associated PB vector. The PB vector value is adapted such that the prediction error for the next sensory-motor state sm_{t+1} is minimized, while the higher level forward model *F* predicts the next PB vector value, PB_{T+1}.

sory-motor flow is segmented into sequences of reusable primitives by the gate opening mechanism (Tani & Nolfi 1998). A higher level forward model is now introduced to the system (Tani & Nolfi 1998). The higher level forward model learns to predict the gate opening sequences by observing them. Therefore, the original sensory-motor flow is reconstructed in terms of sequences of pre-acquired primitives. Our essential claim here is that subjective experiences of sensory-motor flow are objectified by referencing them with their corresponding module IDs which are manipulable at the higher level. It is noted that prediction error is an essential drive to articulating continuous sensory-motor flow into objectified primitives. In other words, subjective experience is segmented by momentary self-accompanying incoherence into a sequence of consciously retrievable events that constructs the narrative self. At this very moment, the objective time level might appear from the pre-empirical level (see more precise descriptions in Tani 2004).

8 A distributed representation scheme is now introduced where similar but more psychologically plausible explanations can be made (see Figure 1b). Tani and colleagues (Tani, Ito & Sugita 2004) have proposed a neural network model, the so-called "recurrent neural network" with parametric biases (RNNPB), that can learn multiple forward dynamics models in a distributed way within a single recurrent neural network (RNN). In this model, an RNN is associated with additional units, the so-called "parametric biases" (PB). The PB play the role of bifurcation parameters for the forward dynamics realized by the RNN. By modulating the values of the PB vector, the forward dynamics generates diverse dynamic patterns by going through successive bifurcations. The learning in RNNPB is considered as a process of determining an optimal synaptic weights matrix that embeds all the target dynamic patterns and a set of PB vectors specific to each of the target dynamic patterns. As the result of learning, a mapping between the PB vector and the dynamic patterns is self-organized. In the RNNPB, it is considered that each distinct sensory-motor structure is objectified by its corresponding PB vector value. If a higher level RNN is introduced in order to learn sequences of PB vector shifting, a corresponding switching of sensory-motor structures can be obtained at the lower level that seems to be analogous to the gate switching shown in the local representation scheme.

9 However, there is a distinct advantage to the generalization capability of the RNNPB, which originates from its distributed representation characteristics. In the mapping of PB, if the hamming distance between two PB vectors is short, dynamic patterns generated from these two PB vectors become similar. In other cases, they become different from each other. In this manner, the PB mapping can provide a continuous functional space with generalization, while the gating networks cannot attain such a generalization capability because their functional space is partitioned discretely by a finite set of local modular functions.

10 Such generalization characteristics have been demonstrated by an RNNPB-implemented humanoid robotics experiment in manipulating different shapes of objects (Nishide et al. 2008). As has been said in the current essay by Butz (§51), different objects entail different sensory-motor structures. Nishide et al. (2008) trained two types of mappings where one was a PB mapping to the motor trajectories of the robot arms and the other was from visual images of objects to the PB vector. As a result of simultaneous training of these two mappings, when the robot sees one of the trained objects, the visual mapping generates a corresponding PB vector, which turns out to generate the correct motor trajectory for manipulating the object. When the robot was asked to manipulate a novel object for which the visual feature is between two of the pre-trained objects, the motor trajectory was adequately generated as interpolation between two motor trajectories trained for these objects. This generalization capability for novel objects results from the fact that the objectified entities are still represented in the low-dimensional metric space of the PB. Furthermore, when the robot arm was guided by researchers to move using pre-trained motor patterns, the corresponding mental imagery of the visual object was generated because mapping from proprioception to vision through the PB is established by means of the inverse computation. This might be a possible implementation of Butz's idea (§59) of bidirectional mapping between the dorsal processing, specialized for bodyspace encoding, and the ventral processing for object identification during tool use.

11 In the current commentary, two possible neuronal mechanisms have been proposed to account for the psychological pathways of the development of self-consciousness from its reflexive stage to the reflective one, as proposed by Butz. Although both the local representation scheme and the distributed one are shown to be capable of mapping from subjective sensory-motor experiences to objectified entities, the latter might provide a more psychologically-plausible mechanism because the objectified entities still remain in a metric space. Because these objectified entities that appear in the PB space are not like the arbitrary shapes of tokens (Harnad 1990) but preserve metricity, they could have inherently natural interfaces with the sensory-motor reality in the shared metric space.

Anticipation of Motor Acts

Good for Sportsmen, Bad for Thinkers

John G. Taylor
King's College London
<john.g.taylor@kcl.ac.uk>

1 This paper is full of stimulating and creative ideas. It posits that an anticipatory drive is what guides the development in the brain of a set of internal motor models, specifically a set of inverse and forward models. Through these models becoming increasingly complex, a conscious self develops. This is a simple and important thesis, if true. But is it? As my title suggests, it may be so for sportsmen, with their emphasis on ever more refined motor responses. However, those of a more cerebral nature may find themselves burdened by all those coupled internal motor models and not able to think as clearly as they would like. This is not to say that prediction isn't a useful property to possess, both for finance (especially now) and in one's general living patterns. But the question I wish to consider is: What sort of predictive model can lead to thinking?

2 There is a further difficulty with this paper: it promised an answer to "Why consciousness?" Consciousness is claimed to arise from the increasing plethora of internal

models in ever more complex environments. But even if consciousness does arise that way, what is its function? That is hidden inside the increasing complexity of these internal models.

3 My commentary starts by noting that there can be internal models both for motor acts (with good experimental supporting evidence) and for sensory attention (with less evidence, although the existence of an inverse model to generate attention movement control signals is strong and has been observed in the superior parietal cortex through numerous brain imaging experiments, such as those reported by Hopf et al. 2000). But these latter internal models have been completely ignored in the paper. Thus for the author to say (§98) that he "completely agrees with Taylor's (2002) perspective" is to totally miss that perspective. The author has completely concentrated on motor control internal models, in line with the current fetish with embodiment. However we all know of the experience of being completely conscious even though we are not moving a muscle (or our eyes, in a paradigm using covert attention). The perspective of my paper in 2002 was that consciousness arose purely through a sensory attention control system, without any motor control internal models being involved; this same perspective is developed more fully in further papers (Taylor 2006, 2007 and earlier references contained therein).

4 Of course the motor and sensory attention control circuits must be suitably fused in the brain. This may not be trivial since there is evidence that they have crucial components in opposite hemispheres (Rushworth et al. 1997, 2001). Such fusion of sensory and motor attention has already been included in recent work, such as that on the mental simulation present in observational learning and beyond (Hartley et al. 2008; Hartley & Taylor 2008). However in no case has there been any need for an anticipatory drive to be used to get the fusion going or get any of the forward/inverse models up and running.

5 In my attention control approach, consciousness arises as a two-component set of activities in the brain: 1) Those coding for the stimuli being experienced, such as the smell and colour of a rose, the taste of the delicious glass of wine, and so on; 2) Activities relating to the owner of the experience of these stimulus representations. It is this latter which was mulled over in depth by philosophers of the school of Western phenomenology (Husserl, Sartre, Merleau-Ponty, Henry and so on, well-reviewed and updated by Zahavi 2006). As these thinkers pointed out (and as was described in detail in a number of my papers referred to above), without an owner there would be no experience, no consciousness. This is completely missing from the author's discussion.

6 There is one (among many) feature of great importance in the nature of consciousness: that of "immunity to error through misidentification of the first person pronoun" (Shoemaker 1968). As Wittgenstein pointed out, if I tell you, "I am in pain," you cannot ask me the question, "Are you sure it is you who is in pain?" I just am sure. I know it is I who is suffering the pain, not anyone else, such as you. It is this certainty of oneself (which may disintegrate in some forms of schizophrenia) that can be teased out from the attention–based model I proposed in the above references. The crucial component of this model was the extension of the attention control model to include a copy of the attention control signal, and also an associated forward model for rapid and early prediction of the next attended state.

7 I proposed that the experience of the owner is generated by that copy of the attention movement control signal. The copy was further proposed to prevent distracters getting in the way of awareness arising of any attended stimulus, so that awareness arises with certainty of the attended stimulus being the one desired and expected. Immunity to error about I, acting as the sentry to the gate for access to awareness of content, has thus been attained.

8 Why, then, does consciousness exist? In the author's approach, it would appear to be an epiphenomenon, arising through an accumulation of increasingly complex coupled internal models. In other words a bit of a miracle! So he does not come through on the "Why" in his title. On the other hand my "attention copy" model of consciousness has a very important function for consciousness. In internal motor models there is also good evidence for the existence of a copy of the motor control signal being used in a predictor (forward model) so as to allow for fast error correction. In the attention-based approach to consciousness that I advocate (the attention copy model), the existence of the attention copy signal allows both for fast error correction, mentioned above, as well as for speeding up access to the relevant working memory site of the attended stimulus input. So both components of consciousness – content and owner – function together to speed up the error-free access of the stimulus representation to reporting. There is consequent use of this reportable stimulus representation throughout the higher reaches of the brain. Thus the "I" is a necessary "speeding-up" and error-correcting component for speeding stimulus representations into report mode, with clear survival value.

