IEEE CIS

AMD NEWSLETTER

The Newsletter of the Autonomous Mental Development Technical Committee

Developmental Robotics
Machine Intelligence
Neuroscience
Psychology

Volume 10, number 2
Fall 2013

Editorial

Pierre-Yves Oudeyer

Inria and Ensta
ParisTech, France
Editor
pierre-yves.oudeyer@inria.fr

During the last decade of cognitive sciences, language has often played a pivotal role in the organisation of scientific enquiry. This is today the case in the study of development. Cognitive development in infants happens as a complex dynamical process, and for this reason many have argued for the necessity to use computational and robotic models as tools to refine our understanding. This makes sense, but only if a true and deep dialogue takes place between specialists of human development and specialists of computational modelling. Establishing such dialogue is perhaps the main challenge for our research community. Work around language development has been an example to follow. As illustrated in this issue of the newsletter, linguists, developmental psychologists and roboticists have made great steps forward to work in the same garden, growing together the trees of knowledge, and learning from each other.

Dialog initiated by Peter Ford Dominey investigates how structured embodied meanings, such as the temporal enfolding of events, and grammatical constructions to talk about these meanings, can be developed. Brian McWhinney, William Croft, Alistair Knott, Michael Arbib and Victor Barrès, and Juyang Weng propose their own perspective. A particularly illuminating hypothesis is that structured conceptual and linguistic meanings form within the flow of structured social interaction loops, which themselves act as a frame that guides the infant. This process of learning shall not be conceptualized as a process where building blocks or atoms of conceptual structures are first learnt, then combined, but rather as a process starting from learning how to pragmatically deal with global situations, which are then later on decomposed into subunits. Further, it is actually suggested that linguistic grammatical constructions interact with a sensorimotor system which has itself a full-blown grammatical organization, which opens in itself stimulating research avenues.

Such interaction between the formation of linguistic and conceptual/sensorimotor structures is taken as the starting point of Katerina Pastra's new dialog initiation, entitled "Autonomous acquisition of sensorimotor experiences: Any role for language?". In particular, she formulates a bold hypothesis: language as a communication system may have evolved as a byproduct of language as a tool for (self-)organizing conceptual structures. Those of you interested in reacting to this dialog initiation are welcome to submit a response by March 15th, 2013. The length of each response must be between 600 and 800 words including references (contact pierre-yves.oudeyer@inria.fr).

The issue of the role and study of social interaction loops as structures bootstrapping efficient social learning is also the core topic of the latest special issue of the IEEE TAMD journal, entitled "Microdynamics of interaction: Capturing and modeling infants' social learning" and coordinated by Katharina Rohlfing and Gedeon Deak.

Last but not least, I would like to introduce Fabien Benureau, PhD researcher at Inria Flowers (France), as the new editorial assistant of the newsletter, taking over from Adrien Baranes, and thank Fabien for accepting this job essential to the newsletter, as well as Adrien for all the great work achieved in the previous years.

Seizing the opportunity of the approaching tenth anniversary of the Newsletter with continuously renewed scientific dialogues, a new "look and feel" was designed by Fabien, and I hope everyone will enjoy it.

AMDTC Chair's Message

Angelo Cangelosi

University of Plymouth,
England
AMDTC Chair

acangelosi@plymouth.ac.uk

The Third Joint IEEE ICDL-EpiRob Conference, held in Osaka on August 2013, was a great success in terms of quality of papers and number of participants. All plans are now in place for the next joint conference that will be held in Genoa, Italy, on 13-16 October 2014 (paper submission deadline in April 2014). In parallel, the TC will organise a Special Session on "Cognition and Development" at the WCCI-2014 Conference in Beijing, 6-11 July 2014 (paper submission deadline 20 December 2013). The IEEE TAMD journal continues its success with a new impact factor of 2.17.

Following our "lively" brainstorming on the new name, the IEEE Computational Intelligence Society (CIS) has recently set up an "Ad-Hoc Committee on Bolstering the AMD Community". This has the task of listening to the wider AMD and developmental robotics community to identify a new name for the TC, the TAMD journal, and the joint conference. The Ad-Hoc Committee will report to the IEEE CIS AdCom in the coming months. Finally, as I am at the end of my 2-year tenure as AMD TC Chair, I would like to thank all the members of the Technical Committee, the Task Force Chairs and members, and the TC Deputy Chairs Yukie and Matt for their commitment and contribution to the TC life, and in general to the AMD and developmental community.

Table of Contents

Dialogue

New Dialogue Initiation

IEEE TAMD Table of Content

Dialogue

How are Grammatical Constructions Linked to Embodied Meaning Representations?

Peter Ford Dominey

CNRS & INSERM U846
Stem Cell and Brain
Research Institute
Robot Cognition
Laboratory
Lyon, France

peter.dominey@insem.fr

There is a strong tendency in modern cognitive neuroscience to adopt the perspective that the comprehension of meaning is achieved in the brain through the mental simulation of that meaning. In this embodied meaning context, brain networks involved in the active perception of an event would also participate in the re-presentation of such an event when reading a sentence describing that event. A wealth of behavioral and neurophysiological studies, including those arguing for a human "mirror system" support this position. These mental simulations have been referred to in a number of contexts including Johnson-Laird's mental models (Johnson-Laird 1980, 1983, Garnham et al. 1982), Barsalou's perceptual symbol systems (Barsalou 1999, Barsalou et al. 2003), and situated simulation models (Barsalou 2009). Major open issues remain, however.

The first issue concerns the development of language: how children learn to use grammar that allows the specification of the temporal unfolding of events in simulation? Does the progressive increase in the complexity of grammatical constructions that are used in development correspond to a developing capability to mentally represent? Do these capacities for language and simulation co-develop, and is there a dependency relation? We have attempted to address these issues of co-development (Dominey & Boucher 2005a, 2005b), arguing for a form of "conceptual bootstrapping" (Dominey 2002), where the conceptual system provides structure on which the grammatical system is built (Dominey 2000, 2003). But what if man's unusually developed simulation capability owes part of its power to language? It is also potentially problematic that our method of linking language to meaning required a propositional link between the perceptual system and the language system.

This raises a second important question concerning how these simulations are linked to language. In particular, is there a direct link between language and simulation? Or must the target of simulation be coded symbolically? Can the language system directly access simulations via multimodal convergence zones (Meyer & Damasio 2009)?

A third major open issue concerns the details of how these simulations are managed, the unrolling of these simulations in time, and how grammatical structure orchestrates the "internal film" of mental simulation. Madden and colleagues (Madden et al. 2010) hold that "language allows the speaker to ''direct the film", to precisely control the initiation, unfolding and termination of appropriate simulations in the mind of the listener, through precise grammatical mechanisms that have evolved for this purpose". Bergen and Chang (Bergen & Chang 2005) have begun to address how verb aspect can influence simulation.

Thus, it is likely that comprehension involves a coordinated cooperation between amodal linguistic representations, and modal simulations, as proposed in the Language and Situated Simulation theory of Barsalou (Barsalou et al. 2008). This raises important questions for future research (Pezzulo et al. 2011): How does grammar connect to mental simulations, to turn them on and off, to fast forward them? How does grammatical aspect (which allows the specification of the completion status of actions) interact with situated simulations?

Johnson-Laird, P. Mental Models in Cognitive Science, Cognitive Science, 4, pp. 71-115, 1980.

Garnham, A., Oakhill, J. and Johnson-Laird, P. N. Referential continuity and the coherence of discourse, Cognition, 11, pp. 29-46, Jan 1982.

Johnson-Laird, P. N. Mental models: Towards a cognitive science of language, inference, and consciousness, Harvard University Press, 1983.

Barsalou, L. W. Perceptual symbol systems, Behav Brain Sci, 22, pp. 577-609; discussion 610-60, Aug 1999.

Barsalou, L. W., Simmons, W. Barbey, A. and Wilson, C. Grounding conceptual knowledge in modality-specific systems, Trends Cogn Sci, 7, pp. 84-91, Feb 2003.

Barsalou, L. W. Simulation, situated conceptualization, and prediction, Philos Trans R Soc Lond B Biol Sci, 364, pp. 1281-9, May 12 2009.

Dominey P. and Boucher, J. Learning to talk about events from narrated video in a construction grammar framework, Artificial Intelligence, 167, pp. 31-61, 2005.

