Which Structures Are Out There?

Learning Predictive Compositional Concepts Based on Social Sensorimotor Explorations

Martin V. Butz

How do we learn to think about our world in a flexible, compositional manner? What is the actual content of a particular thought? How do we become language ready? I argue that free energy-based inference processes, which determine the learning of predictive encodings, need to incorporate additional structural learning biases that reflect those structures of our world that are behaviorally relevant for us. In particular, I argue that the inference processes and thus the resulting predictive encodings should enable (i) the distinction of space from entities, with their perceptually and behaviorally relevant properties, (ii) the flexible, temporary activation of relative spatial relations between different entities, (iii) the dynamic adaptation of the involved, distinct encodings while executing, observing, or imagining particular interactions, and (iv) the development of a — probably motor-grounded — concept of forces, which predictively encodes the results of relative spatial and property manipulations dynamically over time. Furthermore, seeing that entity interactions typically have a beginning and an end, free energy-based inference should be additionally biased towards the segmentation of continuous sensorimotor interactions and sensory experiences into events and event boundaries. Thereby, events may be characterized by particular sets of active predictive encodings. Event boundaries, on the other hand, identify those situational aspects that are critical for the commencement or the termination of a particular event, such as the establishment of object contact and contact release. I argue that the development of predictive event encodings naturally lead to the development of conceptual encodings and the possibility of composing these encodings in a highly flexible, semantic manner. Behavior is generated by means of active inference. The addition of internal motivations in the form of homeostatic variables focusses our behavior — including attention and thought — on those environmental interactions that are motivationally-relevant, thus continuously striving for internal homeostasis in a goal-directed manner. As a consequence, behavior focusses cognitive development towards (believed) bodily and cognitively (including socially) relevant aspects. The capacity to integrate tools and other humans into our minds, as well as the motivation to flexibly interact with them, seem to open up the possibility of assigning roles — such as actors, instruments, and recipients — when observing, executing, or imagining particular environmental interactions. Moreover, in conjunction with predictive event encodings, this tool- and socially-oriented mental flexibilization fosters perspective taking, reasoning, and other forms of mentalizing. Finally, I discuss how these structures and mechanisms are exactly those that seem necessary to make our minds language ready.

Keywords

Anticipatory behavior | Compositional concepts | Cooperation | Embodiment | Event segmentation theory | Free energy principle | Homeostasis | Language | Predictive encodings | Sensorimotor learning | Social interactions

1 Structuring the Generative, Predictive Mind

The predictive mind (Hohwy 2013), which may be viewed as continuously "surfing" on its currently active predictions and the involved uncertainties about its environment (Clark 2013; Clark 2016), gives a very intuitive and integrative view on how our mind works. However, many details of this perspective remain to be determined. The free energy-based inference principle offers a mathematical framework to specify implementational details (Friston 2010), addressing the question how predictions may interact and how predictive structures may be learned in the first place. Furthermore, goal-pursuance

has been successfully integrated by formulations of active inference, which is anticipatory in that it takes probabilistic differences between expected and desired future states into account when inferring current behavior, thus yielding goal-directed and epistemic (that is, information seeking) behavior. In sum, the free energy principle allows the mathematical formulation of slower structural learning and faster activity adaptations (Friston et al. 2011) as well as anticipatory, active inference-based goal-directed behavior.

While all three inference aspects have been implemented successfully, the implementations so far have not come anywhere close to yielding a scalable learning system, that is, a system that is able to successfully and computationally efficiently learn in and interact with complex, real-world environments. Moreover, the learning of conceptual structures and behaviorally-relevant abstractions from continuous sensory-motor information has not yet been accomplished. Nonetheless, the available proofs of principle show that the free energy-based inference approach and the resulting conceptualization of the mind as a predictive encoding and processing system has very strong merits and seems neuro-computationally as well as cognitively plausible (Butz 2016; Clark 2016; Hohwy 2013).

One reason why scalability is still out of reach may lie in the fact that current formalizations and implementations rely on particular, hand-designed representations, within which formalizations of uncertainties, probabilistic information processing, predictive estimations, motor activities, and sensory feedback unfold. From machine learning and optimization theory perspectives, however, it is well-known that learning can only make efficient progress when particular structures can be expected in the addressed learning or optimization problems (cf. no free lunch theorem (NFL), Wolpert and Macready 1997). Thus, the theory implies that it is mandatory to uncover the structures that can be found in our environment and which a learning and optimization system should `look for.' Technically, this means that formalizations of free energy-based learning and inference should work on integrating those structural learning and information processing biases that are maximally suitable to learn from and interact with our world most effectively. Presumably, evolution has integrated those biases into our brain's free energy-based inference processes (including, for example, physiological growth and neural wiring mechanisms).

Fortunately, we do not need to start from scratch when exploring which structures are there and we do not need to be very speculative, either. Rather, psychological and cognitive science research offers various clues about fundamental structural components, which our brain tends to process in distinct manners. One very important aspect is the fact that predictive encodings must ultimately serve the purpose of flexibly and effectively planning and controlling interactions with the world. Thus, the mentioned structural learning biases need to be behavior-oriented. For example, behavioral predictions and goal-directed manipulations of entities can be encoded much more effectively by entity-relative spatial arrangements and local interactions between entities in contrast to global spatial localizations.

Thus, I have suggested that three fundamental types of predictive encodings[1] should be distinguished, which are *spatial*, *top-down*, and *temporal predictive encodings* (Butz 2016). The separation of these will lead to the development of (i) universal spatial mappings — and probably the possibility to think spatially in the first place — of (ii) higher-level, multisensory integrative perceptual encodings of entities and their particular properties, and of (iii) temporal predictive encodings, which enable the anticipation of future events on various time scales. Moreover, temporal predictive encodings enable goal-directed behavioral decision making and control as well as goal-oriented attention by means of active inference.

For abstracting the predictive encodings further, event segmentation theory (Zacks and Tversky 2001) offers an additional fundamental structural principle: the segmentation of the continuous sensory-motor experiences into *event encodings* and *event boundary encodings*. It appears that when learning focusses on the processing, detection, and induction of events, fundamental conceptualiza-

1 Please note that I use the term "predictive encodings" to explicitly distinguish such encodings from "representations." Predictive encodings are not representations per se. Rather, they are encodings of predictions about the activity state of other predictive encodings. Their partial representational properties are only a result of what is actually encoded.

tions of the environment can develop, which come in the form of spatial, property-oriented, and temporal force-based conceptualizations (Butz 2016).