9 Besides these doubts about the neglect of sensory attention control systems and the existence of an "owner," there is the more basic question: from whence comes this inherited anticipatory drive? It is not one of the primary drives of sex, hunger, thirst and so on. Does it have more of a secondary nature, like curiosity? That also does not seem likely. Indeed this anticipatory drive is a very complex one, needing the internal models to be created before anticipation or forward prediction can be achieved.

10 How does the anticipatory drive help to create the relevant internal models? These can be trained, without the drive, by use of purely unsupervised learning by STDP. What is added to that by the drive? Is there more activation in suitable neural regions by the drive, so speeding up learning? Recent models of observational learning fit experimental data on infant learning without the need for such a drive (Hartley et al. 2008; Hartley & Taylor 2008). So what is the specific experimental data that requires this drive?

11 As can be seen, I am sceptical of the existence of such an anticipatory drive, in the absence of hard evidence. Of course the same scepticism can be directed at my attention copy model approach to consciousness. However that has the immunity to error property in its favour, as well as experimental data of a very specific form from analysis of the attentional blink (Fragopanagos, Kockelkoren & Taylor 2005) and from an early (200 msec) signal in the temporal lobe (following a sharp signal in the parietal lobe some tens of milliseconds earlier), as observed by MEG in attention control tasks associated with the creation of the N2pc (Hopf et al. 2000).

Author's Response

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

Introduction

1 First of all, allow me to articulate my gratitude to all the colleagues that provided this very exciting and diverse feedback. Truly, each and every one has broadened my perspective on the topic and has also made me aware, once again, of how difficult it is – if not impossible – to state things with crystal clarity. In fact, I feel that several of the concerns raised were due to misunderstandings and although I would often even agree perfectly with the objections raised in the commentaries, I see the points raised not really as objections but as issues that should be integrated. Thus, my response in general tries to integrate rather than contrast, while obvious misunderstandings will be clarified.

2 To do so, I will now start with a clarification of the actual intentions of the original article, then proceed with an attempt to further clarify the terminology used, move on to further differentiations of different aspects of the anticipatory drive, consider different aspects of sensorimotor and sensory-dynamic structures, touch shortly on mirror neurons once again, and end with a discussion on the relation with several types of consciousness.

Target article intentions

3 Although I intended to propose that the anticipatory drive is one principle of consciousness, I never had the intention to propose answers to how and why there is consciousness. This is also the reason why the title states "How and why the brain lays the foundations for a conscious self" and not "… constructs a conscious self," or similar. The article was intended to propose how brain structures emerge in which a conscious self may be embedded. The article only touches on the possible mechanisms (binding mechanisms, involved attention, etc.) of how the conscious self is actually embedded (cf. §§93ff) and I feel quite confident that it is simply not sufficiently well understood, as yet, how consciousness actually works.

4 Nonetheless, the article's intention was to propose one of the fundamental mechanisms that leads to the construction of the media, that is, brain implementations of behavioral, perceptual, and further abstracted concepts, in which a conscious self can be embedded – or rather, upon which the mental activity that constitutes the conscious self may work. However, although I believe that the concept of an anticipatory drive has the potential to unify several different brain theoretical aspects, mechanisms, and representations, I do not believe that the anticipatory drive is the only mechanism that controls brain development and functionality – which was also touched upon in the target article when I emphasized embodiment (cf. §§8ff) as well as motivational drives such as hunger (cf. §26).

5 Moreover, I certainly did not want to claim that the anticipatory drive is a special capability of the human cognitive apparatus. In fact, I believe that implementations of the anticipatory drive lie at the heart of the development of any existing brain. In more complex life forms, however, the drive interacts with increasingly sophisticated bodies, already developed brain modules (in somewhat stage-wise phases of development), and complex social and cultural worlds.

6 Finally, I want to emphasize that it was not my intention to show or prove that anticipatory mechanisms (defined in cf. §20) take place in the brain. The evidence is simply overwhelming (cf. §2, §15ff) that brain mechanisms do lead to the encoding of operands of sensorimotor dynamics (and many others) and that brain operations work on these operands to form representations of (immediate and also further distant) potential futures. It should be emphasized again, though, that the evidence is not that of a naïve observer made from daily experience (Maturana, §2, §13, cf. also my discussion on the problem of an observer §12), but was made by careful and diverse psychological and neuroscientific experiments (behavioral and cognitive) and it was modeled by various computational approaches. If this is denied, then the argument would at least require a complete theory as to how these phenomena may come about with alternative, computationally specified mechanisms.

Terminology

7 Several commentaries hit the terminology problem and my inevitable imprecision in terminology. Particularly, there were elaborate concerns with the usage of the terms "representation," "code," and "inner reality."

What is "representation"?

8 In particular, Bettoni (§§2ff), but also von Glasersfeld (§4) and Schubotz (§7), have criticized my loose usage of the term "representation" in various contexts. Bettoni (§§4–5) puts forward several quotations by von Glasersfeld concerning the problem that representations are generally considered structures that are isomorphic to the original. The general understanding of such "represented" structures, however, may imply that the structures must be real in some form. While this concern might be strongly important for radical constructivists (is there a reality that goes beyond our bodily confines?), I do not feel the same concern. That is, I believe that it is not essential for my arguments whether the perceived reality exists independently from our perceptions or not. What is important, however, is which representations of perceived sensory and sensorimotor correlates form and are differentiated during learning and development. Moreover, I agree that I should have better distinguished which representations are implementations of operations and which are implementations of operands, to which operations can be applied (§10).

9 For clarification purposes and further elaborations on "representation," I feel the need to clarify the highlighted formulations:

- Substituting "inner representations" with "mental constructs" (Bettoni §7) may trigger mental associations that are, in fact, slightly closer to the point since "representations" was meant in a rather broad sense that could also include, for example, motor control programs and implementations of dynamic processes in general (as put forward perfectly by Schubotz §§2–3).
- "The anticipatory drive leads to … the generation of mental structures that refer to properties of the self and that distinguish the self from the other." (cf. §4)
- "Moreover, the suitability of the inverse structures strongly depends on the particular implementations of state and goal concepts." (cf. §24)

- "The motor cortex encodes (amongst many other things) body postures." (cf. §40)
- "For more elaborate object concepts to emerge, more complex interactions with objects will be necessary." (cf. §51)
- "…, the anticipatory drive as the basic learning mechanism that underlies brain structuring has now created brain modules and mechanisms that include implementations of various concepts of the self." (cf. §81)

Certainly I would be interested in discussing these formulations further – as they never can be absolutely precise.

10 Another important point about "representation," however, comes from a developmental perspective, too. The concept of object permanence only arises at rather advanced developmental stages (Langer et al. 2003; Piaget 1975). Thus, representations of items also emerge and structure themselves during development by the experience that things do usually not simply disappear, even if they are not perceived any longer. Thus, when talking about the presence of a squirrel (Glasersfeld §3), it may not only be the seeing of the squirrel but also the knowing of its *existence*, that is, the assumption of its usual (temporarily extended) permanence and activity in the world, even once it is out of sight.

What is "code"?

11 Somewhat similarly to "representation," von Glasersfeld raises the concern that the terms "code" and "information" are used too loosely and without further specifications (§4). Von Glasersfeld puts forward that "code" means an item or list of items that are semiotically linked to something else, something to which they are not otherwise related (§5). Being a computer scientist, I think about codes in terms of structures that can be used to activate programs, that is, mechanisms, that execute the commands in the code. In this sense, codes are particularly linked to something by being an implementation of properties of this something else. For example, sensorimotor codes may specify how perceptual input changes when I move my body.

12 Von Glasersfeld (§6) particularly queries the following formulation (cf. §44): "Unlike sensory information sources, motor information activates predictive sensorimotor codes, which predict changes in body perception that are dependent on the executed motor commands." Since this sentence was never meant to exclude the possibility that sensory information alone can also trigger sensorimotor codes (Schubotz §4), I should have better formulated this by stating, "Unlike sensory activities, motor activities directly activate predictive sensorimotor codes that invoke predictive representations of changes in body perception, dependent on the executed motor commands."

13 In conclusion, von Glasersfeld requests specific neurobiological definitions of the terms "code" and "information." This, however, is currently impossible since it is still completely unknown what exactly neurons encode and which algorithms are actually at work, as also pointed out by Schubotz (§3). From an anticipatory drive perspective, codes in the brain may be regarded as embodied knowledge about how future perceptions and internal body states (such as homeostatic states) may change. Activated codes, that is, information about the current state of affairs relevant to the body, consequently inherently co-activate potential future states of affairs. It should be remembered in general, though, that regardless of whether I have named them "codes," "representations," "information," or similar, all of these are clearly confined to the experiential realm of the acting and observing agent (Bettoni §5, Glasersfeld §3).

