Martin V. Butz and Daniel Zöllner

Towards Grounding Compositional Concept Structures in Self-organizing Neural Encodings

Martin V. Butz martin.butz@uni-tuebingen.de
University of Tübingen, Faculty of Science,
Department of Computer Science,
Cognitive Modeling

Daniel Zöllner daniel.zoellner@gmx.net
University of Tübingen, Faculty of Science,
Department of Computer Science,
Cognitive Modeling

Abstract

While the symbol grounding problem of agreeing on a mapping between symbols and sensory or even sensorimotor grounded concepts has been solved to a large extent, one possibly even deeper open problem remains: How do concepts and compositional concept structures develop in the first place? Concepts may be described as integrative mental representations that encode certain sensory, motor, or sensorimotor states or events. Compositionality, on the other hand, determines how concepts are associated with each other in a semantically meaningful and highly flexible manner. We argue that progressively complex concepts and compositional structures can be developed starting from very basic perceptual and motor control mechanisms. An experiment with a simple simulated robot gives hints about highly relevant structural ontogenetic prerequisites for their development. In the outlook, we conclude by sketching out the current most pressing challenges ahead.

Keywords: concepts, compositionality, development, symbol grounding, language, neural networks, manifolds, anticipation

1 Introduction

Symbols are “placeholders” standing for other entities. In a dictionary, and often in conversation, symbols are explained through other symbols. This is a potentially endless process called “semiosis” by the philosopher Charles Sanders Peirce: Symbols are described by symbols, which are described by symbols – and so on. But how can this endless process be ultimately grounded, how “is symbol meaning to be grounded in something other than just more meaningless symbols?” (Harnad 1990, p. 340). This is what Harnad (1990) calls the “symbol grounding problem”.

While Steels (2008) states that the basic symbol grounding problem has been solved, it was also pointed out that yet a deeper symbol grounding problem needs to be addressed (cf. Barsalou 2009, Harnard 1990, Sugita & Butz 2011). The robotic agents in Steels' works are able to come to an agreement about a symbol convention for particular communication realms (such as gestures, colors, etc.). That is, a common language is developed where particular symbols or utterances are associated with particular perceptions or perception-action complexes. The challenge of the deeper symbol grounding problem lies in the development (a) of compositional concept structures from sensorimotor control capabilities and (b) of associations between those structures and grammatical, symbolic, i. e. linguistic structures. Only when these two challenges are accomplished, formal semantics may be actually grounded in sensorimotor codes.

The study of both the developmental progression that led to the grounding of compositional concepts and the nature of the involved structures and associations is expected to provide insights on how "Cognitive Semantics" (Johnson 1987, Lakoff 1987, Lakoff & Johnson 1980) actually pre-determine formal semantics and most likely even structural properties of the universal grammar (Chomsky 1965). Most recently, the idea of cognitive semantics led to the proposition of a Minimalist Action Grammar (Pastra & Aloimonos 2012), which was directly related to the Minimalist Program by Noam Chomsky (1995). The Minimalist Action Grammar is a generative grammar that enables both proper generation and parsing of sentences about physical interactions. It binds an interaction by its final goal, combining tool complements, which are about the acting force, with object complements, which are about the affected object, context- and goal-dependently.

We are particularly interested in how such a Minimalist Action Grammar may develop starting purely from embodied, sensorimotor interactions – in the hope to contribute to the deeper symbol grounding problem sketched-out above. The aim is to develop a self-motivated system that solely perceives its environment via sensory stimulations and that probes its environment by motor activities, where sensors and motors are coupled by the bodily morphology. Ultimately, such a model may show that many structures present in the Universal Grammar are grounded in sensorimotor interactions with the environment that are realized by an embodied agent. Meanwhile, such a line of research is expected to also shed light on why and how grammatical structures in language are structured in the way they are – hints of which can also be found in the Minimalist Action Grammar.

Various researchers now strongly believe that sensorimotor structures and the selective simulation of particular sub-structures set the stage for the development of compositional concept structures (Barsalou 2008, Grush 2004, Pastra & Aloimonos 2012, Pezzulo 2011). How such structures are developed and how these structures may then be coupled with higher level cognitive, symbolic encodings is still an open question, though. While the claim that the compositionality of language may be grounded in the compositionality inherent in interaction competencies is not new (Johnson 1987, Lakoff 1987), how such grounding may be learned and how compositionality may be represented by means of sub-symbolic structures remains an open question. Arbib (2005) proposed a developmental pathway that leads from interactions, the mirror neuron system, and imitation capabilities over several further stages to linguistic competence. We believe that these stages are important components in the development of concepts and compositional concept structures. However, several other prerequisites appear mandatory.