Dominey P. and Boucher, J. Developmental stages of perception and language acquisition in a perceptually grounded robot, Cognitive Systems Research, 6, pp. 243-259, 2005.

Dominey, P. F. Conceptual grounding in simulation studies of language acquisition, Evolution of Communication, 4, pp. 57-85, 2002.

Dominey, P. F. Conceptual grounding in simulation studies of language acquisition, Evolution of Communication, 4, pp. 57-85, 2000.

Dominey, P. F. A conceptuocentric shift in the characterization of language: Comment on Jackendoff, Behavioral and Brain Sciences, 26, pp. 674-674, 2003.

Meyer, K. and Damasio, A. Convergence and divergence in a neural architecture for recognition and memory, Trends Neurosci, 32, pp. 376-82, Jul 2009.

Madden, C. Hoen, M. and Dominey, P. F. A cognitive neuroscience perspective on embodied language for human-robot cooperation, Brain Lang, 112, pp. 180-8, Mar 2010.

Bergen, B. and Chang, N. Embodied construction grammar in simulation-based language understanding, Construction grammars: Cognitive grounding and theoretical extensions, pp. 147-190, 2005.

Barsalou, L. W., Santos, A. Simmons, W. and Wilson, C. Language and simulation in conceptual processing, Symbols, embodiment, and meaning, pp. 245-283, 2008.

Pezzulo, G., Barsalou, L. W., Cangelosi, A. Fischer, M. McRae, K. and Spivey, M. J. The mechanics of embodiment: a dialog on embodiment and computational modeling, Front Psychol, 2, p. 5, 2011.

Using Perspective to Construct Mental Models

Brian MacWhinney

Canergie Mellon
University
Pittsburgh, US

macw@cmu.edu

Peter Dominey asks whether grammatical constructions are linked developmentally to embodied meaning representations. As he notes, the idea that sentence comprehension involves the construction of a cognitive simulation has been an ongoing theme in Cognitive Linguistics since at least Chafe (1974) and MacWhinney (1977). More recently, psycholinguistic and cognitive neurolinguistic investigations have demonstrated ways in which these simulations are constructed by the brain (MacWhinney, 2009). But this growing understanding has not yet been systematically applied to the study of linguistic influences on children's cognitive development.

On a general level, Vygotsky, Bruner, Nelson, Rogoff, and others have made it clear that, by learning the scripts of narratives, folk tales, songs, rituals, and conversational sequencing, human children are able to develop and elaborate mental simulations that go beyond those available to non-human primates. But, underneath these more elaborate structures are a myriad of constructional formulas that respond to particular interactional configurations. At first, these are limited to "please" and "want"; then children pick up "please give me" and "let's play"; and eventually they include forms such as "I understand that" or "why didn't you ask me"? The literature on child language acquisition shows in detail how various item-based positional patterns (MacWhinney, 1975) give rise to generalized constructions that are eventually linked into fuller discourse. Basically, development emerges from the interplay between episodic rote storage and subsequent generalization (MacWhinney, in press).

Between the ages of 3 and 8, children are still struggling to link words and constructions into coherent discourse (Karmiloff-Smith, 1986). In large part this is because they do not yet have the devices and the coordination between devices needed to express the perspective shift required by narratives.They may encode a few stories such as *Three Little Pigs* in detail, but they drop segments, lose perspective, miscode anaphora, and cannot generalize the sequential patterns in these stories to attain overall linguistic control of cognitive simulations.

Dominey asks how mental simulations are managed in real time. The Perspective Hypothesis (MacWhinney, 2008) claims that this is done by using grammatical devices (passive, extraposition, cleft, relativization, aspect, tense, definiteness, anaphora, etc.) in both production and comprehension to mark and decode perspective maintenance and shifting within the internalized mental stage. This means that, to properly compose narratives, interactional plans, and social persuasion, children need to acquire as many as 28 major syntactic constructions (such as passive, extraposition, relativization, complementation, etc.) and must learn to link them together fluently to mark perspective shifting.

Constructions target representations in mental model space. Dominey asks whether these representations are coded symbolically and whether they involve multimodal convergence zones. We can best address this question by extending the theory of embodied cognition to deal with generalized semantic/thematic roles, such as Agent or Location. Although the specific realizations of thematic roles vary markedly from language to language, each system can be characterized in terms of embodied grounding. These generalized roles involve typical stances between actors, recipients, and objects that are undoubtedly also available to other mammals. There is no need for genetic specification of such general role processors, because they will emerge uniformly in all mammals through continual perception of the results of actions on the world (Piaget, 1954).

Chafe, W. (1974). Language and consciousness. Language, 50, 111-132.

Karmiloff-Smith, A. (1986). Language development after age 5. In P. Fletcher & M. Garman (Eds.), Language acquisition. Cambridge: Cambridge University Press.

MacWhinney, Brian. (1975). Pragmatic patterns in child syntax. Stanford Papers And Reports on Child Language Development, 10, 153-165.

MacWhinney, Brian. (1977). Starting points. Language, 53, 152-168.

MacWhinney, Brian. (2008). How mental models encode embodied linguistic perspectives. In R. Klatzky, B. MacWhinney & M. Behrmann (Eds.), Embodiment, Ego-Space, and Action (pp. 369-410). Mahwah: Lawrence Erlbaum.

MacWhinney, Brian. (2009). The emergence of linguistic complexity. In T. Givon & M. Shibatani (Eds.), Linguistic complexity (pp. 405-432). New York: Benjamins.

MacWhinney, Brian. (in press). Language development. In L. Liben & U. Mueller (Eds.), Handbook of Child Psychology and Developmental Science, 7th Edition. New York: Wiley.

Piaget, Jean. (1954). The construction of reality in the child. New York: Basic Books.

Do We Need Propositional Representations Between Language and Embodied Meanings?

William Croft

Department of Linguistics, University of New Mexico

wcroft@unm.edu

The vast majority of linguists assume the necessity of a separate propositional/symbolic representation (henceforth PSR) intervening between language and the reality of an embodied human being or robot interacting with the world. Even linguists who advocate an embodied semantic representation find themselves positing an intermediate PSR (Bergen & Chang 2005, Madden *et al.* 2010). It would be better if we could dispense with that intervening level, and employ direct interpretation (Montague 1974), albeit with a very different linguistic representation and model than Montague envisioned. What is it that the PSR buys us that would have to be achieved in some other way in a theory that directly links linguistic expressions to simulation models (or actual human interactions)?

First, a PSR is intended to provide a "cleaned-up" version of a syntactically messy, grammatically complex, and semantically multiply ambiguous utterance into a straightforward conceptual structure that can be directly related to the world (or a mental simulation). Second, a PSR is intended to impose some structure on a situation/simulation, expressed mainly by predicate-argument relations employed to decompose a scene and to categorize individual entities, and connectives linking predicate-argument structures to each other. Third, an important side effect of this propositional analysis of a situation is the ability to recombine the elements of the structure in order to conceptualize imagined experience. Fourth, a PSR is intended to provide a relatively stable abstraction across the indefinitely large number of interactions that a human (or robot) has engaged in and will engage in, with other agents in the world.

The "messy" nature of human language syntax is a consequence of the evolution of grammatical constructions in language use. One can capture the symbolic relation between a linguistic expression and a (nonpropositional) function by recognizing the existence of constructions as primitive units of complex grammatical structure. The reality of human language is that constructional forms are highly variable across languages (Croft 2001, Croft 2013) and that constructions are very specific low-level combinations of morphemes and categories, albeit related to each other in networks of similarity. The computational model of construction grammar that comes closest to a Radical Construction Grammar representation (Croft 2001, Croft 2013) is Fluid Construction Grammar (Steels 2011).

A construction is a complex unit made up of parts (elements). These elements and the constructional whole symbolize a conceptual representation - more precisely, a conceptualization (construal) of a human experience - but that representation need not be propositional in order to represent the structuring of a situation. For example, predicate-argument structure is linguistically expressed in argument structure constructions, and the components can be represented geometrically in terms of the actions of the participants unfolding over time and their causal interactions (Croft 2012). The geometric representation's schematic structure is the structure of a mental simulation, not an amodal symbolic proposition. If the nonpropositional representation is decomposable, then it is also recombinable in order to represent imagined experience.