What are "inner realities"?

14 A slightly different consideration requires my usage of the term "inner reality." Stewart suggest that my stance must be consequently that of an "internalist" (§1). Here it seems we are falling into another interpretation trap. Maybe the right question to ask for clarification would be, "Where does the "inner" start?" Certainly, it is not restricted to the confines of the brain, but it starts with the body and also with the environment, with which the body interacts. Considering the comparison between leg activity versus walking and brain activity versus cognizing as an example (§3,4), leg activity is embedded into a very complex system with antagonistic muscles, tendons, bones, surfaces on which walking takes place, etc. As I have argued, brain activity is also embedded into an even more complex body system with particular body morphologies, genetically predisposed learning and activity patterns, hierarchies, and modularity. Thus, I want to emphasize that my stance is also neither "internalist" nor "externalist" (Stewart §5), but rather that of an "interactionist." In fact, I would rephrase Stewart's statement that "reality […] is *co-constructed* in the interaction between organism and environment" (§8) and state that our inner realities are *pro-constructed* both during and for the interaction with the environment.

15 Since inner realities are thus grounded in brain–body–environment interactions, they can never be totally autonomous from sensory input or the external world (Pezzulo & Castelfranchi §1) because they will always be embedded in structures whose primary focus lies in the control of body–environment interactions. However, during meditation it is clearly possible to detach the self from the present and current body awareness, but, nonetheless, as proposed, even these detached thoughts must ultimately originate from codes of sensorimotor interactions and dynamics. Thus, cognition is embodied into a particular body morphology, which not only determines particular sensorimotor patterns, but also purely sensory dynamic patterns, such as the smell and look of a rose (Taylor §5). In the same vein, I neglected the appreciation of "dynamics" in these inner realities, most likely because I pre-assumed that sensorimotor codes are inherently dynamic codes. However, it should be acknowledged again that perceived, purely sensory dynamics can also electively invoke (somewhat matching) dynamic sensorimotor structures (Schubotz §4).

16 Another important point put forward by Osvath (§3) and also, from a different perspective, by Rieger (§§1ff) is that inner realities do not necessarily match actual reality (whatever the latter might be). Osvath puts forward several indicators that constructed inner realities, such as beliefs about the future or confabulations about the self, are purposeful, rather than truthful – which is in accordance with an anticipatory drive at work. This "purposeful," however, is of course very hard to prove since the actual real purpose, which drives the self and also the construction of inner realities, has too many facets (working on relaxing muscles or the mind, satisfying motivational needs (hunger, thirst, etc.), finding an acceptable place in society, maintaining a consistent inner reality, etc.). Rieger, on the other hand, points out that inner realities

can also be detrimental, which manifests itself, for example, in psychological disorders (§1). Feelings of helplessness may indeed stem from inappropriately tuned inner realities. In this respect, I perfectly agree that anticipation may not always be beneficial (Rieger §3; cf. §5 and also the recent review on benefits and possible drawbacks of anticipations in Butz & Pezzulo 2008). Still, I was pleased to learn that it has even been shown that inaccurate anticipations can actually lead to vicious cycles in behavioral patterns (Rieger §4), which also strengthens the point that the anticipatory drive is at work (§5).

The anticipatory drive revisited

17 Despite several attempts and a full section on the anticipatory drive (cf. §§21–29), I did not succeed in perfectly clarifying what the anticipatory drive exactly is, where it may be located in the brain, and where it comes from (Pezzulo & Castelfranchi §4; Swarup §3; Taylor §9). The term "drive" should not be too strongly associated with "sexual drive" or "hunger" or "thirst." These drives I consider as distinct mechanisms, which play their important parts. The anticipatory drive, however, as Pezzulo & Castelfranchi suspect, is meant rather in a metaphorical sense (§4). The anticipatory drive refers to the driving force that controls brain structuring during development and learning (the main point of the paper), and it is also involved in guiding actual brain activity (a side point of the paper, which needs to be further elaborated on in the future). The target article admittedly uses the concept of an anticipatory drive in many forms; however, the tendency to form associative forward-inverse structures also seems to be ubiquitously present in the brain. Thus, while considering the "anticipatory drive" as the "breath of life" (Schubotz §7) was never my intention, I strongly believe that it is an important concept, with which the emergence of many structures and functions of the brain can be explained.

18 Pezzulo & Castelfranchi also suggest distinguishing between the mechanism, the function, and the adaptive advantage of the anticipatory drive. Although I wish to be able to do this in more precision, it remains a big challenge and would be beyond the capacity of this journal article. This is not only because of the current lack of knowledge in science but also because the anticipatory drive interacts in different brain modules with different pieces of sensory, motor, sensorimotor, and memory information structures and activities. In general, the mechanism biases learning towards the generation of suitable forward-inverse structures (operands) for efficient and flexible interaction with the environment. Moreover, it is part of the operation principle. In this case, the anticipatory drive can have various beneficial effects on behavior, which include increased stable, flexible, and adaptive behavioral control, interactions with objects and tools, as well as social interactions including communication (cf. §5). Thus, I had proposed that the anticipatory drive has definite positive effects on, if it does not principally direct, behavior, in which I include overt behavior, such as body control, but also covert mental behavior such as (re-) directing the focus of attention (as suggested by Taylor §8) or controlling cognitive processes during speech comprehension and generation (Neumann §§2–6).

19 Maturana suggests that anticipatory mechanisms and the anticipatory drive stand in contrast with his theory on "structurally determined systems" (§5, §18 – the principle of the cosmos) that are operating "in dynamic structural coherence with the medium in its niche" (§18). However, Maturana puts forward neither any actual contrast nor any differentiations between anticipatory mechanisms and his nomenclature. I did not find any argument in his commentary that shows that there cannot be any anticipatory drive except for his direct statements (without evidence) against it. Seeing that control theory, too, shows that predictive control can be highly effective, I do not see any reason why organisms may not have evolved an anticipatory drive that realizes the maintenance of structural coherence by the support of anticipatory processes (and bear in mind the overwhelming evidence that this is actually happening).

20 Due to the various functions and various qualities of the mechanism, which depends on the types of inputs and outputs processed, I refrained from proposing actual algorithms that can realize the drive. Nonetheless, I am grateful that Tani suggests two recurrent neural network algorithms – one local dynamic gating algorithms of local experts (§7) and a more powerful distributed dynamic representation scheme, which is realized by recurrent neural networks with a parametric bias (RNNPB, §§8ff). Both may realize the emergence of the proposed object interaction codes and differentiate between different objects by implementing a type of anticipatory drive that enforces the distinction of different sensorimotor dynamics for the generation of accurate predictions. However, since RNNPBs may preserve the metricity in the dynamics of different object interaction patterns, RNNPBs may additionally offer a natural translation of dynamic patterns onto a lower-dimensional state space, as suggested by Tani (§11). Taylor suggests an associative learning algorithm through time, that is, spike-time dependent plasticity (STDP), which he contrasts with the anticipatory drive (§10), but which I would consider another potential implementation of a type of anticipatory drive. In fact in §23, I propose that the anticipatory drive biases learning towards the formation of "associative relations over time," which STDP essentially realizes.

21 In sum, while I would not call the anticipatory drive an actual acausal mechanism in the strong sense, I agree with Swarup (§10) that the details of the actual mechanisms in the brain that constitute the anticipatory drive still need to be properly understood and implemented on computers for modeling purposes. Certainly, highly potent implementation approaches can be found in RNNPBs as well as in STDP (the former using forms of back-propagation learning while the latter using purely associative learning). However, I hope that most readers will agree that the concept of an anticipatory drive per se is a handy concept (Swarup §11).

Sensorimotor, attentional, and sensory dynamics

22 I have to say that I was rather shocked when I realized that I had not cited Schubotz's work in the original article (since I know it well and it fits very well) and I am grateful to her for pointing out her work. It is definitely an important point that actions can be generalized to events, types of actions, and types of dynamics (Schubotz §§2–3). I tried to hint at these issues when I detailed the importance of learning body control and particularly also sensorimotor (dynamic) forward models on how motor dynamics change sensory inputs (cf. §§35–38, §§43–46, §§48). Sensorimotor structures may sound like static, state-action-

effect structures, but they comprise many rather fluid, dynamic structures.

23 I also stated one of the fundamental reasons for learning sensorimotor (dynamic) models of the own body: "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." (cf. §35). Thus, I perfectly agree that the main issue in order to realize goal-directed behavior is to first know what will when be where (Schubotz §1) – and I regret that I did not succeed in making this sufficiently explicit in the target article. In sum, encodings of sensorimotor dynamics are not exclusively active only when corresponding motor activity occurs, but they are predominantly shaped dependent on motor activity during learning and development.