The aim of this paper is to sketch out a path by means of which complex, compositional concept structures are action-grounded. We propose that in order to explain the human capacity to generalize, to draw inductions, and to develop compositionality, it is not necessary to resort to innate structures. Rather, as increasingly many robotic architectures and even more so simulations with neural networks imply, compositional concept structures can be developed by a brain "from scratch", departing from sensorimotor contingencies. Endorsing the "Cognitive Semantics" of Lakoff and Johnson (1980), we propose to make the next step to confirm this theory by identifying the ontogenetic ingredients that appear necessary to develop such semantics. Thus, we are interested in the architectural constraints and learning biases necessary for developing compositionality based on sensorimotor interactions.

In this way, the paper also takes a stand in the nature/nurture-debate about concepts. In particular we propose that structures, which rationalists tend to regard as purely innate, are actually derivatives of sensorimotor experiences and developmental constraints. Thus, we propose a nature-constraint "nurture" process, in which genetically determined bodily and brain developmental constraints stream cognitive development towards the acquisition of compositional concept structures and language readiness. However, only with the additionally necessary environmental interactions including linguistic communication can the language capacity develop. Consequently, concepts are grounded in the experienced interactions, but genetic predispositions bias the cognitive developmental process towards concept acquisitions.

We argue that purely innate structures leave no flexibility and are generally extremely questionable due to the immense depth of the necessary structures and due to the fact that even innateness needs to be somehow couple such structures to perceptions and actions. Thus, a core claim of this paper is that the Symbol Grounding Problem (Harnad 1990) can only be solved by an empiricist approach to concept acquisition. In contrast to Fodor's (1975, 2008) radical claim that concepts cannot be learned, we suggest that a theory of concept learning is essential for a complete theory of cognition and the mind.

In the following, we first detail a neural network architecture with which it has recently been shown that representational separations and multiplicative interactions between modules are essential ingredients for the development of compositional concept structures. We detail the type of compositional structures that were developed and how thus compositionality was grounded in embodied sensorimotor interactions. We discuss the implications of this study, but also its limitations and current most pressing challenges. Finally, we put the insights gained into the broader perspective on how concepts and compositionality may develop.

2 An Experiment with a Simulated Robot Platform

In a neural network simulation setup, it was shown that a second-order neural network with parametric bias neurons (sNNPB) is able to develop generalized behavioral control routines, presenting the system solely with typical sensory-motor time series data (Sugita, Tani, & Butz 2011). This study essentially offers tentative answers to the question: How can compositional concept structures self-organize based on experienced sensorimotor interactions? Additional ingredients will be necessary to scale this approach to more complex environments and interaction capabilities.

In the experiment, a simulated robot interacted with colored objects. The robot was equipped with two wheels for controlling motion and a camera that scanned the surrounding in front of the robot. In particular, the camera reported the perceived dominant hue and color intensity values covering an area of 120° in front of the robot. The covered areas were partitioned into nine equally spaced sectors. The robot learned two types of interactions: *move-to* and *orient-towards* a particularly *colored object*. In the move-to interaction, the robot had to move to the object and stop in front of it. In the orient-towards interaction, the robot had to simply orient itself towards an object at a specific angular offset; five offsets were trained. One or two colored objects were