Clearly, our body-grounded motivational system is the driving force that makes us interact with and explore our world in the first place. Hunger and other bodily signals, as well as social needs, determine our behavior from birth onwards, and to some extent even before that. Thus, encodings need to develop that are able to predict when and how certain motivations are typically satisfied. As a result, the predictive encodings sketched-out above can be expected to be further shaped by motivational influences. Moreover, behavior will be determined to a large extent by the bodily motivational system, such that the gathered sensory-motor statistics about the world will be strongly motivationally biased.

Finally, besides tendencies towards particular modularizations and segmentations, the human mind has developed highly versatile behavioral and social capacities. *Tool usage* is unprecedented and relies on the ability to flexibly integrate different tools into our own postural body schema. *Social interactions* require the integration of other humans into our cognitive apparatus — with the tendency to assign similar capabilities (physical and mental) to them. I thus emphasize the importance of our social abilities, and particularly cooperation and perspective taking in relation to predictive encodings. By interacting with others and acknowledging that others and even the society as an imaginary entity watch and evaluate us — and even determine further interactions with us dependent on these observations — our minds integrate us into a bigger social reality, within which any participant can take on particular roles during particular interactions (Tomasello 2014). I will discuss what this implies for our actual perceptions, interactions, and actual thoughts about our world, our 'selves', and our knowledge and cognitive capabilities.

In conclusion, I argue that while our minds are individually shaped in their details, the overall structure reflects the structures found in our environment, including physical, biological, as well as social and cultural structures. As a result, our abilities for role taking and flexible concept compositions come naturally to us and significantly contribute to the beauty and the peculiarities of our mind. No wonder that a universal grammar has been detected (Jackendoff 2002): it is the grammar of pre-linguistic human thought, which enables us to learn any available human language as a child.

2 How Do We Comprehend Compositional Concepts? An Illustrative Example

Let us look at an example of how our mind seems to combine entities and associated knowledge about these entities into a consistent concept composition. Interestingly, artificial intelligence has been struggling to do just that and has recently proposed the so-called *Winograd Challenge* (Levesque et al. 2012; Levesque 2014), which is named after the AI researcher Terry Winograd , who created a rather intelligent software in the 1970s (Winograd 1972), which was able to produce meaningful sentences and interactions in a blocks world. The derived challenge basically focusses on common sense reasoning, wrapped into the challenge of pronoun disambiguation. Take for example the following statement:

The ball fits into the suitcase, because it is small [large].

Clearly the pronoun 'it' refers to the ball or to the suitcase, depending on whether the adjective is 'small' or 'large'. Grammatically ball or suitcase could be chosen (and in natural English language it is actually more likely that the adjective (small or large) refers to the subject (ball). The Winograd Challenge explicitly gives a 50% chance of choosing the referenced noun correctly when no semantic information is considered. Our common sense knowledge helps us solve this pronoun disambiguation problem. In particular, as put forward by Barsalou and others (Barsalou 1999; Barsalou 2008; Butz 2008; Butz 2016; Gallese and Goldman 1998), we probably imagine the situation in some form of conceptualized, anticipatory simulation. Figure 1 illustrates some aspects of the simulation that may be activated in our minds. There is the entity 'ball' and the entity 'suitcase', which are probably encoded by means of distributed, conceptual, predictive encodings. Moreover, the verb phrase "fits into" implies

that the subject can be placed into the object, which furthermore implies that the object is a kind of container. We thus have the concept composition of a ball that can be placed into the suitcase — abstracted over spatial entities; the composition may specify that one entity is smaller than the other entity, such that it can be placed into the hollow area of the other entity.

The sentence continuation then makes a statement about why the described situation is true, as indicated by the word "because", and further specifies the spatial relationship between the two, confirming the situation of the first part. For "fits into" to be applicable, the first entity needs to be smaller than the second entity, thus, in order to maintain a consistent overall simulation, "it" must refer to the ball [suitcase] depending on the adjective.

Note that we are processing the information online while reading. This can be nicely illustrated when considering the altered sentence:

The suitcase fits into the ball.

When considering the situation described in this sentence, something feels wrong. From the description above it is not so hard to identify that the concept "fits into" is facing inconsistent objects: a ball is not a common container — especially not for objects! Moreover, it is unlikely that one has ever experienced a suitcase being placed inside a ball. However, the former aspect is probably the one that makes the sentence feel incorrect, because it calls upon our common sense knowledge and essentially makes us think "how can a suitcase be put into a ball!?"

Let us consider one more sentence modification:

The ball fits into the suitcase, because it is light [heavy].

What happens in our mind in this situation? Clearly, the adjectives "light" and "heavy" are referring to a weight concept, which is not tapped into in the first part of the sentence. Lightness and heaviness do usually not affect the concept of 'fitting into' something. However, before fully dismissing this sentence as semantically incorrect, we may develop explanations, such as weight restrictions, which may then allow us to use 'fitting into' in a more metaphorical manner. In this case, the metaphor transfers the 'container' and 'fitting into' concepts from the spatial volume realm into the weight scale realm, assuming there is a particular weight restriction (e.g. for traveling on an airplane). Interestingly, a common magnitude representation has been proposed, which may help to understand this metaphor (Walsh 2003).

Throughout the rest of the paper I will refer back to this example to highlight the importance of the involved predictive encoding structures and the activation mechanisms, which determine the currently most active predictive encodings.

3 Fundamental Types of Predictive Encodings

When considering predictive encodings, it seems worthwhile to contrast particular types of encoded predictions. Before introducing these types, however, I want to clarify how I use the terms *predictive encodings* and *currently active predictive encodings*.

By predictive encodings I loosely refer to a set of neurons with their neural connections, which encode particular (predictive) relationships. Usually, a particular predictive encoding will be implemented by a set of neurons in the brain. For simplicity, however, it suffices to think of an encoding as a neuron with its axon and its dendritic tree, which essentially encode how information is transferred from the pre-synaptically connected neurons via the dendritic tree, axon hillock, and axon to the post-synaptically connected neurons, without considering further details on how this works exactly (not to mention the important involvements and dynamics of the neural transmitters, oxygen supply, etc.). The encodings are predictive in nature because they structure themselves for the purpose of predicting other neural activities.

Currently active predictive encodings are those encodings that are currently actively firing in that they are determining the currently unfolding cognitive processing dynamics. The simplest correspondence in the brain may be a neuron that is producing an action potential. However, the active encodings addressed in this paper are probably realized in the brain by a suitable (probabilistic) combination of well-timed firing neurons. For simplicity reasons, I write about *sets of active encodings*, which may seem to imply that an encoding can only be either on or off. However, what I am actually addressing is those encodings that are currently firing sufficiently strongly to influence the unfolding cognitive dynamics — thus, for example, activating the imagining of — or 'thought' about — a particular concept.