Although the number of instances of a conceptual category is indefinitely large, exemplar approaches to linguistic categorization treat semantic categories in terms of relations between specific instances distributed in a conceptual space rather than as a schematic symbolic unit (Croft 2010, Bybee 2013). These instances need not be propositional, of course: they may be mental simulations, and their "decomposition" is actually implicit in their position in different dimensions of the conceptual space.

In other words, one need not mediate the relationship between a linguistic expression and a mental simulation (or actual perceptual-motor activity) with a PSR. The purposes to which a PSR is put in theories of language comprehension and production can be achieved by a radical constructional approach to syntactic structure, a direct representation of conceptual structure in terms of (mental simulations) of the structure of events, and representation of categories in terms of relationships among exemplars of events in conceptual space. In fact, linguistic expressions themselves should be represented in the same way: instances of complex perceptual-motor actions or simulations, organized by their distribution in the conceptual space of auditory-articulatory events.

Bergen, B. and Chang, N. Embodied construction grammar in simulation-based language understanding, Construction grammars: Cognitive grounding and theoretical extensions, pp. 147-190, 2005.

Madden, C. Hoen, M. and Dominey, P. F. A cognitive neuroscience perspective on embodied language for human-robot cooperation, Brain Lang, 112, pp. 180-8, Mar 2010.

Montague, R. Formal philosophy: selected papers of Richard Montague, Harvard University Press, 1974.

Croft, W. Radical Construction Grammar: syntactic theory in typological perspective, Oxford University Press, 2001.

Croft, W. Radical Construction Grammar, The Oxford handbook of construction grammar, pp. 211-232, 2013.
Steels, L. Design patterns in Fluid Construction Grammar, John Benjamins, 2011.
Croft, W. Verbs: aspect and causal structure, Oxford University Press, 2012.
Croft, W. Relativity, linguistic variation and language universals, CogniTextes 4.303, http://cognitextes.revues.org/303/, 2010.
Bybee, J. L. Usage-based theory and exemplar representations of constructions, The Oxford handbook of construction grammar, pp. 49-69, 2013.

Can Chomsky's Minimalism Help Us Express a Model of Embodied Meaning Representations in Language?

Alistair Knott

Dept of Computer
Science,
University of Otago,
New Zealand
alik@cs.otago.ac.nz

Dominey's questions are very timely. If, as many cognitive neuroscientists believe, linguistic constructions can activate simulations of sensorimotor experience, (i) how are these representations linked in the brain? and (ii) how do children learn the relationship between between these representations in their own mother tongue? I have recently published a book, *Sensorimotor Cognition and Natural Language Syntax* (MIT Press, 2012), which squarely addresses these questions.

My proposal adopts a Chomskyan model of syntax, Minimalism (Chomsky, 1995). Minimalism assumes a language-independent level of syntactic structure: each sentence has a *phonetic form (PF)*, encoding its surface structure, but also a *logical form (LF)* that encodes its semantic structure. While PF structures vary widely between languages, LF representations are largely invariant over translation. Minimalism's postulation of a language-independent LF structure makes it an interesting vehicle for embodied models of syntax. If syntax interfaces closely with the sensorimotor system, we expect there to be similarities between the syntax of different languages, reflecting their close relationship to the sensorimotor system, which is uncontroversially the same for all humans. While Chomskyans traditionally assume that cross-linguistic generalisations expressed at LF reflect the structure of innate *language-specific* knowledge, another possibility is that these generalisations have a sensorimotor origin. This is basically my claim. My proposal, more specifically, is that the LF of a sentence describing a concrete episode in the world can be interpreted as a *direct description of the sensorimotor processes through which this episode was experienced*.

The book illustrates this proposal through a detailed study of an episode in which a man grabs a cup. The processes involved in experiencing grasp actions have been particularly thoroughly studied. I develop a model of these processes, synthesising evidence from many separate studies. The key idea is that experiencing a grasp episode involves a canonical *sequence* of sensorimotor operations. First, the agent of the action is attended to. If the observer is himself the agent, 'attending to the agent' happens through an early *decision to act*, that occurs before a detailed decision about which action to perform (Haggard, 2008). If the observer is a third party, attending to the agent involves allocating attention to an external agent, and configuring the mirror system (Gallese et al., 1996) for action observation. Next, the target of the action is attended to. If the observer is the agent, the target is fixated in order to activate its grasp affordances (Fagg and Arbib, 1998). If the observer is watching the agent, she fixates the agent's intended target early during action monitoring, using joint attention and other postural cues (Webb et al., 2010). Finally, an action monitoring routine is initiated, in which the motor programme 'grasp' is activated and used to encode the hand's trajectory onto the target. If the observer is the agent, the motor programme is selected as a learned behavioural response. If the observer is watching the agent, the motor programme is activated through the circuitry of the mirror system for action perception (Gallese et al., 1996). In either case, I argue that monitoring the action involves *reattending* to both agent and patient, in different sensorimotor modalities. The agent, initially attended to as a static entity, is now reattended to as a dynamic entity, with characteristic patterns of movement. At the end of the action, the target, initially attended to in the visual modality, is reattended to in the haptic modality, as a motor state of the hand/arm. I argue that episode-perception routines are essential for learning cross-modal representations of agents and objects, and incorporate mechanisms for learning these correspondences.

The idea that sensorimotor experiences are naturally organized into sequences is a key part of Ballard et al.'s (1997) model of *deictic routines*. I argue that experiencing any concrete episode involves a sequentially-structured deictic routine. The building blocks of deictic routines are *deictic operations*: attentional or motor actions that function to change the observer's physical relation to the environment. In my model, observers not only use deictic routines to experience concrete episodes, but also to *represent* episodes in working memory, as prepared sequences of sensorimotor operations (see Takac and Knott, 2012 for a computational model). Thinking of episode representations as prepared sequences has two advantages. Firstly, we know a lot about how prepared action sequences are stored in the brain, so the proposal subsumes episode representations within a category of neural representation that is relatively well understood. Secondly, thinking of episode representations as prepared action sequences lends itself well to a simulationist account of meaning, since a prepared action sequence is naturally something that can be executed in simulation.

The concept of basic building blocks also features in Minimalism. LF structures are recursively formed from instances of the *X-bar schema*, predominantly organized into right-branching structures. I propose that each X-bar schema in the LF of a sentence reporting a concrete episode describes a single deictic operation in the sensorimotor routine through which the episode is experienced, and that the right-branching structure of X-bar schemas in the LF structure indicates the sequential order of the associated deictic operations. In summary, in response to Dominey's first question, I propose there is a very general, direct relationship between syntactic and sensorimotor representations, which can be stated within a Minimalist framework, as a correspondence between the basic building blocks of each type of representation.

The second question concerns how infants learn a mapping between sensorimotor and syntactic structures. In Minimalism, syntactic development is largely a matter of parameter-setting. At LF, the subject and object each appear at two positions, and the verb appears at multiple positions; infants must learn which of these alternative positions correspond to the 'surface' positions of subjects, objects and verbs. The book introduces a computational model of this parameter-setting process, expressed within a neural network for sentence generation (Takac et al., 2012). In this model, generating a sentence reporting a given episode involves *replaying* the deictic routine representing this episode, in a special mode in which sensorimotor signals can trigger overt phonological outputs. There are two opportunities to pronounce the subject and object, corresponding to the two agent and patient representations activated in the sensorimotor routine, and many opportunities to pronounce the verb, which is read from tonically active representations in the prefrontal circuit storing the prepared routine (Averbeck et al., 2002). The model's training inputs are mature sentences in a given language paired with deictic routines: from this data it can learn languages with any possible constituent order (SVO, VSO and so on). Interestingly, the model can also learn idiomatic surface structures in the exposure language, of the kind that are problematic for traditional Minimalism. I argue that a sensorimotor interpretation of LF allows insights from Minimalism to be combined with insights from empiricist models of language development in a way that has not previously been possible.