24 While dynamic sensorimotor models are certainly structured by the anticipatory drive, I have admittedly neglected the importance of attention-dependent anticipatory models. Taylor suggests that I have actually totally missed his perspective and "completely concentrated on motor control internal models, in line with the current fetish with embodiment" (§3). Again, I would like to adhere to my integrative standpoint and would first like to point out again, as just discussed, that even purely sensory dynamic activities are inevitably linked to sensorimotor codes in premotor areas (see Schubotz §§2–3). Moreover, I would like to point out also that clearly all sensory experiences are embodied and so any attention processes that inevitably must work at least on some processed form of sensory information (if not sensorimotor) are also confined to the experiential realm and thus embodied.

25 More importantly, though, it has been shown that eye saccades are preceded by an attention shift to the location that the eye will focus on next (cf. §54 and also Swarup §6). Thus, attention and motor control appear to be strongly intertwined. To investigate this correlation further, it would be interesting to evaluate the ease of shifts in attention: if the anticipatory drive has shaped brain structures, shifts in attention should be more easily executable the more the shift is natural, that is, the more easily it can also be realized by types of body movements (for example, visually by an eye saccade). However, in our highly developed conceptualizing and symbolizing brains (mediated by language, writing, complex social interaction, etc.), attention shifts beyond the motor capabilities (taking different, much more abstracted routes) are certainly also possible. I tried to elaborate on such concepts in the target article in §69ff.

26 There is even further evidence that attention and perception are intertwined with motor control. For example, it was shown that target sizes (such as a softball or a hole in golf) are judged bigger, the stronger the current performance of the player (Witt et al. 2007; Witt & Proffitt 2005). Most recently, Witt & Proffitt (in press) showed that distances are also judged dependent on the current motor behavior available. For example, if participants had to reach a distant target location (more than one arm length away) and had a tool with which the location could be reached, the distance to the target was judged shorter than when the tool was not available. This was even the case when the tool was not held but the subjects were merely instructed to imagine using the tool to reach the target. Another recent study has shown that action preparations, such as the preparation of a power grasp or a precision grasp, bias selective attention in a change detection paradigm (Symes et al. 2008).

27 In sum, while attention-controlled dynamics in the brain are certainly a highly important aspect, which may actually also constitute aspects of self-consciousness (cf. §98), the anticipatory drive also causes the creation of the structures and mappings that link concepts and structures together simply because they can occur in succession and are correlated in some way. This is why even inanimate sensory dynamics can invoke correlated activations in the premotor cortex (Schubotz §2). Thus, I would be very happy if Taylor would agree that his theories on sensory attention may not be as far away from the proposed anticipatory drive after all.

Mirror neurons, empathy, and language

28 I suspected that due to the current hype over mirror neurons, my inclusion of mirror neurons in the construction of the structures for a conscious self would be criticized (Schubotz; Stewart §2). A couple of things have to be clarified in this respect.

29 First of all, I admit that I often tend to use personifications of mere correlates. For example, I state that "… mirror neurons distinguish between different behavioral intentions" (cf. §65; as pointed out by Schubotz §7). With this admittedly slightly sloppy use of language, I never wanted to implicate that the neurons are actual agents or are the source of control. Rather, they are embedded into brain activities and reveal particularly interesting differences in measured activities. In particular, it was shown that differences in mirror neuron activities correlate with differences in current behavioral intentions. It remains to be disputed in which cases more precision may be sacrificed for readability purposes.

30 Regardless of the precision dispute, I perfectly agree with the objection of Stewart (§3) stating that "What is done in current neuroscience is to correlate differences in mental activity with differences in brain activity […] the temptation is strong to believe that we are actually seeing mental activity going on. However, correlation is not cause; and it is important to resist the temptation." The point is, though, that the detected correlates suggest that the neural activities in the (embodied) brain distinguish between the discussed concepts. This proves that explicit encodings of such concepts exist in the brain. The concept of the anticipatory drive can intuitively explain how such encodings may emerge in the first place. For example, representations of the intentions of others exist because it is vital to comprehend the behavior of others in order to be able to improve mutual interactions.

31 Furthermore, it is important that intentions of others are represented within the same neural structures that represent own behavior. This, again, was pointed out perfectly by Schubotz (§5), stating that mirror neurons are involved in predicting changes that are the result of own motor activity – so that we are, for example, not surprised when we move our own arm. However, this was the whole point of discussing mirror neurons in the first place: Before introducing mirror neurons, I pointed out the importance of sensorimotor self representations, which are necessary to achieve behavioral competence (cf. §§35ff). Since neural correlates of (usually intentional) interpretations of the behavior of others do recruit such sensorimotor self

representations – then becoming mirror neurons that represent both types of (intentional) behavior of the self, and similar types of observed behavior of others – it becomes necessary to develop representations that allow a distinction to be made between mirror activities that occur due to own behavior and similar activities that are due to the observed behavior of others (cf. §§63ff).

32 Admittedly then, it is still not clear to what extent mirror neurons per se are necessary to be able to experience empathy, or rather, to be empathetic (Schubotz §6). However, there are several lines of research that support this claim (cf. §79). Moreover, even if no single neurons can ever be identified that distinguish between (that is, whose activity patterns correlate with) different emotional behaviors of others, we know that we do empathize (at least sometimes). Thus, there must be sorts of activities in the brain that associate observed emotional patterns of others with own emotional capabilities. This alone proves the point in question and requires, consequently (due to the anticipatory drive), that the emotional self needs to be distinguished from the observed emotional other in some way.

I am more than pleased that the only commentary on the computational language part of my argument basically agrees with it and, even more so, offers a theory that supports the argument from the computational linguistics side (cf. Neumann). I would like to highlight a couple of issues that Neumann points out in his commentary, nonetheless: (1) The presence of forward-inverse processing structures in the language facility improves communication bidirectionally – for phrasing own utterances and for comprehension – while using the principle of item and grammar sharing (§§2–7). In fact I believe that such structural sharing forward-inverse representations may be the key to flexible and adaptive behavioral and attentional decision making and control. (2) Redundancies or alternatives (inherent degrees of freedom) during language parsing and generation allow proper language adaptation dependent on social factors (§9), and parsing and generation are also mutually exploited to support the other (§5). That is, the better communicators we are, the more we (a) consider and adapt a mental model of the partner in a communication and (b) use this model to constrain the inherent degrees of freedom in order to improve mutual understanding during communication (during language parsing and generation). (3) Also, the distinction between blind preferences, which stream behavior on a stereotypic path, and intelligent control strategies, which guide goal-oriented behavior (§10), is certainly not only present in language, but also in other behavioral patterns. This has been, for example, suggested by Möller and Schenck (2008), who use forward simulations for categorization purposes and inverse control models to stream long-term predictions. (4) The symbol grounding problem, that is, neural symbolization, remains the biggest challenge (§11). Here, I believe that sensorimotor structural grounding can be expected to be most fruitful (cf. §76). In fact, this grounding may actually lead to the natural emergence of a language grammar. If this could be verified, then it would actually be proven that there is no abstract "universal grammar" but that universal grammatical properties only exist due to the embodiment of language.

33 Grounding symbols on experienced sensorimotor flow and bidirectional parsing-generation mechanisms is, from my perspective, actually in coherence with Maturana's statement that "objects, entities, notions, ideas, concepts etc, arise as coordinations of coordinations of doings" (§22). Grounding words on the available dynamic sensorimotor representations essentially grounds concepts of words, objects, etc. on coordination codes, which are sensorimotor codes. However, Maturana does not put forward any developmental pressures (such as the proposed anticipatory drive) that may lead to the formation of such structures – especially also with respect to the formation of "special configurations of inner feelings" (§25). The further elaborations on types of reflexive and reflective self consciousness below may help to differentiate these special configurations of inner feelings more conceptually.

Consciousness

34 I think it should have became clear that I did not make claims about how consciousness works but rather which mechanisms may develop the media in which consciousness is embedded. Thus, most of the following points of discussion are further thoughts on aspects of consciousness.

The reflexive self and consciousness

35 Osvath makes the case for the phenomenon that there are perceptions that we do not become conscious of (§6). For example, he points out that there are blind sight subjects, who perceive stimuli for action but do not have conscious access to them. They can, for example, directly insert a letter into a letter box in front of them although they are not able to report the orientation of the slot. This suggests that there are independent systems for conscious sensation and unconscious behaviorally-relevant perception (§§7–8). Other studies have also shown that there are differences in, for example, weight estimates between behavioral components and conscious weight judgments using the size-weight illusion as their paradigm (Flanagan & Beltzner 2000): while grasping behavior adapts appropriately to weight knowledge, weight judgments stick to the size-weight illusion.

36 On the other hand, there are many studies that suggest that consciousness is somewhat fooled by behaviorally-relevant clues. As already discussed above, it was shown that the intention, or even only the imagination of using a reaching tool, leads to shorter distance judgments compared to if tool use was neither intended nor imagined (Witt, Proffitt & Epstein 2005; Witt & Proffitt, in press). Also, it has been shown that attention can be biased by behavioral intentions: dependent on the intention to execute a precision or a power grasp, changes in smaller or larger objects are better detected, respectively (Symes et al. 2008). Further studies in this direction will be necessary to disentangle the causality in the observed influences and the exact type of influence. Nonetheless, the experiments strongly suggest that conscious sensations can be influenced by behavior intentions.