From research in psychology and cognitive science three fundamental types of predictive encodings have often been considered separately — albeit they are certainly strongly interactive: *spatial predictive encodings*, *top-down predictive encodings*, and *temporal predictive encodings* (Butz 2016; Goodale and Milner 1992; Holmes and Spence 2004). In the following paragraphs, I detail their distinction and their most typical interactions.

3.1 Spatial Predictive Encodings

Spatial encodings have been distinguished from recognition-oriented encodings since the seminal papers of Mishkin et al. 1983, and later of Goodale and Milner 1992. However, we still do not know how our minds actually generate predictive spatial encodings from sensorimotor experiences. When I refer to spatial predictive encodings, I mean spatial mappings that map different frames of references onto each other, essentially predicting that the information perceived or encoded in one frame of reference is related to the one perceived in the other frame of reference. The most basic forms of such mappings are concerned with our postural body schema (Butz 2014; Holmes and Spence 2004). At birth, and probably even before birth, a baby shows signs that it has some knowledge about its own body (cf. Rochat 2010). Indeed, such spatial mappings are important not only to map different sensory sources onto each other, but also to enable spatially-oriented interactions with the environment. For example, very early and rudimentary spatial mappings appear to enable fetuses to insert their thumb into their mouth even in the womb.

Modeling such capabilities for enabling a continuous information exchange between different sensory information sources (including visual, proprioceptive, and tactile) has shown that spatial mappings reflect the structure of the external three-dimensional space — or six-dimensional when also considering the orientation of an object or a limb, relative to, for example, the body mid-axis (Ehrenfeld and Butz 2013; Ehrenfeld et al. 2013; Schrodt and Butz 2015). The spatial encodings enable the dynamic activation of the currently applicable spatial mappings. Although it has not yet been shown rigorously, it may be the case that spatial mappings that are learned by means of the free energy-based inference principle, especially when enforcing sparse, compact encodings, develop a universal spatial encoding system that enables the mapping of any frame of reference imaginable onto any other, related frame of reference, and which thus reflects the dimensionality of the outside environment.

From the psychological perspective there are various indicators that spatial encodings play a fundamental role in abstract spatial reasoning and spatial problem solving (Kneissler et al. 2014). Moreover, there are various indicators that suggest that objects are encoded in terms of relative spatial constellations — rather than fully visually. Furthermore, the perception of or the thought about unfolding spatial dynamics — such as rotations — appears to be encoded distinctly from the actual sensors and entities, which may actually cause the perception of the unfolding spatial dynamics. The result is an intermodal crosstalk between the involved modalities, including tactile, visual, and motor modalities as well as the mere thought about some dynamics — such as the mental rotation of oneself or an object (Butz et al. 2010a; Janczyk et al. 2012; Liesefeld and Zimmer 2013; Lohmann et al. 2016).

3.2 Top-down Predictive Encodings

Top-down predictive encoding is the most basic type of predictive encoding. It has been investigated in detail in the neuro-vision literature (Chikkerur et al. 2010; Giese and Poggio 2003; Rao and Ballard 1998). From the psychological perspective, however, it seems that deeper differentiations in the top-down encodings develop only after the fundamental spatial encodings are sufficiently well structured. Possibly one of the first characterizations of top-down predictive encodings comes from the Gestalt psychologists (cf. Koffka 2013), investigating to what extent we can deduce and imagine whole figures from particular sensory input. Point-light motion figures are a well-known example, in which we tend to perceive, for example, a walking human person although we only see a few points of light that are attached to particular body parts. Even the size, agility, gender, and emotions of the person can to some extent be deduced solely by observing the motion dynamics (cf. e.g. Pavlova 2012).

Note that top-down predictive encodings typically encode predictions of particular aspects of a stimulus while ignoring others. For example, specialized areas in the visual cortex have been identified that selectively process color, visual motion, or complex edges. Indications of similar somatosensory property encodings in a corresponding ventral stream have been identified as well. Top-down predictive encodings may thus be generally characterized as predictions about typical perceivable properties of some entity — and these property predictions may address any available modality — including bodily and motivational signals — as well as combinations of modalities. In combination with spatial mappings, the currently active top-down encodings may be flexibly projected onto the currently relevant sensory-grounded frames of reference — such as onto the correct position in the retinotopic frame of reference or onto the correct position in a tactile, somatosensory body map.

3.3 Temporal Predictive Encodings

The third types of predictive encodings, which strongly interact with spatial and top-down predictive encodings, are temporal predictive encodings. Essentially temporal predictive encodings predict activity changes in the two other types of fundamental predictive encodings. The first temporal predictive encodings that develop in our brain are probably those concerning our own body — which muscular activities have which body postural and other perceptual (mainly proprioceptive) effects? For example, it has been shown that the visual effect caused by saccades and the opening and closing of one's eyes is anticipated by an information processing loop through the thalamus (Sommer and Wurtz 2006). Underlying this processing principle is the *reafference principle* (von Holst and Mittelstaedt 1950): corollary discharges of motor activities are converted into sensory predictions of the consequent effects. This principle was further spelled-out by the theory of anticipatory behavioral control and related theories from cognitive psychology (Prinz 1990; Hoffmann 1993).

Note that temporal predictive encodings — irrespective of where they apply and what exactly they predict — process information over time and thus predict upcoming changes. The nature of the changes, however, may differ in very fundamental ways, ranging from sensory changes (such as a change in color, in stiffness, in vibration, in loudness etc.) to abstract property changes (such as weight, content type, amount, etc.) and spatial changes (such as displacements and changes in orientation). Temporal predictive encodings can be expected to be active together with any top-down and spatial predictive encodings, effectively encoding the currently imaginable potential property or spatial changes, respectively.

Take the example of the ball with some of its activated, characteristic spatial, temporal, and top-down encodings shown in Figure 1a. Top-down predictive encodings will, for example, generate visual predictions of roundness, and possibly more concrete images, such as those of a soccer ball. Pre-activated temporal predictions may anticipate the consequences of the ball interacting with other objects as well as typical motion dynamics (e.g. starting to roll, bounce, fly, react to a kick in a certain way, etc.). Finally, without additional information, active spatial predictive encodings may predict a typical standard ball size and a somewhat central location in front of us.

Imagine now the situation where we watch a ball lying outside — say on a sidewalk — and it suddenly begins to role in one direction seemingly without any cause. What would be the constructed explanation? Probably we come up with the explanation that it must be a windy day and the wind must have forced the ball to start moving (possibly supported by an additional suitable slope of the sidewalk). The activation of the rolling temporal predictive encoding requires the presence or the activation of a force. Initial experiences of such forces are generated by our own motor behaviors (experiencing bodily restrictions in the womb, for example). Over time, top-down predictive encodings can be learned that generalize more behaviors to forces, which may be generated by our motor behavior but also by other means. These encodings in turn predict the activation of temporal predictive encodings of the expected changes (including spatial and property changes) that are typically caused by the encoded forces. For example, a "pushing force" encoding can develop that predicts, on the one hand, motor behavior activities that can generate such a force (a top-down prediction) and, on the other hand, changes in motion of the entity that is being pushed by the force (a temporal prediction).