Averbeck, B. et al. (2002). Parallel processing of serial movements in prefrontal cortex. PNAS, 99(20), 13172–13177.
Ballard, D., et al. (1997). Deictic codes for the embodiment of cognition. BBS, 20(4), 723–767.
Chomsky, N. (1995). The Minimalist Program. MIT Press, Cambridge, MA.
Fagg, A. and Arbib, M. (1998). Modeling parietal-premotor interactions in primate control of grasping. Neural Networks, 11(7/8), 1277–1303.
Gallese, V. et al. (1996). Action recognition in the premotor cortex. Brain, 119, 593–609.
Haggard, P. (2008). Human volition: Towards a neuroscience of will. Nature Reviews Neuroscience, 9, 934–946.
Takac, M. and Knott (2013). A neural network model of working memory for episodes. Proceedings of CogSci, 1432-1437.
Takac, M., et al. (2012). Mapping sensorimotor sequences to word sequences: A connectionist model of language acquisition and sentence generation. Cognition, 125, 288–308.
Webb, A. et al. (2010). Eye movements during transitive action observation have sequential structure. Acta Psychologica, 133, 51–56.

Are Grammatical Constructions Linked to Embodied Meaning Representations?

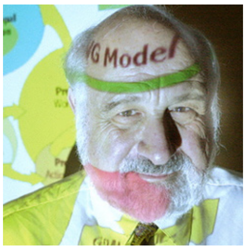
Michael A. Arbib & Victor Barrès

The First Two Sentences

Dominey asks "How are Grammatical Constructions Linked to Embodied Meaning Representations?" We ask "*Are* Grammatical Constructions Linked to Embodied Meaning Representations?" and our answer is "Maybe. Sometimes." In a forthcoming paper (Arbib, Gasser, & Barrès, 2014), we defend this view at length, but here we illustrate the point by analyzing Dominey's first two sentences.

(1) *There is a strong tendency in modern cognitive neuroscience to adopt the perspective that the comprehension of meaning is achieved in the brain through the mental simulation of that meaning.*

Dominey proceeds without ever calling this *tendency* into question. Moreover, this notion of *simulation* is not defined. Consider the sentence "He unlocked the door." How did you understand it? Did you simulate *seeing* someone unlock a door? Was this simulation amodal (purely visual) or was there a "sound track"? And did the door opener have a full



Michael A. Arbib

Computer Science,
Brain Project,
Neuroscience Graduate Program,
University of Southern California, US
arbib@usc.edu

Victor Barrès

Brain Project,
Neuroscience Graduate Program,
University of Southern California, US
barres@usc.edu

head of black hair? Surely you would have noticed had you observed an actual event? Or did you simulate the *motor performance*, the feel of a key in your hand as you twisted a key in a lock and then pulled on the door handle? What if there were no handle, and you pushed the door to open it? Which was the "right" simulation? And if you ran the motor simulation and you are female, how did you simulate the maleness of the door opener? Suddenly, the notion of modal simulation seems woefully inadequate – or overloaded with extraneous detail.

(2) *In this embodied meaning context, brain networks involved in the active perception of an event would also participate in the re-presentation of such an event when reading a sentence describing that event.*

Even if this were true, it would not prove the tendency correct, for all it asserts is (2a) If a sentence describes an event which has been [or *could be*] perceived, then reading the sentence will [to some extent] activate the brain networks involved in the active perception of the event.

But this does not tell us that such activation in itself constitutes the full process of understanding the meaning of even this limited class of sentences. Moreover, (2) does not describe an event and thus is silent about its own meaning.

Rather than addressing such concerns, Dominey raises three issues: 1) How do children learn to use grammar that allows the specification of the temporal unfolding of events in simulation? 2) Is there a direct link between language and simulation? 3) How are these simulations managed, and how does grammatical structure orchestrate the unrolling of the "internal film" of mental simulations? Since space is limited and other commentators will surely address these issues, we turn to a glaring omission: Even though (1) mentions "modern cognitive neuroscience" there is no neuroscience in Dominey's piece, other that an uncritical mention of mirror neurons, *en passant*. This is very strange since Dominey is himself a gifted contributor to computational neuroscience and the testing of models through human brain imaging. We thus turn to a brief exposition of Dominey's neural modeling of constructions.

On Dominey's Neural Construction Grammar

Dominey and his colleagues (Dominey, Hoen, Blanc, & Lelekov-Boissard, 2003; Dominey, Hoen, & Inui, 2006; Dominey & Inui, 2009) have sought to understand how *brain models* can simulate the learning and use of the form-meaning mappings in constructions. They anchor modeling of the language system on previous modeling of the macaque oculomotor system which demonstrated the central role played by cortico-striatal connections in sequence processing (Dominey, Arbib, & Joseph, 1995). In order to link construction processing to sequential cognition, Dominey et al. (2006) indexed constructions by the sequence of function words they contain. For example, the passive construction used in "The ball is kicked by the boy" was characterized by the sequence "X Y by Z", while the object-cleft construction used in "It is the ball that the boy kicked" was associated with the sequence "It is X that Y Z". The authors showed (in principle) how an interlinked set of cortical and subcortical neural nets could learn to detect sequences of function words associated with different constructions, and then use the construction so retrieved to map content words onto their proper thematic roles (X onto Patient, Y onto the Action, and Z onto Agent in the passive construction). This work does not address the challenge of linking constructions hierarchically to generate complex sentences, but it offers a constructive [sic] first step towards a computational neurolinguistics. More recent work (Hinaut & Dominey, 2013) simulates prefrontal area BA47 as a recurrent network with plastic connections between cortex and striatum. The system is trained on sentence-meaning pairs, where meaning is coded as activation in the striatum corresponding to the roles that different nouns and verbs play in the sentences, and learns an extended set of grammatical constructions. Franklin Chang and colleagues (Chang, Dell, & Bock, 2006; Chang, Janciauskas, & Fitz, 2012) have developed somewhat related connectionist models assessed in relation to psycholinguistic data.

Although inspired by the oculomotor system, there is nothing specifically "embodied" in Dominey's neural models. Thus, Dominey seems not to side with those who see language processing as primarily within the brain's capacity to mediate embodiment, and so the link to embodiment of this model – for those sentence which do report embodied sensorimotor experience – remains very much open. It is thus an intriguing challenge for computational neuroscience to make further contact between Dominey's work and three other flavors of computational construction grammar: *Embodied construction grammar*, which is rooted in motor schemas and the extension of "embodied cognition" via metaphor (Feldman & Narayanan, 2004; Narayanan, 1997); *template construction grammar*, which addresses the production and comprehension of descriptions of visual scenes (Arbib & Lee, 2008; Barrès & Lee, 2013); and *fluid construction grammar*, which uses evolutionary games between embodied agents as the basis for studying the formation and dynamics of shared meaning (Beuls

& Steels, 2013).

Arbib, M. A., Gasser, B., & Barrès, V. (2014). Language is Handy but is it Embodied? Neuropsychologia, in press.
Arbib, M. A., & Lee, J. Y. (2008). Describing visual scenes: Towards a neurolinguistics based on construction grammar. Brain Research, 1225, 146-162.
Barrès, V., & Lee, J. Y. (2013). Template Construction Grammar: From Visual Scene Description to Language Comprehension and Agrammatism. Neuroinformatics, DOI 10.1007/s12021-12013-19197-y.
Beuls, K., & Steels, L. (2013). Agent-Based Models of Strategies for the Emergence and Evolution of Grammatical Agreement. PLoS One, 8, e58960.
Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. Psychol Rev, 113, 234-272.
Chang, F., Janciauskas, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning. Language and Linguistics Compass, 6, 259-278.
Dominey, P. F., Arbib, M. A., & Joseph, J.-P. (1995). A model of corticostriatal plasticity for learning oculomotor associations and sequences. Journal of Cognitive Neuroscience, 7, 311-336.
Dominey, P. F., Hoen, M., Blanc, J.-M., & Lelekov-Boissard, T. (2003). Neurological basis of language and sequential cognition: Evidence from simulation, aphasia, and ERP studies. Brain and language, 86, 207-225.
Dominey, P. F., Hoen, M., & Inui, T. (2006). A neurolinguistic model of grammatical construction processing. J Cogn Neurosci, 18, 2088-2107.
Dominey, P. F., & Inui, T. (2009). Cortico-striatal function in sentence comprehension: Insights from neurophysiology and modeling. Cortex, 45, 1012-1018.
Feldman, J., & Narayanan, S. (2004). Embodied meaning in a neural theory of language. Brain Lang, 89, 385-392.
Hinaut, X., & Dominey, P. F. (2013). Real-Time Parallel Processing of Grammatical Structure in the Fronto-Striatal System: A Recurrent Network Simulation Study Using Reservoir Computing. PLoS One, 8, e52946.
Narayanan, S. S. (1997). Knowledge-based Action Representations for Metaphor and Aspect (KARMA). University of California at Berkeley.