The frame problem

37 While none of the commentaries mentions the frame problem explicitly, several discus how many representations are parts of consciousness. Stewart talks about contextual conditions and asks about the location of consciousness. He proposes that the location is nebulous because it is distributed not only in the brain but also in the body and the environment interacted with (§§6–7). Rieger distinguishes between the strength of the anticipatory drive and the content of anticipation

as key factors for the construction of the self (§1, §6). Osvath emphasizes the importance of the capability to detach sensations from current perceptions in order to develop a sensing self (§3). Moreover, he asks if the sensing self may mainly be an adaptation for anticipation (§4).

38 In all these considerations, the frame problem comes into play. That is, how many and which neural representations are activated by the sensing self? And when accepting that most of these representations are dynamic sensorimotor (in various abstracted forms), then it also needs to be considered how far the sensing self will look into the future. In this case, the confidence in predictions must have a strong influence. It has been shown that confidence estimates do play an important role when multiple sources of information are combined in the brain (cf. §50). Moreover, it is likely that the feeling of being in control is an important aspect of self consciousness. Rieger points out that a perceived absence of control over the future may lead to depression (§2). Thus, it appears that an important part of the sensing self is the control of which possible futures are pre-activated and considered for guiding behavior. In retrospect, it becomes important which potential (controlled) futures are actually imaginable.

39 Both points lead back to the available anticipatory representations and the anticipatory drive plus attention (Taylor §3), which control the activity flow within the representational structures. Taylor poses the problem of how it is that we ourselves are always (if things do not go wrong) and inevitably the owner of our conscious experiences (§§5–6). With respect to his attention control system, he proposes that a copy of the attention control signal, which allows for rapid and early predictions of the next attended states (§6), may generate the experience of the owner (§7). Again, I wholeheartedly agree with this description and would like to emphasize that I pointed out that the anticipatory drive has positive influences on attention, including filtering and predictive attention (cf. §5 and §30; see also discussion above).

Reflective self-consciousness

40 Reflective self-consciousness may be termed "coordinations of doings that involve the distinction of the doer of the doings as the observer of the doings being done" (Maturana §27). This distinction was already put forward by Kant (cf. my §82). The target article discussed how distinctions of the inner self such as the observing self and the observed self may be possible in our brain-body-environment coupling self-systems (cf. §§82ff).

41 I agree with Osvath that in order to realize such reflective stages of self-consciousness, it seems particularly important that the sensing self is able to detach itself from current sensations. Language might be very helpful in this respect as well as the representations that distinguish the self from other environments discussed in the target article. The narrative self (Tani §5) may play an important role in considering future states and imagining future possible interactions within the environment as well as remembering past episodes.

42 Interestingly, Tani takes this thought one step further and discusses notions of consciousness awareness. While in coherent behavioral phases everything goes as planned almost automatically, once things go wrong, surprise mechanisms kick in and the system becomes consciously aware of the current error (§5). The notion of surprise, which depends on differences between predicted and actual sensory feedback, is certainly of high importance – also as a means to draw attention (again confirming the strong correlations between anticipations and attention). While these notions of surprise and unsatisfied anticipations may be an important mechanism that invokes types of conscious experience, conscious awareness is probably not only dependent on the notion of surprise. As Taylor points out, self consciousness can be active even when not moving any muscles. Thus, there are further states of consciousness that should be distinguished.

Conclusion

43 Many of my responses have tried to clarify which type of representation, code, activity, or even consciousness I was referring to in the target article. As offered by Bettoni (§14), it might in fact be fruitful to look through the target article in even further detail and rephrase ambiguous or even misleading passages. However, it seems to me that part of the problem lies in the available vocabulary. Discussions on consciousness may need to differentiate types and states of consciousness and give each of the identified types particular names. I suggest utilizing the prerequisites in structures (operands) and mechanisms (operations) that are necessary to invoke particular types of consciousness as a distinguishing criterion (such as surprise mechanisms that can lead to types of conscious awareness). In doing so, I suspect that the anticipatory drive and particularly the consequently emerging anticipatory structures and mechanisms will play an important role and hope that the target article can serve as the basis for such an endeavor.

Acknowledgments

I would like to thank Alexander Riegler for his immensely careful editing, coordination, and communication during the production of the target article and the response. Moreover, I would like to thank again all the commentary writers and reviewers for their contributions. I hope I managed to consider most of the raised points to an accurate and satisfactory degree. Last but not least, I would like to thank all my colleagues at the Department of Cognitive Psychology – especially for suggesting additional relevant literature as well as for their comments on earlier drafts of the target article. This work was supported by the Emmy Noether program of the German research foundation (grant BU1335/3-1).

Collected References

Ach N. (1905) Über die Willenstätigkeit und das Denken: Eine experimentelle Untersuchung mit einem Anhang über das Hippsche Chronoskop. Vandenhoeck & Ruprecht, Göttingen.

Anastasio T. J., Patton P. E. & Belkacem-Boussaid K. (2000) Using Bayes' rule to model multisensory enhancement in the superior colliculus. Neural Computation 12(5): 1165–1187.

Andersen R. A., Snyder L. H., Bradley D. C. & Xing J. (1997) Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annual Review of Neuroscience 20: 303–330.

Arbib M. A. (2001) Co-evolution of human consciousness and language. In: Marijuan P. C. (eds) Cajal and consciousness: Scientific approaches to consciousness on the centennial of ramón y cajal's textura. New York Academy of Sciences 929: 195–220.

Arbib M. A. (2002) The mirror system, imitation, and the evolution of language. In: Dautenhahn K. & Nehaniv C. L. (eds.) Imitation in animals and artifacts. MIT Press, Cambridge MA: 229–280.

Arbib M. A. (2005) From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. Behavioral and Brain Sciences 28: 105–167.

Baars B. J., Ramsøy T. Z. & Laureys S. (2003) Brain, conscious experience and the observing self. Trends in Neurosciences 26: 671–675.

Banissy M. J. & Ward J. (2007) Mirror-touch synesthesia is linked with empathy. Nature Neuroscience 10: 815–816.

Berlyne D. (1960) Conflict, arousal and curiosity. McGraw-Hill, New York.

Braitenberg V. (1984) Vehicles: Experiments in synthetic psychology. MIT Press, Cambridge MA.

Brook A. (2008) Kant's view of the mind and consciousness of self. In: Zalta E. N. (ed.) The Stanford encyclopedia of philosophy. Retrieved from http://plato.stanford.edu on 31 July 2008.

Brooks R. A. (1990) Elephants don't play chess. Robotics and Autonomous Systems 6: 3–15.

Brooks R. A. (1991) Intelligence without representation. Artificial Intelligence 47: 139–159.

Buckner R. L. & Carroll D. C. (2007) Self-projection and the brain. Trends in Cognitive Sciences 11: 49–57.

Butz M. V. (2008) Sensomotorische Raumrepräsentationen. Informatik-Spektrum 31: 237–240.

Butz M. V. & Hoffmann J. (2002) Anticipations control behavior: Animal behavior in an anticipatory learning classifier system. Adaptive Behavior 10: 75–96.

Butz M. V. & Pezzulo G. (2008) Benefits of anticipations in cognitive agents. In: Pezzulo G., Butz M. V., Castelfranchi C. & Falcone R. (eds.) The challenge of anticipation: A unifying framework for the analysis anddesign of artificial cognitive systems. Springer, Berlin: 45–62.

Butz M. V., Herbort O. & Hoffmann J. (2007) Exploiting redundancy for flexible behavior: Unsupervised learning in a modular sensorimotor control architecture. Psychological Review 114: 1015–1046.

Butz M. V., Reif K. & Herbort O. (2008) Bridging the gap: Learning sensorimotor-linked population codes for planning and motor control. In Dillmann R., Sandini G. & Maloney C. (eds.) Proceedings of the 2008 international conference on cognitive systems. University of Karlsruhe, Karlsruhe: 1–7.

Butz M. V., Sigaud O. & Gérard P. (2003a) Anticipatory behavior: Exploiting knowledge about the future to improve current behavior. In: Butz M. V., Sigaud O. & Gérard P. (eds.) Anticipatory behavior in adaptive learning systems: Foundations, theories, and systems. Springer, Berlin: 1–10.

Butz M. V., Sigaud O. & Gérard P. (eds.) (2003b) Anticipatory behavior in adaptive learning systems: Foundations, theories, and systems (LNAI 2684). Springer, Berlin.

Ceccato S. & Zonta B. (1980) Linguaggio, consapevolezza, pensiero. Feltrinelli, Milano.

Chomsky N. (1995) The minimalist program. MIT Press, Cambridge MA.

Cole J. (1995) Pride and a daily marathon. MIT Press, Cambridge MA.