4 Event-Oriented Conceptual Abstractions

While the proposed fundamental types of encodings may be able to encode all kinds of predictions, and top-down predictions typically generate abstractions, several additional psychological theories suggest that in order to build conceptual schemata about the environment, the continuous sensorimotor information flow needs to be segmented systematically.

The theory of anticipatory behavioral control (ABC, Hoffmann 1993; Hoffmann 2003), the common coding approach (Prinz 1990), and the theory of event coding (TEC, Hommel et al. 2001) all imply that actions are encoded in close relation to the action effects they tend to produce. ABC further postulates that such commonly encoded action-effect complexes are endowed with the critical conditions that enable the action-effect complex to take place. For example, an object needs to be in reach in order to be graspable, or an object must not be too heavy such that it is still movable. A computation model of the ABC theory has indeed shown high learning capacity and the potential to model typical adaptive behavior, such as latent learning in a maze, which is observable, for example, in rats (Butz 2002; Butz and Hoffmann 2002). On the other hand, the common coding approach, and its further formalization into TEC, emphasizes that action-effect complexes are co-encoded in a common, abstract code, which coordinates, anticipates, and controls the action-effect complex.

All three theories essentially emphasize that our brains seem to develop encodings of distinct motor activity-effect complexes, which can be selectively activated dependent on the current context. As detailed above, temporal predictive encodings often result in predictive codes of motor activities and their effects. Moreover, motor activities may be substituted by force-effect encodings, which may indeed be the type of encoding envisioned by TEC. Thus, an event can be understood as the application of a particular force in a particular situation.

A more general definition of an event originates from studies on event segmentation. The event segmentation theory (EST), which was derived from these studies, characterizes an event as "a segment of time at a given location that is conceived by an observer to have a beginning and an end". (Zacks and Tversky 2001, p. 3). Later, EST was refined further such that event segments were related to unfolding predictions and it was suggested that "[…] when transient errors in predictions arise, an event boundary is perceived". (Zacks et al. 2007, p. 273).

When considering motor activities, a simple event may thus be understood as a simple, unfolding motor activity, such as a grasp. EST has not been closely related to theories that focus on motor actions, such as ABC or TEC. However, when considering all of these theories in the light of predictive encodings, it appears there is a close relationship: an event may be characterized by a particular set of predictive encodings starting with the onset of this set and ending with the offset of this set. The (also predictive) encodings of the situational properties at which point a particular event typically

commences and at which point it may end then characterize the context, the necessary aspects that can bring the event about, and the ones that can stop it.

Reconsidering the ball example, an event encoding may characterize the typical behavior of a rolling ball by temporal predictive encodings of spatial displacements and of accompanying visual motion signals, which indicate the rolling motion. Moreover, motion onset, that is, the prediction of the onset of the particular temporal predictive encoding of spatial displacement, may be predictable by a top-down force encoding, which can be activated given, for example, the impact of any other moving entity — including one's foot but also a strong gust of wind.

Although the exact formulations of how such event-oriented predictive encodings may develop from basic spatial, top-down, and temporal predictive encodings still need to be spelled-out in detail, a recent algorithm that builds such event encodings from signals of *temporary surprise* seems promising (Gumbsch et al. 2016). One critical aspect of the algorithm is the formulation of temporary surprise, which corresponds to a temporal state of large free energy but which is preceded and succeeded by low free energy states in the involved predictive encodings. Interestingly, this approach was additionally motivated by research on hierarchical reinforcement learning and the challenge to automatically form conceptual hierarchies from sensorimotor signals (Butz et al. 2004; Simsek and Barto 2004; Botvinick et al. 2009).

5 Continuously Unfolding Predictive Encoding Activities

When assuming the existence of a complex network of the described predictive encodings, the currently active encodings constitute the currently considered concepts and their composition. While this may sound intuitively plausible, it is still unknown how these concept compositions are selectively activated, maintained, and dynamically adapted over time. One could assume the basic mechanism would be free energy-based inference. But how does the processing unfold concretely?

From modeling flexible behavioral control and the maintenance of a postural body image over time (Butz et al. 2007; Ehrenfeld et al. 2013; Kneissler et al. 2014) and from the close relationship of these mechanisms to free energy-based inference (Kneissler et al. 2015), evidence has been accumulated that suggest that at least a three-stage information processing mechanism may continuously unfold over time. First, the currently active encodings may be considered as the prior predictions about the current environmental state, including the state of one's body. Next, sensor information integration may take place, yielding local posterior activity adaptations, taking the uncertainties of the prior predictions and sensory information content into account.

After sensor information integration, the predictive network strives towards global consistency, adapting its active predictive encodings further for the purpose of increasing consistency between the active encodings themselves, that is, to minimize internal free energy. As a result, the active encodings move toward a distributed attractor, which comes in the form of a free energy minimum. Note that it seems impossible to reach a global minimum in such a distributed system, in which only local interactions unfold. Note furthermore that this process naturally takes all knowledge about the environment, which is represented in the learned predictive encodings, into account. As a result, local minima, that is, local attractors, will be approximated, which take the predictive activities only from connected encodings into account. Suitably modularized partitions of information contents, such as the spatial and top-down encodings, may optimize this distributed free energy minimization process.

In relation to the concept of a Markov blanket , which is used by Friston and others to derive the theory that the brain develops a predictive model of the (indirectly perceived) outside environment (Friston 2010), consistency enforcement may be thought of as a process during which distributed, internal Markov blankets are at play. That is, while striving for local consistencies in the currently active network of predictive encodings, local adaptations are made, which take the predictive activities of the connected predecessors, successors, and the predecessors of the successors into account. As a

result, the adaptations yield approximate locally consistent (within the respective Markov blanket) but distributed state estimates, which approximate the theoretical, global free energy minimum.

Finally, following the reafference principle, temporal predictions will be processed, generating the next prior predictions. When this anticipation of next sensory information is expanded to all active temporal predictive encodings — including those that do not depend on one's own motor behavior but that depend on the estimated presence of other forces in the environment and the presence of current motion — then essentially the temporal dynamics of the environment are predicted. Given that the global posterior reflects the actual environmental model rather well and given further that the temporal predictive encodings are relatively accurate, hardly any surprise will be encountered when processing the next sensory information and the system will be so-to-say 'in-sync' with the world. Incorrect temporal, spatial, or top-down predictions, on the other hand, will yield larger surprises, that is, larger free energy when processing the incoming sensory information.