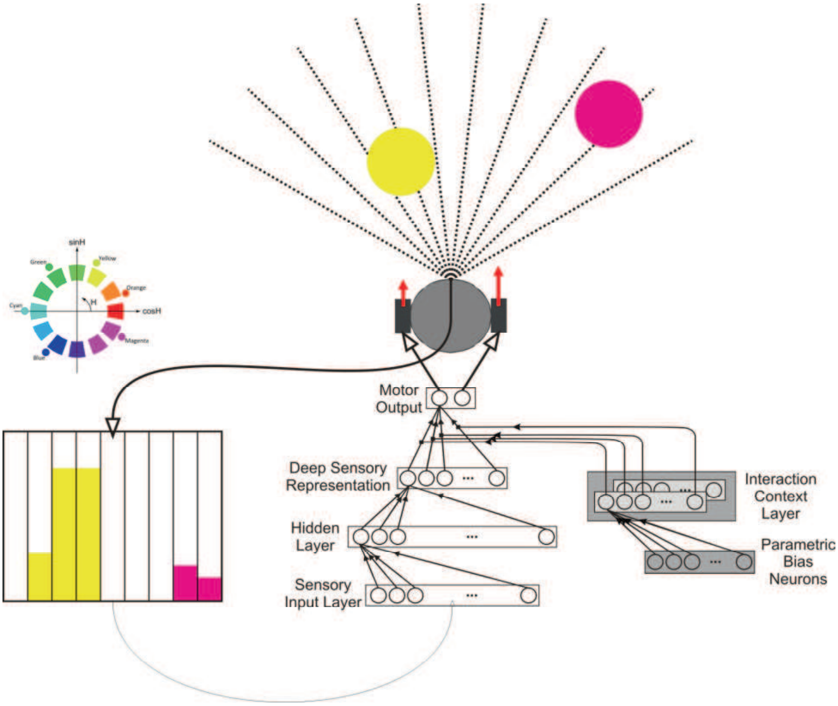

Figure 1: Robot-Environment-sNNPB interaction

present during each interaction trial with the environment. During learning, the actions of the robot were controlled remotely by a hard-coded control program. Figure 1 illustrates the robot, environment, sNNPB interaction.

In the following, we will refer to the two types of interactions as the “verbs” that were trained, to the different colored objects as the “objects” that were addressed in the interactions, and to the offsets in the orient-towards interactions as the involved “modifiers”. Note however that the learning system was not provided with any explicit indicators – neither about the “verbs” nor about the “objects” or the “modifiers” – that may have given clues or induced learning biases towards distinguishing “verb”, “object”, and “modifier” concepts. The only information given to the learning system was the sensorimotor time series data the robot was trained on and the information that particular sets of sensorimotor time series data belonged to the same type of interaction.

The resulting sensorimotor time series data was used to train an sNNPB. An sNN is a traditional neural network, which is trained with backpropagation, which, however, includes some "second-order" neural connections. Second order neural connections essentially are connections whose current weight values are determined by other neural activities. In the conducted simulations, one sub-NN mapped the visual information provided by the camera onto motor output transferring the information over two hidden layers. The connection weights of the connections from the second hidden layer to the motor output, however, were determined by second-order connections. The associated neurons were activated by a second sub-NN with one hidden layer. Input to this network was generated by "parametric bias neurons" (Tani 2003). Error backpropagation was used to adjust the weights of the sNNPB as well as the activities of the parametric bias neurons. The latter were adjusted interaction-specific, thus maintaining a vector for each type of verb-object-modifier interaction the system was trained on.

After learning, the sNNPB was tested on other object constellations and on other, untrained verb-object-modifier interactions. For example, the sNNPB may have never been trained on "move-to the blue object". Nonetheless, after learning the system was tested if it can generate such interactions. To do so, the activity of the parametric bias neurons was set to activity values that matched a small set of generated interactions best. After that, other constellations were tested applying these PB activities.

The results confirmed that the sNNPB generalized over the provided sensorimotor time series data. It was not only able to generate similar interactions in other environmental constellations, but also to generate interactions that were only compositionally related to those trained on. For example, it was able to orient itself towards a particular colored object at a particular angle, while it only had been trained to move to such a colored object. Thus, behaviorally the network exhibited generalization capabilities that were of a compositional nature. Interactions that corresponded to verb-object-modifier constellations could be generated that were not trained – as long as a sufficiently large and distributed subset of other interactions was trained.

Moreover, analyses of the developed sNNPB showed that a self-organized geometrically-arranged manifold structure had developed, which reflected the behaviorally exhibited compositionality. In particular, the activity vectors of the parametric bias neurons were considered for further analysis. A principal component analysis showed that the first principal component differentiated the interactions with respect to the modifier. The second principal component differentiated move-to from orient-towards. The third and fourth principal component revealed a color ring encoding, akin to the one found in the hue-based color encoding provided to the sensory input layer. Thus, activities in the parametric bias neurons self-organized via backpropagation learning into a compositional manifold structure, where the individual dimensions in the manifold corresponded to the verb, object, and modifier components of the individual interactions. The manifold structure enables the sNNPB to flexibly activate any meaningful verb-object-modifier interaction type and also allows generalizing to untrained interaction types. The geometric, orthogonal arrangement was akin to a compositional concept structure because the orthogonality enables flexible interaction concept combinations and the deducible geometric distances can be viewed as indicating concept similarities.

