IEEE CIS

AMD NEWSLETTER

The Newsletter of the Autonomous Mental Development Technical Committee

Developmental Robotics
Machine Intelligence
Neuroscience
Psychology

Volume 11, number 1
Spring 2014

Editorial: Development as a Complex Adaptive Field of Research

Pierre-Yves Oudeyer

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

This issue of the AMD Newsletter is very special: we are celebrating its 10 years of biannual publication! It has progressively become a place of lively and inspiring scientific dialogues spanning an incredible network of topics, with the contributions of many key actors of computational developmental sciences. It has been for me a fantastic honour to be editor of the newsletter in the last seven years, and I sincerely thank all those who accepted to play the game of dialogues through my solicitations.

These dialogues have vividly explored the role of intrinsic motivation and social interaction in development, the mechanisms of cumulative learning, the processes of perceptual growth and organisation, the issue of language acquisition and symbol grounding, and the application of developmental and social robotics as a tool to address developmental disorders in humans. Several dialogues also discussed the epistemological foundations of modelling development, rethinking why and how the study of robot and human development can feed each other.

Most texts expose arguments that are still valid, and reading them again is a great opportunity to observe at the same time what has been achieved and what remains to be done (see http://www.cse.msu.edu/amdtc/amdnl/). Indeed, developmental processes are probably among the most complicated phenomena under scientific inquiry, and we are just beginning to uncover them. So many great questions are still unanswered. For example, understanding how infants create meaningful and usable abstractions over their multimodal sensorimotor flow, how they build skills upon each other in an open-ended manner, how they come to participate so flexibly in complex interactions with adults without understanding many details, and how to achieve such capabilities with robots, are still largely beyond our reach.

But great progress is being made, and especially in realising that development is a complex multi-scale dynamical system, the study of which requires the deep interconnection between concepts and techniques coming from biology, psychology, neuroscience, physics, computer science and mathematics. The study of developmental systems is itself a complex field of research, which needs to find adaptive strategies to grow beyond the traditional disciplinary walls built within academic institutions. Over the years, the gathering of researchers of very diverse and complementary background within this newsletter, within TAMD, within ICDL-Epirob conference and within multiple workshops, shows the successes of this enterprise. Large-scale European projects and wide-spread research platform such as the ICub across Europe are other examples.

The dialogue in this new issue of the newsletter reflects again the integrated approach needed to understand fundamental cognitive processes like the mutual influences of language and action at both ontogenetic and evolutionary scales. Katerina Pastra launched a discussion on the interaction between the formation of linguistic and conceptual/sensorimotor structures, and in particular on the hypothesis that language as a communication system may have evolved as a byproduct of language as a tool for organising conceptual structures. We can read here the responses of Rick Dale, Katharina Rohlfing and her colleagues, Gary Lupyan, Catriona Silvey, Emmanuel Dupoux, and Kerstin Fisher, each highlighting a different facet of the problem, as well as crucial open questions.

In the years to come, one of our most

important challenge is education. Taking an integrated and interdisciplinary approach requires to handle with dexterity concepts and methods from these diverse fields. How can we grow a community of young researchers mastering the latest advances? How can we teach them to establish cross-disciplinary collaboration and impact? These are the questions raised in the new dialogue initiation proposed by Katharina Rohlfing, Britta Wrede and Gerhard Sagerer, entitled "Trained on everything". Those of you interested in reacting to this dialogue initiation are welcome to submit a response by September 15th, 2014. The length of each response must be between 600 and 800 words, including references (contact pierre-yves.oudeyer@inria.fr).

AMD TC Chair's Message

Matthew Schlesinger

Dept. of Psychology, University of Southern Illinois, USA

AMD TC Chair

matthews@siu.edu

As the 2014 chair of the AMD Technical Committee, I am excited and humbled to serve the international community of researchers that I have both worked with and admired for over a decade. In particular, over the last two years Angelo Cangelosi has provided steady and solid leadership of the TC, and I very much hope to maintain the same spirit of service that he and the past TC chairs have generously offered.

Though already a quarter behind us, the *year of the horse* promises to be a fruitful and productive one for AMD members. First, without a doubt the highlight of 2014 is our next annual ICDL-EpiRob meeting, to be held October 13–16 in Genoa, Italy (general chairs for the event are Giorgio Metta, Mark Lee, and Ian Fasel). Keynote presentations will be provided by four invited speakers: Dana Ballard (UT Austin, USA), Cristina Bechio (Univ of Turin, Italy), Harold Bekkering (Radboud Univ, The Netherlands), and Tetsuro Matsuzawa (Kyoto Univ, Japan). I should also note that the final euCognition meeting will be held immediately after ICDL-EpiRob on October 17-18 at the same location.

Second, other major upcoming events include: Preconference workshop on "Computational Models of Infant Development," held at the International Conference on Infant Studies in Berlin on July 2, from 2–6 pm, organized by Jochen Triesch, Matias Rolf, Minoru Asada, and a little help from me!
Brain-Mind Summer School and International Conference on Brain-Mind, running through June 16 to August 1 in Beijing, and organized by John Weng. The conference is co-located with and will directly follow WCCI-2014.
Planning for the 2015 ICDL-EpiRob meeting has received preliminary approval to be hosted at Brown University, in Providence, Rhode Island (USA), and will be co-organized by myself and Dima Amso.