The implications of the sketched-out process are diverse. Let us reread the first part of the sentence about the ball and the suitcase above. What will actually happen according to the sketched-out process when considering a word-by-word processing level? The article "the" prepares the predictive encodings for a concrete concept, generating predictions about the next word. Spatial predictive priors will be uniform, or rather will estimate an uncertain central default location. Top-down predictive priors remain unspecified (for example, a uniform distribution in the respective available encodings). Next, the noun "ball" triggers the activation of a more or less concrete conceptual encoding of some sort of ball with its typical properties. Some of the different types of active predictive encodings are sketched-out in Figure 1a. Note also how the article will be merged with the ball, leading to the assumption that the sentence refers to one concrete ball. Thus, on the one hand, one concrete entity code needs to be activated, while, on the other hand, this entity may be associated with all imaginable top-down, predictive ball entity properties. Because these more concrete encodings are very diverse, the activated focus will remain on the abstract but particular ball concept. It remains to be shown how a concrete, but unspecified entity is imagined by our brain. Adhering to our distinct types of predictive encodings, spatial predictive encodings should restrict the imagined entity to one location, while all imaginable ball properties are associated with this activated location via the activated top-down predictive encodings characterizing the term "ball".

Next, "fits into" will imply that the ball can be put or is inside another entity. Thus, another relative spatial predictive encoding must be activated, which relates a yet undefined entity and the ball, such that the dimensions of the ball fit into the other entity's container part. In fact, the activation of the entity will most likely activate predictive encodings that characterize a *containment property*, as sketched-out in the temporal interaction consequence encodings in Figure 1b for the suitcase. Moreover, active temporal predictive encodings will generate the next prior, which essentially leads to the expectation that the container entity will be specified.

Finally, "the suitcase" indeed makes the expected container concrete, causing the activation of the suitcase concept as a particular example of a container. Thus, the container concept is integrated into the suitcase concept. Moreover, the other predictive encodings are verified and adjusted. For example, the predictive encodings that characterize the size of the ball will be adjusted such that the size of the ball is smaller than the size of the suitcase, essentially increasing the consistency of the imagined whole concept composition.

6 Motivations, Goals, and Intentions

I have focused on the information processing and neural activity adaption mechanisms in the above paragraphs. When considering action decision making, however, active inference mechanisms become necessary. Free energy-based formulations of active inference take imaginable future states into account. The tendency to achieve particular futures comes from the principle that internal predictions expect unsurprising outcomes. Combined with homeostatic internal states, the result is a system that strives to sustain internal homeostasis, because very imbalanced drives cause very 'surprising' and thus unfavorable signals.

This principle is closely related to the principle of living systems as autopoietic systems (Maturana and Varela 1980). While the general principle is closely related to reinforcement learning, the formulations avoid the introduction of a separate term that characterizes reward. Rather, reward is integrated into the free energy formulations, such that extrinsic reward is akin to a successful avoidance of overly surprising signals, which can be, for example, unfavorable signals about one's own bodily state. Various artificial system implementations have successfully generated agents that are self-motivated and that are both curious and goal-oriented at the same time. For example, latent learning in rats is fostered by an inherent curiosity drive, while the gathered knowledge can then be used to act in a goal-directed manner (Butz et al. 2010b). The maintenance of a maximally effective balance between the two components, however, remains to be controlled by an appropriate choice of parameter values (Butz et al. 2010b; Friston et al. 2015; Hsiao and Roy 2005).

As a result, from the perspective of the described predictive encoding network with its unfolding, dynamically adapting current activities, it seems necessary that the currently active encodings about the considered situation also need to enable the partial simulation of potential futures forward in time. The proposed hierarchical activity encoding based on event encodings seems to be ideally suited for this matter, because it enables planning and reasoning on abstract predictive encoding levels. For example, when we want to drink out of a glass in front of us, we first need to reach for and grasp the glass and then transport it suitably to our mouth, tilt it properly, and finally drink out of it in a coordinated manner. Thus, the final goal pre-determines the successive subgoals of being in control of the glass and so forth (Gumbsch et al. 2016). Interestingly, cognitive psychological modelling literature suggests that the encodings of spatial relations between entities (such as the hand, the glass, and the mouth) and their potential space-relative interactions seem to be particularly well-suited for cognitive reasoning (Kneissler et al. 2014). It may indeed be the case that spatial predictive encodings, which must have evolved primarily for the control of flexible interactions with objects, tools, and other entities in space, are recruited by our brains to enable other planning and reasoning processes as well. Note that such spatial encodings need to be highly swiftly and flexibly adaptable to the current circumstances, to changes in these circumstances, and for enabling the consideration of such changes caused be own motor activities. As a result, the imagining of — or thoughts about — entity interactions becomes possible in a similarly swift and flexible manner.

Combinations of the present and considered futures then lead us to act — to the best of our knowledge — in our own best interest. We act intentionally to minimize uncertainties about achieving our desired goals. Moreover, we do this on all levels of abstraction that are imaginable by our minds. Priorities certainly vary and depend on the extent to which we prefer particular states and dislike others. Nonetheless, this point of view gives us an inherent and highly individual intentionality, which essentially determines our character.

Let us go back to the ball again. When we want to project the ball into the goal, we may attempt to kick it when we are in range — because we want to produce a flying ball event which may end, if suitably executed, with the ball entering the goal's interior. Similarly, when we go on a trip and pack a suitcase, we activate predictions about the presence of the items we pack into the suitcase wherever we intend to go with the suitcase. If we expect to play with the ball when we have reached the destination, we may consider packing the ball and thus consider if the ball fits into the suitcase. Given it fits, we may actively put the ball into the suitcase by using, for example, our hands. Note again how first the displacement event is activated, which then causes the activation of particular means, such as particular motor actions, that are believed to cause the displacement. Note also that as learning strongly depends on the types of experiences that have been gathered, the developing predictive encodings will strongly depend on the generated active inferences, which, vice versa, depend on the available predictive encodings, that is, the gathered knowledge about the world and about our ability to interact with it and to manipulate it. As a result, we will choose those means to put the ball into the suitcase that we are most used to (for example, by means of our hands or our feet).