Interestingly, also the structure of the second hidden layer – the one that maps to motor output via the second-order neural connections – was analyzed. Strongly behavior-oriented sensory encodings were found. For example, one neuron switched its behavior from off to on when an object is in the center and very close – resulting in breaking behavior when the move-to interaction is activated in the parametric bias neurons. Other neural activities revealed activities that may be compared to gain fields in neurons (Salinas & Sejnowski 2001, Graziano 2006): neurons responded, for example, in a sinusoidal fashion with respect to color but that response was linearly modulated by the direction where the color was perceived from. In effect, this encoding allowed the flexible activation of particular color-respective encodings for approaching and orienting the robot towards particular colors, dependent on the activated mapping given particular parametric bias activity. From a broader perspective it can be said that object-relative encodings developed that encoded "object affordances" (according to Gibson 1979), in the sense that the encodings afforded to reach a particular orientation towards a particular object or to stop moving when coming close to an object. Providing yet another interpretation, spatial, object-relative encodings were developed that could be directly mapped towards motor activities, yielding a flexible Braitenberg vehicle (Braitenberg 1984).

The network succeeded in developing these compositional concept structures without the provision of any semantic cues besides the ones that were inherent in the sensorimotor time series data. Seeing that various other neural network architectures could not yield similar generalizations, it was concluded that (a) goal-oriented encodings need to be separated from sensorimotor, control-oriented encodings and (b) a multiplicative approach is best-suited to project the goal-oriented encodings onto the sensorimotor encodings for realizing flexible and compositional goal-oriented behavioral control. In the emergent, interaction-specific, goal-oriented encodings the mentioned compositional concept structures could be found, whereas in the processed sensory encodings behavior-oriented signals could be found. Both were shown to be mutually dependent on each other – the former selecting the actual interaction that should be executed; the latter providing potential interaction options.

Seeing that various other neural network architectures were not able to generate comparable compositional behavioral generalization capabilities – let alone actual identifiable compositional structures as the one characterized above – the results suggests that sensory-to-motor mappings should be separated from interaction selection encodings to enable the development of compositional concept structures. Essentially, the interaction selection corresponds to the goal that is to be achieved, with considerations of the component that bring each particular goal about – such as moving to a particularly colored object. While various researchers have suggested that such separations are behaviorally necessary (Cisek 2007), we believe they have not been sufficiently considered in research on the development and structure of language and cognition.

3 Insights and Open Challenges Deducible from the Robot Experiment

The results of the simulated robot experiment have shown that compositional concept structures could only develop in this setup when the sensory-to-motor mapping was separated from the goal encoding, that is, from the code that determines which sensory-to-motor interaction should actually unfold. Also, the time dynamics had to be different in the two encodings in that one goal activity had to be maintained while one full sensorimotor object interaction unfolded. Moreover, it was necessary that the influence from the goal encoding onto the sensory-to-motor mapping was multiplicative. Finally, the generated sensorimotor time series data had to be separated into distinct sets with respect to particular verb-object-modifier combinations. However, no information about the semantics or symbolic characterizations of these particular combinations had to be provided.

In consequence sensorimotor grounded compositional concept structures and behavior-oriented “Braitenberg encodings” co-developed, that is, encodings which are perfectly suited to be directly mapped onto motor output activities, yielding seemingly goal-directed behavior (Braitenberg 1984). Braitenberg encodings are thus goal-oriented encodings, which can be selectively mapped onto actions for pursuing particular object interactions. Indeed, the compositional concept structures had structural similarities with the emerging Braitenberg encodings, thus enabling the selective activation of particular Braitenberg codes for realizing particular object interactions. Compositionality was achieved by embedding a manifold structure into a higher-dimensional neural representation. The individual dimensions of the lower-dimensional (in the experiment four dimensional) manifold corresponded to the compositional verb-object-modifier structure. The developed "object" concept was encoded on a two-dimensional manifold (actually a circular manifold), mimicking the hue-based color encoding in the simulated sensors. Due to the emerging orthogonal arrangement of the distinct concept structures, the sNNPB was able to flexibly compose any verb-object-modifier interaction, even if it had not been trained. The developed compositional concept structure appeared to be perfectly suited to be associated with a corresponding action grammar.