Condillac É. B. de (1780) Logique. Reprinted in: Thonnard F.-J. (1963) Extraits des grands philosophes. Desclée & Cie, Paris: 536–537.

Deacon T. W. (1997) The symbolic species: The co-evolution of language and the brain. W. W. Norton & Company, New York.

Deneve S. & Pouget A. (2004) Bayesian multisensory integration and cross-modal spatial links. Journal of Physiology, Paris, 98: 249–258.

Desmurget M. & Grafton S. (2000) Forward modeling allows feedback control for fast reaching movements. Trends in Cognitive Sciences 4: 423–431.

Drescher G. L. (1991) Made-up minds: A constructivist approach to artificial intelligence. MIT Press, Cambridge MA.

Edelman G. (1987) Neural Darwinism: The theory of neuronal group selection. Basic Books, New York.

Evans R., Weir D., Carroll J., Paiva D. & Belz A. (2007) Modelling control in generation. In: Proceedings of the 11th European workshop on natural language generation, 17–20 June 2007, Schloss Dagstuhl, Germany: 25–32.

Feynman R. P. (1988) QED: The strange theory of light and matter. Princeton University Press, Princeton.

Fiehler K., Burke M., Bien S., Röder B. & Rösler F. (2008) The human dorsal action control system develops in the absence of vision. Cerebral Cortex (in press).

Flanagan J. R. & Beltzner M. A. (2000) Independence of perceptual and sensorimotor predictions in the size–weight illusion. Nature Neuroscience 3: 737–741.

Fragopanagos N., Kockelkoren. S. & Taylor J. G. (2005) A neurodynamic model of the attentional blink. Cognitive Brain Research 24(3): 568–586.

Fries P. (2005) A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. Trends in Cognitive Sciences 9: 474–480.

Fries P., Nikolic D. & Singer W. (2007) The gamma cycle. Trends in Neurosciences 30: 309–316.

Frith C. D., Blakemore S.-J., & Wolpert D. M. (2000) Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. Brain Research Reviews 31: 357–363.

Gallagher S. (2000) Philosophical conceptions of the self: Implications for cognitive

science. Trends in Cognitive Science 4(1): 14–21.
Gallese V. (2001) The "shared manifold" hypothesis: From mirror neurons to empathy. Journal of Consciousness Studies 8(5–7): 33–50.
Gallese V. (2003) The manifold nature of interpersonal relations: The quest for a common mechanism. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 358: 517–528.
Gallese V. & Goldman A. (1998) Mirror neurons and the simulation theory of mind-reading. Trends in Cognitive Sciences 2: 493–501.
Gentner R. & Classen J. (2006) Modular organization of finger movements by the human central nervous system. Neuron 52: 731–42.
Gibson J. J. (1979) The ecological approach to visual perception. Lawrence Erlbaum, Mahwah NJ.
Giese M. A. & Poggio T. (2003) Neural mechanisms for the recogniton of biological movements. Nature Reviews Neuroscience 4: 179–192.
Glasersfeld E. von (1995) Radical constructivism. A way of knowing and learning. Falmer Press, London.
Glasersfeld E. von (2003) Foreword. In: Butz M. V., Sigaud O. & Gérard P. (eds.) Anticipatory behavior in adaptive learning systems: Foundations, theories, and systems. Springer, Berlin: v–vi.
Glasersfeld E. von (2005) Thirty years radical constructivism. Constructivist Foundations 1(1): 9–12.
Glasersfeld E. von (2008) Who conceives of society? Constructivist Foundations 3: 59–64.
Graziano M. & Gross C. (1998) Spatial maps for the control of movement. Current Opinion in Neurobiology 8(2): 195–201.
Graziano M. S. A. (2006) The organization of behavioral repertoire in motor cortex. Annual Review of Neuroscience 29: 105–134.
Graziano M. S. A. & Cooke D. F. (2006) Parieto-frontal interactions, personal space, and defensive behavior. Neuropsychologia 44: 845–859.
Grill-Spector K. & Malach R. (2004) The human visual cortex. Annual Review of Neuroscience 27: 649–677.
Grush R. (2004) The emulation theory of representation: Motor control, imagery, and perception. Behavioral and Brain Sciences 27: 377–396.
Guerra-Filho G. & Aloimonos Y. (2006) Understanding visuo-motor primitives for motion synthesis and analysis. Computer Animation and Virtual Worlds 17: 207–217.
Guerra-Filho G. & Aloimonos Y. (2007) A language for human action. IEEE Computer Magazine 40: 60–69.
Harnad S. (1990) The symbol grounding problem. Physica D 42: 335–346.
Hartley M. & Taylor J. G. (2008) Towards a neural model of mental simulation. In: Kurkova V., Neruda R. & Koutnik I. (eds.) Proceedings of the international conference on artificial neural networks (ICANN2008) Part II. Springer, Berlin: 968–979.
Hartley M., Fagard J., Esseily R. & Taylor J. G. (2008) Observational versus trial and error effects in a model of an infant learning paradigm. In: Kurkova V., Neruda R. & Koutnik I. (eds.) Proceedings of the international conference on artificial neural networks (ICANN2008) Part II. Springer, Berlin: 277–289.
Hauser M. D., Chomsky N. & Fitch W. T. (2002) The faculty of language: What is it, who has it, and how did it evolve? Science 298: 1569–1579.
Heidegger M. (1962) Being and time. Harper and Row, New York. German original published in 1927.
Herbart J. F. (1825) Psychologie als Wissenschaft neu gegründet auf Erfahrung, Metaphysik und Mathematik. Zweiter, analytischer Teil. August Wilhem Unzer, Königsberg.
Herbort O., Butz M. V. & Hoffmann J. (2005) Towards an adaptive hierarchical anticipatory behavioral control system. In: Castelfranchi C., Balkenius C., Butz M. V. & Ortony A. (eds.) From reactive to anticipatory cognitive embodied systems: Papers from the AAAI fall symposium. AAAI Press, Menlo Park CA: 83–90.
Hesslow G. (2002) Conscious thought as simulation of behaviour and perception. Trends in Cognitive Sciences 6: 242–247.
Hoffmann H. & Möller R. (2003) Unsupervised learning of a kinematic arm model. In: Kaynak O., Alpaydin E., Oja E. & Xu L. (eds.) Artificial neural networks and neural information processing, ICANN/ICONIP 2003, LNCS 2714. Springer, Berlin: 463–470.
Hoffmann J. (1993) Vorhersage und Erkenntnis: Die Funktion von Antizipationen in der menschlichen Verhaltenssteuerung und Wahrnehmung. [Anticipation and cognition: The function of anticipations in human behavioral control and perception.]. Hogrefe, Göttingen.
Hoffmann J., Berner M., Butz M. V., Herbort O., Kiesel A., Kunde W., Lenhard R. & Hoffmann J. (2007) Explorations of anticipatory behavioral control (ABC): A report from the cognitive psychology unit of the University of Würzburg. Cognitive Processing 8: 133–142.
Hofsten C. von (2003) On the development of perception and action. In: Valsiner J. & Connolly K. J. (eds.) Handbook of developmental psychology. Sage, London: 114–140.
Hofsten C. von (2004) An action perspective on motor development. Trends in Cognitive Science 8: 266–272.
Holmes N. P. & Spence C. (2004) The body schema and multisensory representation(s) of peripersonal space. Cognitive Processing 5: 94–105.
Holst E. von & Mittelstaedt H. (1950) Das Reafferenzprinzip. Naturwissenschaften 37: 464–476.
Hommel B., Müsseler J., Aschersleben G. & Prinz W. (2001) The theory of event coding (TEC): A framework for perception and action planning. Behavioral and Brain Sciences 24: 849–878.
Hopf J. M., Luck S. J., Girelli M., Hagner T., Mangun G. R., Scheich H. & Heinze H. J. (2000) Neural sources of focused attention in visual search. Cerebral Cortex 10(12): 1233–1241.
Hull C. L. (1943) Principles of behaviour Appleton-Century-Crofts, New York.
Hume D. (1748) Philosophical essays concerning human understanding. Millar, London.
Humphrey N. (1992) A history of the mind: Evolution and the birth of consciousness. Copernicus, New York.
Humphrey N. (2006) Seeing red: A study in consciousness. Harvard University Press, Cambridge MA.
Husserl E. (1964) The phenomenology of