7 Perspective Taking, Hypothetical Thinking and Role Taking

As our bodies and our 'selves' become "a public affair" over the first years of our lives, so does our behavior and our social interactions with others (Rochat 2010). In fact, it appears that we tend to constitutionalize our social interactions in such a way that an imaginative entity, that is, the 'public' as a whole is the one that is perceived as (partially) watching us (Tomasello 2014). As we have particular expectations about the general knowledge of others — such as that everybody knows how to open a door, how to count, or the common words of his or her mother tongue — we have particular expectations about social rules of interactions and basic tendencies for interactions, even with nearly complete strangers within a society or a particular culture (Tomasello 2014). These social capacities, which are strengthened and shaped further by our versatile communication capabilities, inevitably enable us to view ourselves not only from the geometric (that is, purely spatial) but also form the epistemic perspective of somebody else. And this 'somebody else' does not even need to be a person but may be an imagined knowledge entity.

A somewhat similar capacity develops when mastering tool use. The tool temporarily becomes part of our body, thus *subjectifying* the tool (Butz 2008; Iriki 2006). From a social cooperative perspective, we can act as a tool — as when handing over the butter at the breakfast table. Similarly, we can perceive our body as a tool, as when we attempt to undertake a task with our hands that is usually accomplished with the help of a particular tool. In both of the latter cases, we essentially *objectify* our body — or a part of it — as a tool.

When manipulating and interacting with objects — possibly with the help of tools — as well as when interacting with (and sometimes also manipulating) others, we experience particular interaction roles and the currently involved goals. When we are the one to initiate and control an interaction, we are the intentional actor, while the manipulated object or person is the recipient. Likewise, we may be the recipient and another person may be the intentional actor. This is probably the reason why we tend to *subjectify* objects when they interact with us by chance in peculiar ways — such as the infamous apple that is said to have fallen onto Sir Isaac Newton's head 'giving' him the idea of universal gravity. Due to the peculiarity of the interaction, we tend to *subjectify* the object and thus assign intentions to it — essentially in the same way we attribute intentions to the behavior that we observe in other humans.

It seems that these social interactions and tool manipulation capabilities are thus the key to enable us to take on different perspectives and thus to flexibly switch roles (actor, recipient, means, space, time) mentally. As a further consequence, we become able to learn the human language we are confronted with. This is probably the case because human languages offer a means to satisfy our drive to communicate with others in order to coordinate social interactions successfully. The grammar of a language essentially constrains how we assign roles to the individual entities addressed in a sentence, including their relationship in space and time and the particular interaction addressed. The prior assumption of the pre-linguistic mind is that verbalized states of affairs in the environment and changes in the environment typically refer to particular entities, relative spatial relations between entities, and particular interactions of entities (cf. also Knott 2012). In all these cases, it is important to clarify exactly which entities are referred to and what role they play. The detection of ambiguities during communication leads to the addition of words and further learning of the grammar the developing child is confronted with.

Let us re-consider the original sentence and analyze the second part of it from a reasoning perspective with the ability to take on different perspectives and thus the ability to construct counterfactuals and the involved hypothetical alternative scenarios.

The ball fits into the suitcase, because it is small.

The word "because" suggests that there is a particular property of one of the entities that makes the first sentence true and if it was changed to another property — typically the opposite — then the sentence would be false. To simulate this, we need to be able to change our image of one or the other object. The property put forward is of type "size" and the term "small" implies being "not large". The

property can be directly associated with the constructed mental situation of one entity fitting into another entity. Revisiting the constructed network of active predictive encodings (Figure 1c) and changing the size of either object allows the construction of two counterfactual situations. Changing the ball to a larger ball may make it not fit, because the combined relative spatial and size encoding would yield a spatial overlap of the entities, implying that the ball does not fit. On the other hand, changing the suitcase to a larger suitcase would not change the truth of the first part, because no overlap would occur and the ball would still fit. Thus, it is much more plausible that "the ball" is referenced by the pronoun.

Let us now change the example to show that the same principle generally can apply in a social context with particular persons or a group of persons constituting particular entities. Consider the following nearly equivalent sentence in a social context:

The Smith family fits into the discussion group, because it is insightful.

Note how the sentence specifies two entities, which are actually a group of people: The "Smith family" and the "discussion group". Again, the first entity is said to fit into the second entity — where "fits into", as above, implies that the first entity can be inserted or added to the second entity, where the second entity functions as a container. In fact, a "discussion group" allows the addition of more people, so it can function as a container. Fitting into the social group in this case, however, is not a matter of 'size' but of 'social competence'. This is exactly what the second part of the sentence implies: it argues that the Smith family fits because of its insightfulness. The opposite social property — something like "dull" or "unwise", that is, "not insightful" — would probably not fit into a discussion group, where insight and fruitful discussions are typically sought. Thus, the sentence can generally be processed in a manner similar to the ball and suitcase sentence, constructing an active predictive encoding network (an attractor that indicates consistency) and disambiguating the 'it' by imagining the two counterfactual options, that is, the 'Smith family' or the 'discussion group' being not insightful. As a result, it is much more likely that the 'Smith family' is referenced by the pronoun, although the other interpretation cannot be fully excluded.

8 Conclusions

While predictive coding intuitively makes a lot of sense to many of us, I have argued that it is necessary to further describe the most likely kinds of predictive encodings that are developing in our brains and how they interact. I have argued that three basic types may constitute the building blocks of our thoughts: top-down, spatial, and temporal predictive encodings. Moreover, I have argued that, starting with basic predictive encodings of the encountered sensorimotor experiences, event-oriented abstractions can lead to the development of event and event boundary encodings, which can lead to conceptual encodings of the relevant properties to bring particular events about.

Moreover, I have argued that the simulation of a particular thought, a particular situation, or a particular event in a particular context is essentially encoded by a network of active predictive encodings. While, for example, reading a sentence, a constructive process unfolds, which attempts to activate those predictive encodings that form a maximally consistent, interactive network. The network reflects what is believed to be the sentence's content, including its implications. By changing parts of the activated predictive encodings, it is possible to alter the imagined situation in meaningful ways and to probe the resulting consistency. As a result, it is possible to argue semantically about certain events, situations, and statements. Our developing tool usage and social capabilities and the involved encodings seem to support the necessary perspective taking abilities — and thus the ability to imagine conceptual environmental interactions in the past, the future, and by others (Buckner and Carroll 2007). Due to the event-oriented predictive encodings, such imaginings are conceptual because the involved predictive encodings focus on those aspects of the environment that are believed to be relevant for particular situations in and interactions with the world. Moreover, mental manipulations of such imaginings allow the probing of

situational changes and their effects, including property, spatial, and temporal changes — thus enabling planning, reasoning, and the pursuance of hypothetical thoughts. Finally, the social perspective as well as our tool use abilities enable us to objectify ourselves and thus to develop explicit self-consciousness.

Clearly, details of the proposed theory need to be further defined in the near future, in order to verify the involved hypotheses and to further differentiate the identified kinds of predictive encodings. The best way to go forward may be to develop actual implementations of artificial cognitive systems in virtual realities, in addition to the necessary further interdisciplinary philosophical, neuro-psychological, and linguistic research.