However, at this point language structures have not been successfully associated with developing compositional structures, yet. Sugita & Tani (2005) managed to associate symbolic structures with similar sensorimotor time series data. However, in this case only a more rudimentary action grammar consisting of three possible verbs and six possible colors was learned. Nonetheless, Sugita and Tani (2005) succeeded in mutually shaping both the symbol-based linguistic encoding and the sensory-to-motor mapping. Thus, associating symbolic, linguistic input with developing, self-organizing, more complex action grammars is still a very hard challenge.

Even when focusing only on the challenge of developing pre-linguistic compositional concept structures – without associating symbolic language components – however, additional learning biases and developmental constraints seem mandatory for scalability reasons. At the moment, the sNNPB architecture is still an extremely flexible learning architecture. For developing more complex compositional structures, it seems necessary that the learning processes are further guided by additional learning biases. However, overly constraint learning may not give enough room for the emergence of compositional concept structures, such as the manifold structure identified in the robot experiment. Thus, complex compositionality is likely to emerge only if a good balance between learning biases on the one hand and self-organization on the other hand is maintained.

Another challenge lies in the fact that sets of sensorimotor time series data had to be explicitly distinguished when training the sNNPB, while the more autonomous separation of different types of interactions is desirable. While similarity thresholds may distinguish the sensorimotor time series data, it is very hard to find the right distance metric that could suitably distinguish different time series in a semantically meaningful way. The self-organized topology in the PB neurons of the sNNPB is likely to be the best candidate, but the development of it relied on the distinctness information in the first place.

We believe that several of the following ingredients will be mandatory to develop learning systems that can autonomously produce emergent compositional concept structures in more complex environments. First, the incorporation of an *anticipatory drive* (Butz 2008) that stresses the capability of predicting the future based on state, context, and motor (force) activities seems necessary. Such an anticipatory drive may guide learning first towards identifying the most obvious sensorimotor contingencies in the sensory and motor information available to the system. Further distinctions starting from basic sensorimotor flow may then lead to the desired progressively more distinct compositional concept structures.

Once sensorimotor contingencies are identified, sensorimotor topologies can be developed within which particular interactions can unfold. In the simulated robot experiment, a topology was implicitly developed in the deep sensory encodings, providing Braitenberg codes. Similar, but further modularized encodings are necessary to enable the even more flexible and selective interaction with the environment using different means, different pathways through the environment, etc.

Furthermore, active, information-seeking, curious behavior, caused by the anticipatory drive, may enable the more direct identification of relevant concept structures, that is, of sensory and motor information necessary for predicting particular consequences reliably. The consequent identification of contextual "concepts" that separate states into concepts that are relevant for particular behaviors – such as free versus occupied, heavy versus light, etc. – will be the result.

Besides these learning biases derived from the anticipatory drive, the challenge of removing the requirement of providing distinct sets of sensorimotor time series data may be accomplished by introducing internal motivations. Such internal motivations may serve as the distinctness indicators – identifying a distinct interaction by its distinct effect on the internal motivational state. Thus, distinct positive and negative reinforcement may serve as a critical additional clue to distinguish interactions further into meaningful concepts.

Finally, it seems somewhat unsatisfactory that the activity in the parametric bias neurons cannot be internally self-activated. To do so, the activity of the parametric bias neurons may be partially activated by sensory input as well – potentially enabling the selective activation of those interaction codes that can actually unfold in the current circumstances. For example, a potential interaction with a red object may only be activated if a red object is present. Furthermore, the mentioned internal motivations may be associated with those parametric bias neuron activities that previously had led to a corresponding change in the internal motivational state. Consequently, the interaction choice may be co-determined by the internal motivations and the goals currently possible in the environment.

4 Conclusions

The robot experiment described above contributes to the solution of the symbol grounding problem, and also illuminates concept learning. One of the most vexing problems regarding this topic is Fodor's problem of concept acquisition. Fodor (1975, 2008) essentially questions that fundamental concepts – those that cannot be further partitioned into smaller conceptual entities – can be learned. And presuming that they cannot be learned, he concludes that they must be innate. The details of Fodor's argument are beyond the scope of this article. It suffices to state that according to most recent philosophical considerations, "it appears that Fodor's problem of concept acquisition remains a puzzle for philosophers and psychologists to solve" (McCaffrey & Machery 2012, p. 275).