internal time consciousness. (Translated by James S. Churchill). Indiana University Press, Bloomington IN. German original published in 1928.
Ingvar D. H. (1979) "Hyperfrontal" distribution of the cerebral grey matter flow in resting wakefulness: on the functional anatomy of the conscious state. Acta Neurologica Scandinavica 60: 12–25.
Iriki A. (2006) The neural origins and implications of imitation, mirror neurons and tool use. Current Opinion in Neurobiology 16: 660–667.
James W. (1950) The principles of psychology. Dover, New York. Originally published in 1890.
Jiang W., Wallace M. T., Jiang H., Vaughan J. W. & Stein B. E. (2001) Two cortical areas mediate multisensory integration in superior colliculus neurons. The Journal of Neurophysiology 85(2): 506–522.
Jordan J. S. & Hunsinger M. (2008) Learned patterns of action–effect anticipation contribute to the spatial displacement of continuously moving stimuli. Journal of Experimental Psychology: Association Human Perception and Performance 34: 113–124.
Kant I. (1974) Kritik der reinen Vernunft. Edited by W. Weischedel. Suhrkamp, Frankfurt am M. Originally published in 1781.
Kant I. (2003) The project Gutenberg etext of the critique of pure reason. Retrieved from http://www.gutenberg.org/etext/4280 on 12 September 2008.
Karlsson M. M. (1996) Do we think with our brains? In: Mémoires, inscriptions, actions, individuelles et collectives. Séminaire Interdisciplinaire de Sciences Cognitives et Epistémologie, Université de Technologie de Compiègne, France, 22–26 January 1996. pp. 33–45.
Keysers C. & Perrett D. I. (2004) Demystifying social cognition: A Hebbian perspective. Trends in Cognitive Sciences 8(11): 501–507.
Knoblich G., Thornton I. M., Grosjean M. & Shiffrar M. (eds.) (2005). Human body perception from the inside out. Oxford University Press, New York.
Konczak J. & Dichgans J. (1997) The development toward stereotypic arm kinematics during reaching in the first 3 years of life. Experimental Brain Research 117: 346–354.
Körding K. P., Ku S.-P. & Wolpert D. M. (2004) Bayesian integration in force estimation. Journal of Neurophysiology 92: 3161–3165.
Korsten N. J. H., Fragopanagos N., Hartley M., Taylor N. & Taylor J. G. (2006) Attention as a controller. Neural Networks 19: 1408–1421.
Kuhn J. (2000) Processing optimality-theoretic syntax by interleaved chart parsing and generation. In: Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL-2000), Hongkong: 360–367.
Kuhn S. (2008) Prisoner's dilemma. In: Zalta E. N. (ed.) The Stanford encyclopedia of philosophy. Retrieved from http://plato.stanford.edu/on 31 July 2008.
Kunde W., Koch I. & Hoffmann J. (2004) Anticipated action effects affect the selection, initiation, and execution of actions. The Quarterly Journal of Experimental Psychology. Section A: Human Experimental Psychology 57: 87–106.
Làdavas E., Zeloni G. & Farnè A. (1998) Visual peripersonal space centred on the face in humans. Brain 121: 2317–2326.
Langer J., Rivera S., Schlesinger M. & Wakeley A. (2003) Early congitive development: Ontogeny and phylogeny. In: Valsiner J. & Connolly K. J. (eds.) Handbook of developmental psychology. Sage, London: 141–171.
Lee T. S. & Mumford D. (2003) Hierarchical Bayesian inference in the visual cortex. Journal of the Optical Society of America 20(7): 1434–1448.
Legrand D. (2007a) Pre-reflective self-as-subject from experiential and empirical perspectives. Consciousness and Cognition 16: 583–599.
Legrand D. (2007b) Subjectivity and the body: Introducing basic forms of self-consciousness. Consciousness and Cognition 16: 577–582.
Legrand D., Brozzoli C., Rossetti Y. & Farnè A. (2007) Close to me: Multisensory space representations for action and pre-reflexive consciousness of oneself-in-the-world. Consciousness and Cognition 16: 687–699.
Levin J. (1999) Qualia. In: Wilson R. A. & Keil F. C. (eds.) The MIT encyclopedia of the cognitive sciences. Cambridge MA: MIT Press: 693–694.
Maravita A., Spence C. & Driver J. (2003) Multisensory integration and the body schema: Close to hand and within reach. Current Biology 13: 531–539.
Maturana H. & Varela F. J. (1987) The tree of knowledge. Shambhala, Boston.
Mehta B. & Schaal S. (2002) Forward models in visuomotor control. Journal of Neurophysiology 88: 942–953.
Mel B. W. (1991) A connectionist model may shed light on neural mechanisms for visually guided reaching. Journal of Cognitive Neuroscience 3: 273–292.
Möller R. & Schenck W. (2008) Bootstrapping cognition from behavior. A computerized thought experiment. Cognitive Science 32: 504–542.
Neumann G. (1998) Interleaving natural language parsing and generation through uniform processing. Artificial Intelligence 99(1): 121–163.
Neumann G. (2004) Domänenadaptive Kerntechnologien für die Analyse und Synthese von natürlicher Sprache. Habilitation in Computational Linguistics, Saarland University. Retrieved from http://www.dfki.de/~neumann/publications/new-ps/habil-GN.pdf on 20 October 2008.
Newell A. (1990) Unified theories of cognition: The William James lectures. Harvard University Press, Cambridge MA.
Nishide S., Ogata T., Tani J., Komatani K. & Okuno H. (2008) Predicting object dynamics from visual images through active sensing experiences. Advanced Robotics 22(5): 527–546.
Nitschke J. B., Sarinopoulos I., Mackiewicz K. L., Schaefer H.S. & Davidson R. J. (2006) Functional neuroanatomy of aversion and its anticipation. Neuroimage 29: 106–116.
O'Regan J. K. & Noë A. (2001) A sensorimotor account of vision and visual consciousness. Behavioral and Brain Sciences 24: 939–1031.
Pezzulo G. (2008) Coordinating with the future: The anticipatory nature of representation. Minds and Machines 18(2): 179–225.
Pezzulo G. & Castelfranchi C. (2007) The symbol detachment problem. Cognitive Processing 8(2): 115–131.
Pfeifer R. & Bongard J. C. (2006) How the body shapes the way we think: A new view

of intelligence. MIT Press, Cambridge MA.
Piaget J. (1975) Der Aufbau der Wirklichkeit beim Kinde. Ernst Klett: Stuttgart. Originally published in French as: Piaget J. (1937) La construction du réel chez l'enfant. Délachaux & Niestlé, Neuchâtel.
Piaget J. (1991) Das Erwachen der Intelligenz beim Kinde. Ernst Klett Verlag. Originally published in French as: Piaget J. (1935) La naissance de l'intelligence chez l'enfant. Délachaux & Niestlé, Neuchâtel.
Pinker S. (1994) The language instinct: How the mind creates language. W. Morrow, New York.
Poggio T. & Bizzi E. (2004) Generalization in vision and motor control. Nature 431: 768–774.
Porr B. & Wörgötter F. (2005) Inside embodiment – what means embodiment to radical constructivists? Kybernetes 34: 105–117.
Rao R. P. N. (2005) Bayesian inference and attentional modulation in the visual cortex. Neuroreport 16: 1843–1848.
Redgrave P. & Gurney K. (2006) The short-latency dopamine signal: A role in discovering novel actions? Nature Reviews Neuroscience 7: 967–975.
Ridley M. (1996) The origins of virtue: Human instincts and the evolution of cooperation. Penguin Books, London.
Riegler A. (2005) Editorial. The constructivist challenge. Constructivist Foundations 1(1): 1–8.
Riegler A. (2007) Superstition in the machine. In: Butz M. V., Sigaud O. & Pezzulo G. & Baldassarre G. (eds.) Anticipatory behavior in adaptive learning systems: From brains to individual and social behavior. Springer, Berlin: 57-72.
Riesenhuber M. & Poggio T. (1999) Hierarchical models of object recognition in cortex. Nature Neuroscience 2(11): 1019–1025.
Riesenhuber M. & Poggio T. (2000) Models of object recognition. Nature Neuroscience Supplement 3: 1199–1204.
Ristad E. S. (1993) The language complexity game. MIT Press, Cambridge, MA.
Rizzolatti G. & Craighero L. (2004) The mirror-neuron system. Annual Review of Neuroscience 27: 169–192.
Rizzolatti G. & Fadiga L. (1998) Grasping objects and grasping action meanings: The dual role of monkey rostroventral premotor cortex (area F5). Novartis Foundation Symposium 218: 81–103.
Rizzolatti G., Fadiga L., Fogassi L. & Gallese V. (1997) Enhanced: The space around us. Science 277: 190–191.
Rizzolatti G., Fadiga L., Gallese V. & Fogassi L. (1996) Premotor cortex and the recognition of motor actions. Cognitive Brain Research 3: 131–141.
Rizzolatti G., Riggio L., Dascola I. & Umiltá C. (1987) Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. Neuropsychologia 25(1A): 31–40.
Rosen R. (1985) Anticipatory systems. Pergamon Press, Oxford UK.
Rosen R. (1991) Life itself. Columbia University Press, New York.
Rosenblueth A., Wiener N. & Bigelow J. (1943) Behavior, purpose and teleology. Philosophy of Science 10: 18–24.
Rushworth M. F. S., Ellison A. & Walsh V. (2001) Complementary localization and lateralization of orienting and motor attention. Nature Neuroscience 4(6): 656–661.
Rushworth M. F. S., Nixon P. D., Renowden S., Wade D. T. & Passingham R. E. (1997) The left parietal cortex and motor attention. Neuropsychologia 35(9): 1261–1273.
Salmon W. C. (1998) Causality and explanation. Oxford University Press, New York.
Schacter D. L., Addis D. R. & Buckner R. L. (2007) Remembering the past to imagine the future: The prospective brain. Nature Reviews Neuroscience 8: 657–661.
Schenck W. & Möller R. (2007) Training and application of a visual forward model for a robot camera head. In: Butz M. V., Sigaud O. & Pezzulo G. & Baldassarre G. (eds.) Anticipatory behavior in adaptive learning systems: From brains to individual and social behavior. Springer, Berlin: 153–169.
Schubotz R. I. (2004) Human premotor cortex: Beyond motor performance. MPI Series in Human Cognitive and Brain Sciences (Volume 50). Max Planck Institute for Human Cognitive and Brain Sciences: Leipzig.
Schubotz R. I. (2007) Prediction of external events with our motor system: Towards a new framework. Trends in Cognitive Sciences 11(5): 211–218.
Schubotz R. I. & Cramon D. Y. von (2003) Functional-anatomical concepts on human premotor cortex: Evidence from fMRI and PET studies. NeuroImage 20 (Supplement 1): S120–S131.
Schuster A. (1904) An introduction to the theory of optics. Edward Arnold, London.
Schwartz A. B., Moran D. W. & Reina G. A. (2004) Differential representation of perception and action in the frontal cortex. Science 303: 380–383.
Sebanz N., Bekkering H. & Knoblich G. (2006) Joint action: Bodies and minds moving together. Trends in Cognitive Sciences 10: 70–76.
Seligman M. E. P. (1975) Helplessness: On depression, development, and death. W. H. Freeman, San Francisco.
Seligman M. E. P. & Maier S.F. (1967) Failure to escape traumatic shock. Journal of Experimental Psychology 74: 1–9.
Seward J. P. (1949) An experimental analysis of latent learning. Journal of Experimental Psychology 39: 177–186.
Shanahan M. P. & Baars B. J. (2005) Applying global workspace theory to the frame problem. Cognition 98: 157–176.
Shieber S. M. (1993) The problem of logical-form equivalence. Computational Linguistics 19(1): 179–190.
Shoemaker S. (1968) Self-reference and self-awareness. The Journal of Philosophy 65: 555–567.
Simons D. J. & Chabris C. F. (1999) Gorillas in our midst: Sustained inattentional blindness for dynamic events. Perception 28: 1059–1074.
Singer W. (1999) Neuronal synchrony: A versatile code for the definition of relations? Neuron 24: 49–65.
Skinner B. (1971) Beyond freedom and dignity. Bantam/Vintage, New York.
Smitsman A. W. & Bongers R. M. (2003) Tool use and tool making: A developmental action perspective. In: Valsiner J. & Connolly K. J. (eds.) Handbook of developmental psychology. Sage, London: 172–193.
Sommer M. A. & Wurtz R. H. (2006) Influence of the thalamus on spatial visual processing in frontal cortex. Nature 444: 374–377.
Spence S. A. (2002) Alien motor phenomena: A window on to agency. Cognitive Neuropsychiatry 7: 211–220.