Figure 1: Illustrations of the active predictive encoding networks for the thought about "the ball" and "the suitcase" as well as the composition with the connecting information "fits into". Note the predictive interactions and the consistencies between the illustrated active encodings. The visualized encodings include top-down predictions about the visual appearances, spatial predictions about where, relative to the observer (and relative to the other object), one object may be present, as well as what size it may be. Temporal predictive encodings predict interaction consequences as well as potential motion dynamics. In the concept compositions these may annihilate each other, indicating the thought of the ball lying stably inside the suitcase with unlikely current motion dynamics.

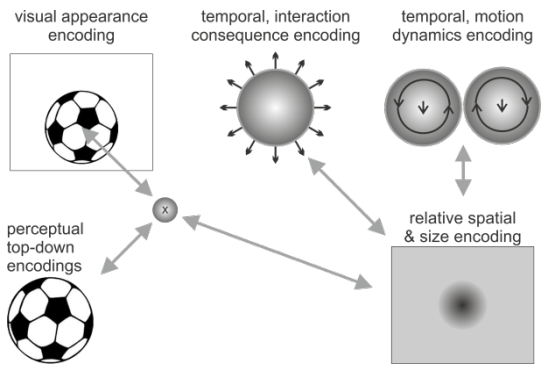

Figure 1 a: Sketch of predictive coding network for "the ball".

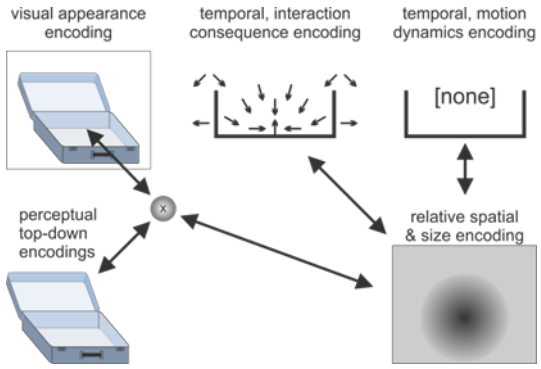

Figure 1 b: Sketch of predictive coding network for "the suitcase".

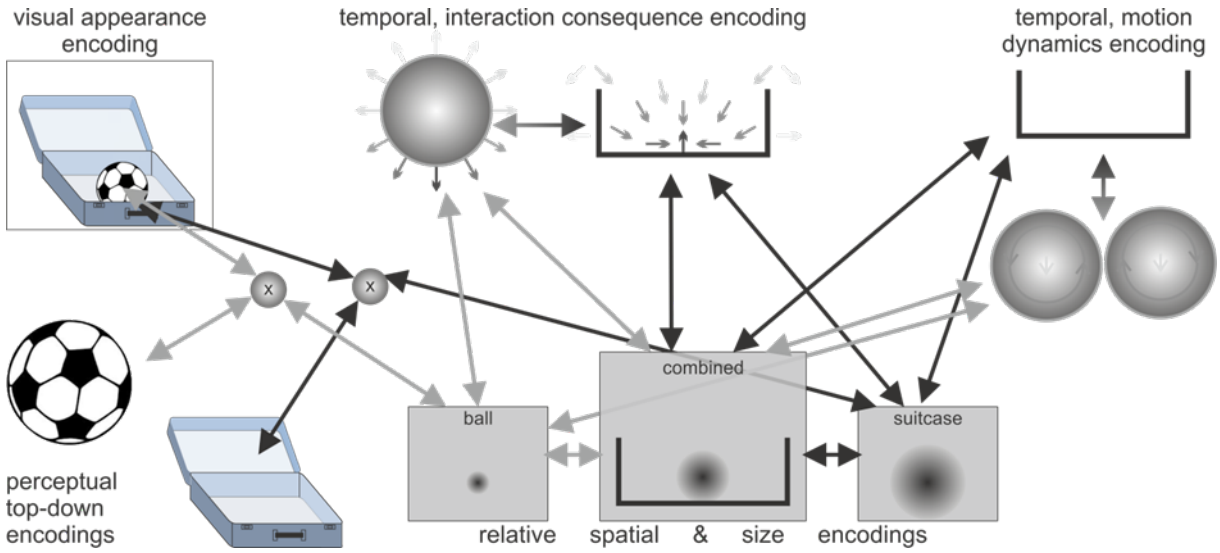

Figure 1 c: Concept composition: "The ball fits into the suitcase."

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–600.

———(2008). Grounded cognition. *Annual Review of Psychology, 59*, 617-645.

Botvinick, M., Niv, Y. & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition, 113* (3), 262 - 280. https://dx.doi.org/10.1016/j.cognition.2008.08.011.

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

Butz, M. V. (2002). *Anticipatory learning classifier systems.* Boston, MA: Kluwer Academic Publishers.

———(2008). How and why the brain lays the foundations for a conscious self. *Constructivist Foundations, 4* (1), 1-42.

———(2014). Rubber hand illusion affects joint angle perception. *PLoS ONE, 9* (3), e92854. Public Library of Science. https://dx.doi.org/10.1371/journal.pone.0092854.

———(2016). Towards a unified sub-symbolic computational theory of cognition. *Frontiers in Psychology, 7* (925). https://dx.doi.org/10.3389/fpsyg.2016.00925.

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., Swarup, S. & Goldberg, D. E. (2004). Effective online detection of task-independent landmarks. In R. S. Sutton & S. Singh (Eds.) *Online proceedings for the ICML'04 workshop on predictive representations of world knowledge* (pp. 10). Online. http://homepage.mac.com/rssutton/ICMLWorkshop.html.

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., Thomaschke, R. Linhardt, M. J. & Herbort, O. (2010a). Remapping motion across modalities: Tactile rotations influence visual motion judgments. *Experimental Brain Research, 207*, 1-11.

Butz, M. V., Shirinov, E. & Reif, K. L. (2010b). Self-organizing sensorimotor maps plus internal motivations yield animal-like behavior. *Adaptive Behavior, 18* (3-4), 315–337.

Chikkerur, S., Serre, T. Tan, C. & Poggio, T. (2010). What and where: A Bayesian inference theory of attention. *Vision Research, 50*, 2233-2247. https://dx.doi.org/10.1016/j.visres.2010.05.013.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Science, 36*, 181-253.

———(2016). *Surfing uncertainty: Prediction, action and the embodied mind.* New York: Oxford University Press.

Ehrenfeld, S. & Butz, M. V. (2013). The modular modality frame model: Continuous body state estimation and plausibility-weighted information fusion. *Biological Cybernetics, 107*, 61-82. https://dx.doi.org/10.1007/s00422-012-0526-2.