These Questions Arose because You Used Symbolic Representations

Juyang Weng

Department of
Computer Science and
Engineering,
Cognitive Science
Program,
Neuroscience
Program,
Michigan State
University, US

weng@cse.msu.edu

Peter Dominey's dialog raised some interesting questions that are faced by many language system researchers.

Dominey's questions arose because he used symbolic representations (Weng 2012). I do know that Dominey used neural networks such as reservoir networks in his models. Partial neural networks and partial symbolic representations still amount to an overall symbolic representation (Weng 2012). For example, in Dominey's representation, each word has one of handcrafted thematic (grammar) roles, Agent, Predicate, Object, and Recipient.

In a Developmental Network (DN) (Weng et al. 2010, Miyan & Weng 2013), all extra-body concepts are considered actions in one of the motor areas and they all should be learned, just like any other actions. A combination of distributed and parallel actions amounts to an emergent state of the brain, not the internal firing state of brain neurons many researchers have thought! These emergent states are teachable, observable, and calibratable through physical interactions (including communications) with the external environment that includes teachers.

For example, in Dominey's model a word Mary is learned, but a concept Agent is not (in the genes?). Then, Mary and Agent belong to different modules, inconsistent with autonomous development.

Dominey asked: "How does grammar connect to mental simulations, to turn them on and off, to fast forward them?"

I have a fundamentally different view on brain's way of doing language. According to my understanding of how the brain works, through a synthesis from a vast amount of literature in biology, neuroscience, psychology, and cognitive sciences, our brain-inspired DN model (Weng et al. 2010, Miyan & Weng 2013) gives the following answer to Dominey's questions:

First, the brain's developmental mechanisms are basically the same for vision and language. I do not mean that every brain area contributes equally to vision and language. However, vision and language are not separable in many settings and they often take place concurrently. Their brain principles are basically the same. Dominey stated "our method of linking language to meaning required a propositional link between the perceptual system and the language system." That should not be the case for autonomous development, as I explain below.

According to our DN theory, the perceptual system and the language system are within each other seamlessly. For example, when DN sees "Mary" in a clutter background, its "What" Motor TM area says "Mary" (as a language) and its "Where" Motor (LM) area has a firing neuron that corresponds to the relative location of Mary in the "retina". Therefore, vision and language can be seamlessly integrated into the same network.

Second, there should not be (A) a grammar module and (B) a mental simulation module separately in developmental system. Grammar rules, if they emerge, are either implicit (e.g., before entering school) or explicit (e.g., after attending school) as parallel firing patterns in the motor areas.

For example, that Mary is associated with Agent and Hit is associated with Predicate in grammar is dealt with by DN just like any other concept, such as Sparrow is associated with Flying and Penguin is associated with non-flying (Miyan & Weng 2013). The DN thinking is considered an "off task" process (Slogi et al. 2013) while the network learns while performing. There are not a separate learning phase and a separate performing phase. Thus, there does not exist a question how (A) interacts with (B). (A) and (B) are emergent properties of DN whose execution and interactions are fully autonomously learned according to experience. In particular, the programmer does not need to answer whether (A) turns (B) on and off. Both interact with each other and many other capabilities in very complicated but autonomous ways — impractical to handcraft.

In summary, for autonomous development, no concept corresponding to the extra-body environment should be handcrafted. Otherwise, one uses symbolic representations. Any handcrafted concepts (e.g., grammar concepts) cannot be autonomously dealt with by a developmental agent. The DN theory has explained schematically why the network can interactively learn an open-ended array of concepts, including tasks, communication protocols, and grammar concepts as special cases.

J. Weng. Symbolic Models and Emergent Models: A Review, IEEE Transactions on Autonomous Mental Development, 4(1), pp.29-53, 2012.
J. Weng, M. D. Luciw and Q. Zhang, "Brain-Like Emergent Temporal Processing: Emergent Open States," IEEE transactions on Autonomous Mental Development, vol. 5, no. 2, pp. 89 - 116, 2013.
K. Miyan and J. Weng, "WWN-Text: Cortex-Like Language Acquisition with 'What' and 'Where'," in Proc. IEEE 9th International Conference on Development and Learning," Ann Arbor, pp. 280-285, August 18-21, 2010.
M. Solgi, T. Liu and J. Weng, "A Computational Developmental Model for Specificity and Transfer in Perceptual Learning," Journal of Vision, vol. 13, no. 1, ar. 7, pp. 1-23, 2013.

Linking Language to Meaning: Integrating Multiple Perspectives

Peter Ford Dominey

CNRS & INSERM U846
Stem Cell and Brain
Research Institute
Robot Cognition
Laboratory
Lyon, France

peter.dominey@insem.fr

It is a great pleasure to analyze these thoughtful comments on the target article. In the target I posed three questions. In this response I will make a structured synthesis of the comments from my distinguished colleagues with respect to these questions, and I will then conclude with a plan for the next steps in this human adventure.

1. How do children learn to use grammar that allows the specification of the temporal unfolding of events in simulation?

Language learning, and the relation between language and conceptual relations (in terms of the possible influence of language on conceptual representations) is a key point addressed by MacWhinney. He notes that Vygotsky, Bruner, Nelson, Rogoff and others have made it clear that the inherent narrative structure of social discourse provides a structuring of the perceived world that allows humans to develop elaborate mental models. Thus language contributes to a mental structure and simulation capability (vastly) superior to that of non-human primates.

MacWhinney crucially introduces the importance of "interactional configurations" in language learning. That is, he highlights the reality of social interaction that pervades human life from the outset, and the constructional formulas that structure these interactions. He notes how this provides the transition from item-based constructions to more generalized constructions in fuller discourse. This is consistent with Croft's radical construction grammar view, which holds that "Utterances are instances of constructions. In other words, from the point of view of the language learner (and the fieldworker), the larger units come first" (Croft 2005).

In order to clearly identify the link between language and learning, Knott proposes a model of language in which semantic representations are sensorimotor sequences, that can be stored and replayed (Takac et al. 2012). He notes two major advantages to this approach to encoding meaning – first we know much about the underlying physiology, and second, by their nature, such sequential representations are appropriate for simulation.

2. How are simulations linked to language?

Arbib and Barres note that though grounded in the corticostriatal sensorimotor sequencing system, our work in grammatical construction processing is not "particularly embodied." Indeed, it is very "structural" and there is at present no contact with semantic content (though we propose such a link in (Lallée et al. 2010, Madden et al. 2010)).

Croft suggests that we dispense with the intervening level of symbolic representations. Radical construction grammar also "discards the assumption that syntactic categories are the primitive elements of syntactic representation. Instead, constructions are the primitive elements of syntactic representation" (Croft

2005). The radical constructional approach to syntactic structure thus suggests "a direct representation of conceptual structure in terms of (mental simulations) of the structure of events."

Looking towards the possible implementation of such a conceptual structure, Knott provides a minimalist response to this question, where phonetic form (PF) encodes surface structure and logical form (LF) encodes semantic structure that is recursively formed from instances of the X-bar schema. Interestingly, Knott links x-bar schemata to sensorimotor experiences (based on Ballard et al. 1997) deictic routines. The right branching structure of the X-bar schemata corresponds to the sequential unfolding of the corresponding experiences coded as deictic representations, thus directly establishing the link between language and meaning.

Weng takes a strong stance on the problems associated with symbolic representations. He notes that rather than introducing an internal role such as "agent" from the outset, the system should instead learn the concept associated with agency. Weng thus suggests a more holistic account, where the language system and the meaning system are not distinct and separate systems, but rather they are seamlessly interlinked. This is implemented in his Dynamic Network (DN) model.

3. How does grammar allow the unrolling and orchestration of simulations?

Arbib and Barres expand on this question: Given "He unlocked the door", they ask how detailed is the simulation. For example, do we actually feel the key in the door?