We propose to overcome Fodor's "radical concept nativism" (cf. Laurence & Margolis 2002) by a different stance towards "innateness". This very ambiguous term may gain a more specific sense if it is related to embodiment. In short, we propose that the innateness of concepts may not be directly genetically imprinted, but concepts and compositional concept structures may be indirectly pre-determined to develop due to (a) the ontogenetic path laid-out in the genes of the organism, (b) the morphological constraints given by the body of the organism, and (c) the environmental reality with which the organism interacts.

Fundamental concepts may indeed be innate – but actually innate in the sense of being behaviorally embodied and pre-destined to be developed. For example, basic reflexes – such as the grasp reflex in infants – can foster the development of particular concepts – such as a concept for grasping. Separating then successful from unsuccessful grasps, a concept structure that specifies the prerequisites for a successful grasp develops, in contrast to contexts were grasps are unsuccessful. Co-developing with such a representation is a concept of graspable entities. Realizing the effects of successful grasps, will expand and differentiate the grasp concept further into entities that are moveable, light versus heavy, spiky versus smooth, etc. The basic reflex may thus lead to the generation of sensorimotor interactions that can be differentiated on the one hand side by their perceptual differences but, and even more importantly so, by their distinct effects.

Essentially we point-out that the combination of an anticipatory drive with an embodied, sensing and acting agent can foster the development of pre-linguistic, compositional concept structures. The anticipatory drive drives the organism to actively search for and learn about predictable and controllable (sensorimotor) structures in the environment (Butz 2008). Due to this self-controlled, embodied developmental process, the developing concept structures are inherently meaningful because the structures determine predictability, controllability, and their relation to changes in internal motivational states. Thus, the combination of the human body morphology with its ontogenetic development of body and brain fosters the development of "innate" but behaviorally acquired compositional concept structures.

Unitizations and differentiations in the sense of Landy & Goldstone (2005) (cf. also Stöckle-Schobel 2012) are fundamental processes that foster the development of compositional concept structures. We propose that these processes are not purely perceptual or sensorimotor, but are developed for predictability, controllability, and achievability purposes. With this proposition we go one step beyond theorists of "neo-empiricism" like Barsalou (2009), Jesse Prinz (2002), and others. We strongly acknowledge that their accounts on perceptually grounded symbols and concepts are highly important in overcoming unworkable accounts of innateness. However, we would like to further stress that cognition and – more specifically, concept acquisition – is not solely shaped by (and for) perception. Rather, it is most important for being able to interact flexibly goal-directedly with objects and other agents.

Moreover, the robot experiment has shown that spatial, object- and body-relative representations should be separated from goal-oriented representations in order to foster the development of compositional structures. Given this separation, particularly the goal-oriented representations appear well-suited for the development of compositionality. Thus, the separation of dorsal and ventral pathway (Goodale & Milner 1992), which is certainly highly behaviorally relevant and mandatory for realizing flexible behavioral control (Cisek 2007, Milner & Goodale 2008), may have actually set the stage for the development of compositional concept structures, that is, structures that allow the development of language in the first place.

Certainly other processes are still highly important as well. In particular, we believe that the development of mirror capabilities and tool use are two fundamental additional ingredients. The capability of mirror neurons, which was first most likely beneficial for improving mutually beneficial interactions with other individuals, fosters the further development of communication between individuals, by, for example, enabling the development of verbal imitations from gestural imitations (Arbib 2005, Rizzolatti &Arbib 1998). The capability of handling tools led to the development of much more intense interactions between the dorsal and ventral processing streams, thus being able to view tools and objects as part of the subject and, in retrospect, also oneself as a tool (Iriki 2006).