Ehrenfeld, S., Herbort, O. & Butz, M. V. (2013). Modular neuron-based body estimation: Maintaining consistency over different limbs, modalities, and frames of reference. *Frontiers in Computational Neuroscience, 7* (148). https://dx.doi.org/10.3389/fncom.2013.00148.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*, 127- 138. https://dx.doi.org/10.1038/nrn2787.

Friston, K., Mattout, J. & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics, 104* (1-2), 137-160. https://dx.doi.org/10.1007/s00422-011-0424-z.

Friston, K., Levin, M. Sengupta, B. & Pezzulo, G. (2015). Knowing one's place: A free-energy approach to pattern regulation. *Journal of The Royal Society Interface, 12* (105). https://dx.doi.org/10.1098/rsif.2014.1383.

Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2* (12), 493–501.

Giese, M. A. & Poggio, T. (2003). Neural mechanisms for the recogniton of biological movements. *Nature Reviews Neuroscience, 4*, 179-192.

Goodale, M. A. & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15* (1), 20-25. https://dx.doi.org/10.1016/0166-2236(92)90344-8.

Gumbsch, C., Kneissler, J. & Butz, M. V. (2016). *Learning behavior-grounded event segmentations* (pp. 1787-1792). Cognitive Science Society.

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.].* Göttingen, GER: Hogrefe.

———(2003). Anticipatory behavioral control. In M. V. Butz, O. Sigaud & P. Gérard (Eds.) *Anticipatory behavior*

in adaptive learning systems: Foundations, theories, and systems (pp. 44-65). Berlin/Heidelberg, Springer-Verlag.

Hohwy, J. (2013). *The predictive mind.* Oxford, UK: Oxford University Press.

Holmes, N. P. & Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cognitive Processing, 5*, 94-105.

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.

Hsiao, K. Y. & Roy, D. (2005). In C. Castelfranchi, C. Balkenius, M. V. Butz & A. Ortony (Eds.) *A habit system for an interactive robot* (pp. 83-90). Menlo Park, CA: AAAI Press.

Iriki, A. (2006). The neural origins and implications of imitation, mirror neurons and tool use. *Current Opinion in Neurobiology, 16*, 660-667.

Jackendoff, R. (2002). *Foundations of language. Brain, meaning, grammar, evolution.* Oxford University Press.

Janczyk, M., Pfister, R. Crognale, M. A. & Kunde, W. (2012). Effective rotations: Action effects determine the interplay of mental and manual rotations. *Journal of Experimental Psychology: General, 141*, 489-501. https://dx.doi.org/10.1037/a0026997.

Knauff, M. (2013). *Space to reason. A spatial theory of human thought.* Cambridge, MA: MIT Press.

Kneissler, J., Stalph, P. O. Drugowitsch, J. & Butz, M. V. (2014). Filtering sensory information with XCSF: Improving learning robustness and robot arm control performance. *Evolutionary Computation, 22*, 139-158. https://dx.doi.org/10.1162/EVCO_a_00108.

Kneissler, J., Drugowitsch, J. Friston, K. & Butz, M. V. (2015). Simultaneous learning and filtering without delusions: A Bayes-optimal combination of predictive inference and adaptive filtering. *Frontiers in Computational Neuroscience, 9* (47). Frontiers Media S.A. https://dx.doi.org/10.3389/fncom.2015.00047.

Knott, A. (2012). *Sensorimotor cognition and natural language syntax.* Cambridge, MA: MIT Press.

Koffka, K. (2013). *Principles of Gestalt psychology.* Abingdon, UK: Routledge.

Levesque, H. J. (2014). On our best behaviour. *Artificial Intelligence, 212*, 27–35.

Levesque, H., Davis, E. & Morgenstern, L. (2012). The Winograd schema challenge. *Knowledge Representation and Reasoning Conference. Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning.* http://www.aaai.org/ocs/index.php/KR/KR12/paper/view/4492.

Liesefeld, H. R. & Zimmer, H. D. (2013). Think spatial: The representation in mental rotation is nonvisual. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39* (1), 167–182.

Lohmann, J., Rolke, B. & Butz, M. V. (2016). In touch with mental rotation: Interactions between mental and tactile rotations and motor responses. *Experimental Brain Research*.

Maturana, H. & Varela, F. (1980). *Autopoiesis and cognition: The realization of the living.* Boston, MA: Reidel.

Mishkin, M., Ungerleider, L. G. & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6*, 414-417. https://dx.doi.org/10.1016/0166-2236(83)90190-X.

Pavlova, M. A. (2012). Biological motion processing as a hallmark of social cognition. *Cerebral Cortex, 22* (5), 981–995. https://dx.doi.org/10.1093/cercor/bhr156.

Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Eds.) *Relationships between perception and action* (pp. 167-201). Berlin Heidelberg: Springer-Verlag.

Rao, R. P. & Ballard, D. H. (1998). Development of localized oriented receptive fields by learning a translation-invariant code for natural images. *Computational Neural Syststems, 9*, 219-234.

Rochat, P. (2010). The innate sense of the body develops to become a public affair by 2-3 years. *Neuropsychologia, 48* (3), 738 - 745. https://dx.doi.org/10.1016/j.neuropsychologia.2009.11.021.

Schrodt, F. & Butz, M. V. (2015). *Learning conditional mappings between population-coded modalities* (pp. 141-148).

Simsek, Ö. & Barto, A. G. (2004). Using relative novelty to identify useful temporal abstractions in reinforcement learning. *Proceedings of the Twenty-First International Conference on Machine Learning (ICML-2004)*, 751-758.

Sommer, M. A. & Wurtz, R. H. (2006). Influence of the thalamus on spatial visual processing in frontal cortex. *Nature, 444*, 374-377.

Tomasello, M. (2014). *A natural history of human thinking.* Cambridge, MA: Harvard University Press.

von Holst, E. & Mittelstaedt, H. (1950). Das Reafferenzprinzip (Wechselwirkungen zwischen Zentralnervensystem und Peripherie.). *Naturwissenschaften, 37*, 464-476.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive*

Sciences, 7 (11), 483–488. https://dx.doi.org/10.1016/j.tics.2003.09.002.

Winograd, T. (1972). Understanding natural language. *Cognitive Psychology, 3* (1), 1–191.

Wolpert, D. H. & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation, 1* (1), 67-82.

Zacks, J. M. & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127* (1), 3–21. https://dx.doi.org/10.1037/0033-2909.127.1.3.

Zacks, J. M., Speer, N. K. Swallow, K. M. Braver, T. S. & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133* (2), 273–293. https://dx.doi.org/10.1037/0033-2909.133.2.273.