Croft addresses this issue in the context of radical construction grammar (RCG) (Croft 2012). Simplifying, verb aspect (different ways of viewing the internal temporal constituency of a situation (after (Comrie 1976)) is manifest lexically and grammatically. RCG then apparently provides a framework for linking the specification of aspect in sentences directly into the semantic (mental simulation) representation of meaning. It remains to attempt to determine how to implement this. A central problem will be the characterization of the conceptual space in which meaning (including aspect) is represented.
Knot's suggestion that the representation of meaning is in the form of sensorimotor sequences based on deictic operations provides the direct link, where the X-bar representation extracted from the surface form will directly link to the associated SM sub-sequence. Extending this, we can consider that the more detailed the surface description, the higher spatio-temporal fidelity of the corresponding simulation.

MacWhinney identifies the grammatical devices (some 28 forms including passive, extraposition, aspect, tense, etc.) that are used to manage the temporal unfolding of simulations via language. In addition to the temporal structure, the thematic structure (who did what to whom) is also encoded grammatically. MacWhinney suggests that because these roles are so evidently available via the continual perception of action in the world, they will necessarily emerge in embodied representations.

Arbib and Barres review our work in the neurophysiological grounding of grammatical constructions. They recall that the basis of these grammatical constructions is in the cortico-striatal representations of the sensorimotor sequencing system. This returns to the sensorimotor simulation framework of Knott. Note that in (Hinaut & Dominey 2013), we demonstrate grammatical generalization to new complex constructions that were not used in training, and so from Knott's perspective this generalization to new grammatical forms will extend to the sensorimotor domain, to new forms of meaning representation.

Summary: This exchange has proved to be extremely fruitful. Part of the motivation behind this exchange is my goal to understand language and cognitive development through building it in a humanoid robot that interacts socially with humans, e.g. (Lallée et al. 2013, Petit et al. 2013). This exchange confirms that the narrative structure of social interaction provides the external scaffolding for the co-development of the conceptual system and the interlaced language system (the two are inextricably linked). By adhering to this embodied context – i.e. language accompanies (social) interaction - the link between sensorimotor representations of meaning and language can be directly exploited. That is, the language the developing child hears is her interactions – we do not learn disembodied sentences that are unrelated to the here and now (at least in early language development). Thus, early language will always be accompanied by physically/socially embodied meaning.

This meaning can be encoded in the corticostriatal sensorimotor sequences that were employed in perception and action. Interestingly, this confirms an earlier statement of such an approach in (Madden et al. 2010) where we indeed suggested that embodied meaning representations are coded in the cortico-striato-thalamo-cortical (CSTC) sensorimotor circuits, grammatical constructions encoded in a newly proposed "language" CSTC circuit, and narrative or situation constructions in a bilateral prefrontal CSTC "situation construction" circuit.

Proposals on the actual linking and control of such representations and sensorimotor simulations to/by verb aspect are provided by Croft, Knott, and MacWhinney, and in (Bergen & Chang 2005). The most specific of these appears to be Knott's, linking the X-bar structure of logical forms to deictic sensorimotor representations.

Our future work will include a more detailed analysis and synthesis of these proposals on linking temporal/narrative structure in grammar to the corresponding spatio-temporal sensorimotor representations, and an attempt to implement this linking whereby language can orchestrate the unfolding of embodied representations in a physical robot, based on its acquired experience with humans. This work has been initiated in the EU FP7 Project EFAA, and will be pursued in the new FP7 Project WYSIWYD.

Bergen B, Chang N. 2005. Embodied construction grammar in simulation-based language understanding. Construction grammars: Cognitive grounding and theoretical extensions: 147-90
Comrie B. 1976. Aspect: An introduction to the study of verbal aspect and related problems. Cambridge University Press.
Croft W. 2005. Logical and typological arguments for Radical Construction Grammar. Construction grammars: Cognitive grounding and theoretical extensions: 273-314
Croft W. 2012. Verbs: aspect and causal structure. Oxford University Press.
Ballard, D., et al. 1997. Deictic codes for the embodiment of cognition. BBS, 20(4), 723–767.
Hinaut X, Dominey PF. 2013. Real-time parallel processing of grammatical structure in the fronto-striatal system: a recurrent network simulation study using reservoir computing. PloS one 8: 1-18
Lallée S, Hamann K, Steinwender J, Warneken F, Martienz U, et al. 2013. Cooperative Human Robot Interaction Systems: IV. Communication of Shared Plans with Naïve Humans using Gaze and Speech. Presented at IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo
Lallée S, Madden C, Hoen M, Dominey P. 2010. Linking language with embodied teleological representations of action for humanoid cognition. Frontiers in Neurobotics
Madden C, Hoen M, Dominey PF. 2010. A cognitive neuroscience perspective on embodied language for human-robot cooperation. Brain Lang 112: 180-8
Petit M, Lallee S, Boucher J-D, Pointeau G, Cheminade P, et al. 2013. The Coordinating Role of Language in Real-Time Multi-Modal Learning of Cooperative Tasks. IEEE Transactions on Autonomous Mental Development 5: 3-17
Takac M, Benuskova L, Knott A. 2012. Mapping sensorimotor sequences to word sequences: A connectionist model of language acquisition and sentence generation. Cognition

New Dialogue Initiation

Autonomous Acquisition of Sensorimotor Experiences: Any Role for Language?

Katerina Pastra

Cognitive Systems
Research Institute &
Institute for Language
and Speech Processing
Athens, Greece

kpastra@csri.gr

Self-exploration of the world starts with the very first body movements, even from within the womb. As the motor system develops, such exploration becomes more complex and more efficient. It becomes also more multisensory, as all perceptual abilities develop radically too. However, some percepts have a special status, a symbolic one; speech, for example, is also there during self-exploration of the world and infants are attentive to it and affected by it, from the very first months of their life (Waxman et al. 2010). Beyond the traditional role of verbal communication for expressing intention and passing on knowledge/information, *does language play any other role in such context? Does it affect, facilitate, or enable this exploration of the world? If so, how? Could verbal communication be the epiphenomenon of more basic functions served by language?*

Recent years have seen an increasing body of experimental evidence suggesting a tight relation between language, perception and action. Part of this evidence sheds light on the role of the (visuo)motor system in language comprehension. For example, motor circuits of the brain have been shown to contribute to comprehension of phonemes, semantic categories and grammar (Pulvermuller and Fadiga 2010). Motor simulation has been found to be activated during language comprehension (Glenberg 2008). At a computational level, there is a large body of research on automatic action-language association (Pastra and Wilks 2004, Pastra 2008), in both intelligent multimedia systems and robotics. The research addresses the semantic gap problem between low-level processes and high-level analyses; its philosophical manifestation is the symbol grounding problem and the related debate on the need for artificial agents to ground symbols to sensorimotor experience for 'grasping' the meaning of the language they analyse or generate (Cangelosi 2010).

However, is such mapping needed only for efficient communication with others? Is it merely a sign of truly knowing the meaning of symbols/words? Is the language-motor system relation merely a one-directional one? What does language contribute to the (visuo)motor system, if anything?

There has been increasingly growing evidence that language contributes significantly to structuring sensorimotor experiences. In particular, it has been shown that in perceptual category formation, infants readily compute correlations between different modalities (Plunket et al. 2008). For instance, they correlate the name/label of an object and its visual appearance. This dual category representation (i.e. linguistic and visual) entails that verbal categories (of concrete concepts) comprise members with perceptual similarity.

Indeed, dual category representation creates expectations when a new object is perceived, or a known label is used. Familiar labels create expectations of the visual appearance of the objects to be applied to, so they allow inferences on the basis of the known label, which has not been shown to be the case when a novel verbal label is used (in the later case inferences are based on appearance only) (Smith et al. 2002). Furthermore, infants generalise familiar labels to object categories according to specific perceptual properties they have and there is universal tendency to do that: from single naming of object instances to generalisation of names of different kinds according to different perceptual properties (Smith et al. 2010).

Furthermore, developmental studies have indicated that when verbal labels are applied as a system (e.g. two different labels name different objects) they facilitate object discrimination, which is not the case with non-verbal labels, such as tones, sounds, and emotions (Lupyan et al.2007). This was shown for infants as young as 3 months old (Waxman et al. 2010). So, verbal categories (of concrete concepts) have distinctive perceptual characteristics, which allow one category to be distinctive in its denotation from another.