Stock A. & Stock C. (2004) A short history of ideo-motor action. Psychological Research 68: 176–188.
Swarup S. & Gasser L. (2007) The role of anticipation in the emergence of language. In: Butz M. V., Sigaud O. & Pezzulo G. & Baldassarre G. (eds.) Anticipatory behavior in adaptive learning systems: From brains to individual and social behavior. Springer, Berlin: 35–56.
Symes E., Tucker M., Ellis R., Vainio L. & Ottoboni G. (2008) Grasp preparation improves change detection for congruent objects. Journal of Experimental Psychology: Human Perception and Performance 34: 854–871.
Tani J. (1998) An interpretation of the "self" from the dynamical systems perspective: A constructivist approach. Journal of Consciousness Studies 5(5–6): 516–542.
Tani J. (2004) The dynamical systems accounts for phenomenology of immanent time: an interpretation by revisiting a robotics synthetic study. Journal of Consciousness Studies 11(9): 5–24.
Tani J. (2007) On the interactions between top-down anticipation and bottom-up regression. Frontiers in Neurorobotics 1(2): 1–10.
Tani J. & Nolfi S. (1998) Learning to perceive the world as articulated: An approach for hierarchical learning in sensory-motor systems. In: Pfeifer R., Blumberg B., Meyer J.-A. & Wilson S. (eds.) From animals to animats 5. MIT Press, Cambridge MA. Reprinted in Neural Networks 12: 1131–1141.
Tani J., Ito M. & Sugita Y. (2004) Self-organization of distributedly represented multiple behavior schemata in a mirror system: Reviews of robot experiments using RNNPB. Neural Networks 17: 1273–1289.
Taylor J. G. (2002) Paying attention to consciousness. Trends in Cognitive Sciences 6: 206–210.
Taylor J. G. (2006) The mind: A user's manual. Wiley, Chichester UK.
Taylor J. G. (2007) On the neurodynamics of the creation of consciousness. Cognitive Neurodynamics 1(2): 97–118.
Tolman E. C. (1932) Purposive behavior in animals and men. Appleton, New York.
Tolman E. C. (1949) There is more than one kind of learning. Psychological Review 56: 144–155.
Toussaint M. & Goerick C. (2007) Probabilistic inference for structured planning in robotics. International Conference on Intelligent Robots and Systems (IROS 2007): 1–6.
Tulving E. (1972) Episodic and semantic memory. In: Tulving E. & Donaldson W. (eds.) Organization of memory. Academic Press, New York: 381–403.
Umiltà M. A., Kohler E., Gallese V., Fogassi L., Fadiga L., Keysers C. & Rizzolatti G. (2001) I know what you are doing: A neurophysiological study. Neuron 31: 155–165.
Varela F., Thompson E. & Rosch E. (1991) The embodied mind. MIT Press, Boston.
Vaziri S., Diedrichsen J. & Shadmehr R. (2006) Why does the brain predict sensory consequences of oculomotor commands? Optimal integration of the predicted and the actual sensory feedback. The Journal of Neuroscience 26: 4188–4197.
Velde F. van der & Kamps M. de. (2006) Neural blackboard architectures of combinatorial structures in cognition. Behavioral and Brain Sciences 29: 37–108.
Ward L. M. (2003) Synchronous neural oscillations and cognitive processes. Trends in Cognitive Sciences 7: 553–559.
Watson J. B. (1913) Psychology as a behaviorist views it. Psychological Review 20: 158–177.
Webster's ninth new collegiate dictionary (1984) Merriam-Webster, Springfield.
Williams J. H. G., Whiten A., Suddendorf T. & Perrett D. I. (2001) Imitation, mirror neurons and autism. Neuroscience and Biobehavioral Reviews 25: 287–295.
Wilson T. D. & Gilbert D. T. (2005) Affective forecasting: Knowing what to want. Current Directions in Psychological Science 14(3): 131–134.
Witt J., Linkenauger S., Bakdash J. & Proffitt D. (2007) Golf performance makes the hole look as big as a bucket or as small as a dime [Abstract]. Journal of Vision 7(9): 291.
Witt J. K. & Proffitt D. R. (2005) See the ball, hit the ball. Apparent ball size is correlated with batting average. Psychological Science 16: 937–938.
Witt J.K. & Proffitt D. R. (in press) Action-specific influences on distance perception: A role for motor simulation. Journal of Experimental Psychology: Human Perception and Performance.
Witt J. K., Proffitt D. R. & Epstein W. (2005) Tool use affects perceived distance, but only when you intend to use it. Journal of Experimental Psychology: Human Perception and Performance 31: 880–888.
Wohlschläger A., Haggard P., Gesierich B. & Prinz W. (2003) The perceived onset time of self- and other-generated actions. Psychological Science 14: 586–591.
Wolff P. (1985) Wahrnehmungslernen durch Blickbewegungen. In: Neumann O. (ed.) Perspektiven der Kognitionspsychologie. Springer, Berlin: 63–111.
Wolpert D. M. & Ghahramani Z. (2004) Computational motor control. Science 269: 1880–1882.
Wolpert D. M. & Kawato M. (1998) Multiple paired forward and inverse models for motor control. Neural Networks 1: 1317–1329.
Womelsdorf T., Schoffelen J.-M., Oostenveld R., Singer W., Desimone R., Engel A. K. & Fries P. (2007) Modulation of neuronal interactions through neuronal synchronization. Science 316: 1609–1612.
Wörgötter F. & Porr B. (2005) Temporal sequence learning, prediction and control. A review of different models and their relation to biological mechanisms. Neural Computation 17: 245–319.
Ximena D. Y. & Maturana H. R. (2008) Habitar humano en seis ensayos de biología-cultural. Juan Carlos Saes Editores, Santiago.
Zahavi D. (2006) Subjectivity and selfhood: Investigating the First Person Perspective. MIT Press, Cambridge MA.