However, we believe that the sketched-out processes will set the stage to be able to ultimately solve the mystery of concept acquisition. By separating goals from spatial topologies and events, flexible goal-directed behavior can be selected and pursued. Current internal goals can be flexibly pursued dependent on the current spatial constraints. Moreover, the availability of potential goals in the environment as well as the context-dependent estimated achievability of such potential goals can yield tremendous behavioral flexibility and effectivity. While the development of such a separation was thus initially most likely purely behavior-driven, it also enabled the development of compositional concept structures. While potential goals and the involved concepts for achieving these goals are detached from the here-and-now, the encodings can be flexibly projected onto the current state in the environment. Meanwhile, state representations must have developed that enable the flexible activation of goals and involved concepts for pursuing particular goals. Object-referenced encodings found in in the parietal cortex (Chafee, Averbeck, & Crowe 2007) support the pro-motor representations found in integrative, multimodal cortical areas. The parietal-frontal interactions with which action goals appear to be transferred into actual movement control support their strong goal- and behavioral relevance (Graziano, Cooke 2008). Arguably, similar correspondences were even proposed to exist between Wernicke's and Broca's areas (Graziano, Cooke 2008). Finally, gain-modulations, which are found nearly ubiquitously in the brain, suggest selective, multiplicative computations in individual neurons (Salinas & Sejnowski 2001), supporting the flexible, goal-oriented selection of maximally suitable sensory-to-motor mappings.

In the minimalist Action Grammar as proposed by Pastra & Aloimonos (2012) goals unify particular actions with objects and further modifiers. Our proposition in this paper gives first hints why goals are crucial both, for the development of grammatical structures and for being able to flexibly combine compositional concept structures to achieve particular goals dependent on their current urgency and achievability. Nonetheless, much future research is necessary to sort the identified puzzle pieces, identify even further pieces, and arrange them in the way the ontogenesis of the brain manages to do so beautifully.

References

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

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1281–1289.

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

Butz, M. V. (2008). How and why the brain lays the foundations for a conscious self. *Constructivist Foundations*, 4, 1–42.

Cangelosi, A., & Riga, T. (2006). An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots. *Cognitive Science*, 30, 673–689.

Chafee, M. V., Averbeck, B. B., & Crowe, D. A. (2007). Representing spatial relationships in posterior parietal cortex: Single neurons code object-referenced position. *Cerebral Cortex*, 17, 2914–2932.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

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

Cisek, P. (2007). Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362, 1585–1599.

Fodor, J. (1975). *The Language of Thought.* New York: Crowell.

Fodor, J. (2008). *LOT 2. The Language of Thought Revisited.* Oxford: Oxford University Press.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15, 20–25.

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.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27, 377–396.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42, 335–346.

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

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination and reason.* Chicago, IL: University of Chicago Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago, IL: University of Chicago Press.

Landy, D., & Goldstone, R. G. (2005). How we learn about things we don’t already understand. *Experimental & Theoretical Artificial Intelligence*, 17, 343–369.

Laurence, S., & Margolis, E. (2002). Radical concept nativism. *Cognition*, 86, 25–55.

McCaffrey, J., & Machery, E. (2012). Philosophical issues about concepts. *WIREs Cognitive Science*, 3, 265–279.

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46, 774–785.

Pastra, K., & Aloimonos, Y. (2012). The minimalist grammar of action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 103–117.

Pezzulo, G., & Calvi, G. (2011). Computational explorations of the perceptual symbol system theory. *New Ideas in Psychology*, 29, 275–297.

Prinz, J. (2002). *Furnishing the mind. Concepts and their perceptual basis.* Cambridge, MA: MIT Press.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience*, 21, 188–194.

Salinas, E., & Sejnowski, T. J. (2001). Gain modulation in the central nervous system: Where behavior, neurophysiology, and computation meet. *The Neuroscientist*, 7, 430–440.

Steels, L. (2008). The symbol grounding problem has been solved. So what’s next? In M. de Vega, A. M. Glenberg & A. C. Graesser (Eds.). *Symbols and embodiment: Debates on meaning and cognition* (pp. 223–244). New Haven: Academic Press.

Stöckle-Schobel, R. (2012). Perceptual learning and feature-based approaches to concepts – a critical discussion. *Frontiers in Psychology*, 3: 93. doi: 10.3389/fpsyg.2012.00093.

Sugita, Y., & Butz, M. V. (2011). Compositionality and embodiment in harmony. In Oudeyer, P.-Y. (Ed.), *AMD Newsletter, IEEE CIS*, 8, 8–9.

Sugita, Y., & Tani, J. (2005). Learning semantic combinatoriality from the interaction between linguistic and behavioral processes. *Adaptive Behavior*, 13, 33–52.

Sugita, Y., Tani, J., & Butz, M. V. (2011). Simultaneously emerging Braitenberg codes and semantic compositionality. *Adaptive Behavior*, 19, 295–316.

Tani, J. (2003). Learning to generate articulated behavior through the bottom-up and the top-down interaction processes. *Neural Networks*, 16, 11–23.