Actually, verbal labels per se have been shown to impose distinctiveness even in cases when perceptual similarity is inconclusive – as a sole criterion – for categorisation of an object to a familiar category. In experiments with 10 month old infants, the use of verbal labels was shown to have an impact on the categorisation of animal cartoon drawings to the extent that led the participants to override perceptual dissimilarities between objects and treat them as more similar to each other (Plunkett et al. 2008). In such case, language was shown to play a *causal role* in perceptual category formation during infancy.

So, what does naming (verbal labelling) of sensorimotor experiences enable? Is it just a communication mechanism? Is communication a by-product of an evolutionary basic functionality of language?

Addressing such questions can shed new light on language analysis itself, as well as on the development of cognitive, artificial agents.

Cangelosi, A. (2010) 'Grounding Language in Action and Perception: From Cognitive Agents to Humanoid Robots', Physics of Life Reviews, vol. 7, no. 2, pp. 139-151.
Glenberg A. (2008) 'Toward the Integration of Bodily States, Language and Action', in Semin G. and E. Smith (eds), Embodied Grounding: Social, Cognitive and Neuroscientific Approaches, ch. 2, pp. 43-70, Cambridge University Press.
Lupyan, G.; Rakison, D. & McClelland, J. (2007) 'Language is not just for talking: Redundant Labels Facilitate Learning of Novel Categories', Psychological Science, vol. 18, pp. 1077-1083
Pastra K. (2008) 'PRAXICON: The Development of a Grounding Resource', In Proceedings of the 4th International Workshop on Human-Computer Conversation, Bellagio, Italy.
Pastra K. and Y. Wilks (2004) 'Vision-Language Integration in AI: a reality check', In Proceedings of the 16th European Conference on Artificial Intelligence (ECAI), pp. 937-941, Valencia, Spain.
Plunkett K., Hu J. and Cohen L. (2008) 'Labels can override perceptual categories in early infancy' Cognition, vol. 106, pp. 665–681.
Pulvermueller F. and Fadiga L. (2010) 'Active perception: sensorimotor circuits as a cortical basis for language' Nature Reviews in Neuroscience, vol. 11, number 5, pp. 351-360.
Smith, L., Colunga, E. & Yoshida, H. (2002) 'Making an Ontology: Cross-Linguistic Evidence' Early Category and Concept Development, Cambridge University Press, pp. 275-302
Smith, L., Colunga, E. & Yoshida, H. (2010) 'Knowledge as Process: Contextually-Cued Attention and Early Word Learning' Cognitive Science, pp. 1-28
Waxman S. (2010) 'Categorization in 3- and 4-Month-Old Infants: An Advantage of Words over Tones' Child Development, vol. 81, pp. 472-479.

IEEE TAMD Table of Contents

Volume 5, Issue 2, June 2013

Brain-Like Emergent Temporal Processing: Emergent Open States

J. Weng, M. D. Luciw, and Q. Zhang

Informed by brain anatomical studies, we present the developmental network (DN) theory on brain-like temporal information processing. The states of the brain are at its effector end, emergent and open. A finite automaton (FA) is considered an external symbolic model of brain's temporal behaviors, but the FA uses handcrafted states and is without "internal" representations. The term "internal" means inside the network "skull." Using action-based state equivalence and the emergent state representations, the time driven processing of DN performs state-based abstraction and state-based skill transfer. Each state of DN, as a set of actions, is openly observable by the external environment (including teachers). Thus, the external environment can teach the state at every frame time. Through incremental learning and autonomous practice, the DN lumps (abstracts) infinitely many temporal context sequences into a single equivalent state. Using this state equivalence, a skill learned under one sequence is automatically transferred to other infinitely many state-equivalent sequences in the future without the need for explicit learning. Two experiments are shown as examples: The experiments for video processing showed almost perfect recognition rates in disjoint tests. The experiment for text language, using corpora from the Wall Street Journal, treated semantics and syntax in a unified interactive way.

A Simple Ontology of Manipulation Actions Based on Hand-Object Relations

F. Wörgötter, E. E. Aksoy, N. Krüger, J. Piater, A. Ude, and M. Tamosiunaite

Humans can perform a multitude of different actions with their hands (manipulations). In spite of this, so far there have been only a few attempts to represent manipulation types trying to understand the underlying principles. Here we first discuss how manipulation actions are structured in space and time. For this we use as temporal anchor points those moments where two objects (or hand and object) touch or un-touch each other during a manipulation. We show that by this one can define a relatively small tree-like manipulation ontology. We find less than 30 fundamental manipulations. The temporal anchors also provide us with information about when to pay attention to additional important information, for example when to consider trajectory shapes and relative poses between objects. As a consequence a highly condensed representation emerges by which different manipulations can be recognized and encoded. Examples of manipulations recognition and execution by a robot based on this representation are given at the end of this study.

An Autonomous Social Robot in Fear

A. Castro-González, M. Malfaz, and M. A. Salichs

Currently artificial emotions are being extensively used in robots. Most of these implementations are employed to display affective states. Nevertheless, their use to drive the robot's behavior is not so common. This is the approach followed by the authors in this work. In this research, emotions are not treated in general but individually. Several emotions have been implemented in a real robot, but in this paper, authors focus on the use of the emotion of fear as an adaptive mechanism to avoid dangerous situations. In fact, fear is used as a motivation which guides the behavior during specific circumstances. Appraisal of fear is one of the cornerstones of this work. A novel mechanism learns to identify the harmful circumstances which cause damage to the robot. Hence, these circumstances elicit the fear emotion and are known as fear releasers. In order to prove the advantages of considering fear in our decision making system, the robot's performance with and without fear are compared and the behaviors are analyzed. The robot's behaviors exhibited in relation to fear are natural, i.e., the same kind of behaviors can be observed on animals. Moreover, they have not been preprogrammed, but learned by real inter actions in the real world. All these ideas have been implemented in a real robot living in a laboratory and interacting with several items and people.

Adaptability of Tacit Learning in Bipedal Locomotion

S. Shimoda, Y. Yoshihara, and H. Kimura

The capability of adapting to unknown environmental situations is one of the most salient features of biological regulations. This capability is ascribed to the learning mechanisms of biological regulatory systems that are totally different from the current artificial machine-learning paradigm. We consider that all computations in biological regulatory systems result from the spatial and temporal integration of simple and homogeneous computational media such as the activities of neurons in brain and protein-protein interactions in intracellular regulations. Adaptation is the outcome of the local activities of the distributed computational media. To investigate the learning mechanism behind this computational scheme, we proposed a learning method that embodies the features of biological systems, termed tacit learning. In this paper, we elaborate this notion further and applied it to bipedal locomotion of a 36DOF humanoid robot in order to discuss the adaptation capability of tacit learning comparing with that of conventional control architectures and that of human beings. Experiments on walking revealed a remarkably high adaptation capability of tacit learning in terms of gait generation, power consumption and robustness.

Reaching for the Unreachable: Reorganization of Reaching With Walking

B. J. Grzyb, L. B. Smith, and A. P. del Pobil

Previous research suggests that reaching and walking behaviors may be linked developmentally as reaching changes at the onset of walking. Here we report new evidence on an apparent loss of the distinction between the reachable and nonreachable distances as children start walking. The experiment compared nonwalkers, walkers with help, and independent walkers in a reaching task to targets at varying distances. Reaching attempts, contact, leaning, and communication behaviors were recorded. Most of the children reached for the unreachable objects the first time it was presented. Nonwalkers, however, reached less on the subsequent trials showing clear adjustment of their reaching decisions with the failures. On the contrary, walkers consistently attempted reaches to targets at unreachable distances. We suggest that these reaching errors may result from inappropriate integration of reaching and locomotor actions, attention control and near/far visual space. We propose a reward-mediated model implemented on a NAO humanoid robot that replicates the main results from our study showing an increase in reaching attempts to nonreachable distances after the onset of walking.

Redundant Neural Vision Systems—Competing for Collision Recognition Roles

S. Yue and F. C. Rind

Ability to detect collisions is vital for future robots that interact with humans in complex visual environments. Lobula giant movement detectors (LGMD) and directional selective neurons (DSNs) are two types of identified neurons found in the visual pathways of insects such as locusts. Recent modeling studies showed that the LGMD or grouped DSNs could each be tuned for collision recognition. In both biological and artificial vision systems, however, which one should play the collision recognition role and the way the two types of specialized visual neurons could be functioning together are not clear. In this modeling study, we compared the competence of the LGMD and the DSNs, and also investigate the cooperation of the two neural vision systems for collision recognition via artificial evolution. We implemented three types of collision recognition neural subsystems - the LGMD, the DSNs and a hybrid system which combines the LGMD and the DSNs subsystems together, in each individual agent. A switch gene determines which of the three redundant neural subsystems plays the collision recognition role. We found that, in both robotics and driving environments, the LGMD was able to build up its ability for collision recognition quickly and robustly therefore reducing the chance of other types of neural networks to play the same role. The results suggest that the LGMD neural network could be the ideal model to be realized in hardware for collision recognition.

Volume 5, Issue 3, September 2013

Microdynamics of Interaction: Capturing and Modeling Infants' Social Learning [Guest Editorial]

Rohlfing, K.J. ; Deak, G.O.

Social learning takes place within an interactional loop. The contributions of this Special Issue exemplify approaches capturing the microdynamics of interaction to provide us with insights into the adaptation and learning processes.

Mothers' Infant-Directed Gaze During Object Demonstration Highlights Action Boundaries and Goals

R. J. Brand, E. Hollenbeck, and J. F. Kominsky

When demonstrating objects to young children, parents use specialized action features, called "motionese," which elicit attention and facilitate imitation. We hypothesized that the timing of mothers' infant-directed eye gaze in such interactions may provide systematic cues to the structure of action. We asked 35 mothers to demonstrate a series of tasks on objects to their 7- and 12-mo-old infants, with three objects affording enabling sequences leading to a salient goal, and three objects affording arbitrary sequences with no goal. We found that mothers' infant-directed gaze was more aligned with action boundary points than expected by chance, and was particularly tightly aligned with the final actions of enabling sequences. For 7- but not 12-mo-olds, mothers spent more time with arbitrary than enabling-sequence objects, and provided especially tight alignment for action initiations relative to completions. These findings suggest that infants may be privy to patterns of information in mothers' gaze which signal action boundaries and particularly highlight action goals, and that these patterns shift based on the age or knowledge state of the learner.

From Action to Interaction: Infant Object Exploration and Mothers' Contingent Responsiveness

C. S. Tamis-LeMonda, Y. Kuchirko, and L. Tafuro

We examined maternal contingent responsiveness to infant object exploration in 190 mother-infant pairs from diverse cultural communities. Dyads were video-recorded during book-sharing and play when infants were 14 mo. Researchers coded the temporal onsets and offsets of infant and mother object exploration and mothers' referential (e.g., "That's a bead") and regulatory (e.g., "Stop it") language. The times when infant or mother were neither exploring objects nor communicating were classified as "off task." Sequential analysis was used to examine whether certain maternal behaviors were more (or less) likely to follow infant object exploration relative to chance, to one another, and to times when infants were off task. Mothers were more likely to explore objects and use referential language in response to infant object exploration than to use regulatory language or be off task, and maternal behaviors were reduced in the context of infants being off task. Additionally, mothers coordinated their object exploration with referential language specifically; thus, mothers' responses to infants were didactic and multimodal. Infant object exploration elicits reciprocal object exploration and informative verbal input from mothers, illustrating the active role infants play in their social experiences.

Young Children's Dialogical Actions: The Beginnings of Purposeful Intersubjectivity

J. Rączaszek-Leonardi, I. Nomikou, and K. J. Rohlfing

Are higher-level cognitive processes the only way that purposefulness can be introduced into the human interaction? In this paper, we provide a microanalysis of early mother-child interactions and argue that the beginnings of joint intentionality can be traced to the practice of embedding the child's actions into culturally shaped episodes. As action becomes coaction, an infant's perception becomes tuned to interaction affordances.

From Language to Motor Gavagai: Unified Imitation Learning of Multiple Linguistic and Nonlinguistic Sensorimotor Skills

T. Cederborg and P.-Y. Oudeyer

We identify a strong structural similarity between the Gavagai problem in language acquisition and the problem of imitation learning of multiple context-dependent sensorimotor skills from human teachers. In both cases, a learner has to resolve concurrently multiple types of ambiguities while learning how to act in response to particular contexts through the observation of a teacher's demonstrations. We argue that computational models of language acquisition and models of motor skill learning by demonstration have so far only considered distinct subsets of these types of ambiguities, leading to the use of distinct families of techniques across two loosely connected research domains. We present a computational model, mixing concepts and techniques from these two domains, involving a simulated robot learner interacting with a human teacher. Proof-of-concept experiments show that: 1) it is possible to consider simultaneously a larger set of ambiguities than considered so far in either domain; and 2) this allows us to model important aspects of language acquisition and motor learning within a single process that does not initially separate what is "linguistic" from what is "nonlinguistic." Rather, the model shows that a general form of imitation learning can allow a learner to discover channels of communication used by an ambiguous teacher, thus addressing a form of abstract Gavagai problem (ambiguity about which observed behavior is "linguistic", and in that case which modality is communicative).

Supporting Early Vocabulary Development: What Sort of Responsiveness Matters?

M. L. McGillion, J. S. Herbert, J. M. Pine, T. Keren-Portnoy, M. M. Vihman, and D. E. Matthews

Maternal responsiveness has been positively related with a range of socioemotional and cognitive outcomes including language. A substantial body of research has explored different aspects of verbal responsiveness. However, perhaps because of the many ways in which it can be operationalized, there is currently a lack of consensus around what type of responsiveness is most helpful for later language development. The present study sought to address this problem by considering both the semantic and temporal dimensions of responsiveness on a single cohort while controlling for level of parental education and the overall amount of communication on the part of both the caregiver and the infant. We found that only utterances that were both semantically appropriate and temporally linked to an infant vocalization were related to infant expressive vocabulary at 18 mo.

SEED Framework of Early Language Development: The Dynamic Coupling of Infant–Caregiver Perceiving and Acting Forms a Continuous Loop during Interaction

P. Zukow-Goldring and N. d. V. Rader

The research and theory described here evolved from fine-grained descriptions of early word learning based on videotapes of infants and their families in the US and Mexico. This naturalistic approach led to theorizing about the perceptual processes underlying the caregiver's role in assisting infants' early word learning. Caregivers educate infants' attention by synchronizing the saying of a word with a dynamic gesture, a show, in which they display the object/referent to the infant. By making this perceptual information prominent, infants can detect an amodal invariant across gesture and speech. Doing so brackets the word and object within the auditory and visual flow of events and constitutes the basis for perceiving them as belonging together. Stemming from the earlier naturalistic work, we designed eye-tracking experiments to test three hypotheses: 1) infants will attend more to an object when the referring word is said if the speaker uses a dynamic, synchronized show gesture, rather than a static or asynchronous gesture; 2) a show gesture will be most effective in drawing attention away from the mouth to the object when the referring word is spoken; and 3) the use of a show gesture will lead to enhanced word learning. These experiments confirmed our hypotheses, establishing that infants detected referent-word relations best when the speaker used a show gesture. These results support the SEED Framework of early language development which delineates how the situated, culturally embodied, emergent, and distributed character of caregiver-infant interaction nurtures communicative behavior. The ability to communicate germinates and takes root during social interaction, as the dynamically-coupled perceiving-and-acting of infants and caregivers forms a continuous loop, each of them unceasingly affecting the other. These findings have implications for the design of cognitive systems in autonomous robots, especially "tutor spotting" and detecting "acoustic packages."

Methodological Considerations For Investigating the Microdynamics of Social Interaction Development

K. de Barbaro, C. M. Johnson, D. Forster, and G. O. Deák

Infants are biologically prepared to learn complex behaviors by interacting in dynamic, responsive social environments. Although the importance of interactive social experiences has long been recognized, current methods for studying complex multimodal interactions are lagging. This paper outlines a systems approach for characterizing fine-grained temporal dynamics of developing social interaction. We provide best practices for capturing, coding, and analyzing interaction activity on multiple -temporal scales, from fractions of seconds (e.g., gaze shifts), to minutes (e.g., coordinated play episodes), to weeks or months (e.g., developmental change).

Editor **Pierre-Yves Oudeyer,** Inria and Ensta ParisTech, France, pierre-yves.oudeyer@inria.fr
Editorial assistant **Fabien Benureau,** Inria and Ensta ParisTech, France
ISSN 1550-1914