Looking forward, there are also a number of significant changes taking place. First, Zhengyou Zhang, who has devoted himself tirelessly as editor-in-chief of our flagship journal, TAMD, will be stepping down this year, and a search will be conducted to find his replacement. Second, in 2013 the TC initiated a proposal to update the term our community has adopted, "Autonomous Mental Development," to a new name that not only reflects the diverse questions we study, but that will also increase the reach of our work to related disciplines. More updates on these issues in the next AMD newsletter!

Table of Contents

Dialogue

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.

Language and Cognition: Occasions for Synthesis, but Not Surprise

Rick Dale

Cognitive & Information Sciences, University of California, Merced

rdale@ucmerced.edu

These are challenging and exciting questions at the scientific forefront of not just language, but the organization of human cognition itself. The questions are timely and already Pastra has surveyed intriguing recent results that offer some answers (e.g., Lupyan et al., 2007). The point of my brief commentary is to say that perhaps a broader role for language should not be so surprising, and that there is much work left to do to articulate that role.

Pastra's dialogue initiation starts with language very broadly, but it ends with a focus on labeling. There is no doubt that reference is a central aspect of language. But it is not the only crucial component of language that may affect cognition and social interaction. Language operates across multiple scales and multiple modalities. For example, while Pastra focuses on labeling, most forms of labeling also rely on a range of other important processes, from intricate motor control, to discourse planning in which labeling is used. Language also exploits multiple modalities. When two people interact in spoken language, they deploy vocal, manual, and oculomotor control, along with visual and auditory perceptual processes, all simultaneously. These remarks may seem obvious, but they have implications that may be under-appreciated by our academic focus on the written form (Linell, 2005), which has perhaps led to a strong focus on certain levels of analysis, primarily words and rules (Pinker, 1998).

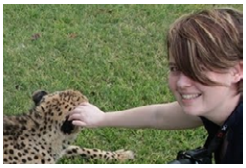
For one, this broader quality leads to Peter Carruthers' provocative discussion of language as a kind of domain-general integration system, "serving to integrate the outputs of a variety of domain-specific conceptual faculties" (Carruthers, 2002, p. 657). But in most theoretical discussions, we often focus on specific loci that some cluster of scholars sees as "true" language. These loci are different depending on one's research program or theoretical leanings. Some propose it to lie in complex manual or vocal motor control, others in sophisticated semantic skills, and still others on some single abstract computational process. However, when looking to what language actually does, in vivo and in situ, the story is considerably more complex, and involves a balancing of a wide range of processes in any bout of linguistic performance. These processes also operate in a surprising lockstep fashion, as if they are operating together as one stable, multidimensional system (Louwerse et al., 2012). Some collaborators and I have argued that this balancing act among components is the harder problem to grasp (Dale et al., 2013).

So language interfaces with many systems, and may systematically map onto and shape processes traditionally seen as part of other systems, from motor control to perception to semantic memory. Only on the backdrop of a historically influential, implausible, and perhaps apocryphal Fodorian modularity might it surprise us that labeling may affect perception (Lupyan & Ward, 2013) and categorization (Lupyan et al., 2007), that fine-grained motor control might relate to phonological awareness (Thomson & Goswami, 2008), that sequential learning might be related to syntax and sentence processing (Misyak, Christiansen, & Tomblin, 2010), that social awareness might assist discourse inferences (Filippova & Astington, 2008), that bilingualism may relate to enhanced forms of cognition including executive control (Bialystok, 2011), and so on. The flow of influence here may be on language or from language, but perhaps just as likely, both. For example, the extended control required to shift registers or languages in bilingualism may both support bilingualism, and through that support, enhance that control. It may be a two-way street.

Given recent work on neural organization, all of this should not surprise us. By all accounts, full-blown human language is a quite recent behavioral innovation in our primate lineage. If it is a new behavioral innovation, then by evolutionary logic language would likely be built upon and reuse (and thus also influence) many other processes that preceded it. Indeed, when large databases of brain activity from language tasks are analyzed, researchers have found that language is one of the most broadly impactful activities, widely recruiting a network of brain areas (Anderson, 2008). Beginning even just with speech perception, the best theories of this process invite a range of brain areas related to motor control, perception, and multisensory integration, functioning across a range of time scales (Giraud & Poeppel, 2012). So we should not be surprised by a deep connection between language and cognition. Perhaps historical contingencies of overly stark theorization lead to a kind of cultural surprise. This commentary and mini-review identified some connections between language and cognition. In the end, I agree wholeheartedly with Pastra that "[a]ddressing such questions can shed new light on language analysis itself, as well as on the development of cognitive, artificial agents." Let's move beyond the debate and achieve this synthesis.

Anderson, M. L. (2008a) Circuit sharing and the implementation of intelligent systems. Connection Science, 20(4):239–51
Bialystok, E. (2011). Reshaping the mind: the benefits of bilingualism. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 65(4), 229.
Carruthers, P. (2002). The cognitive functions of language. Behavioral and Brain Sciences, 25(06), 657-674.
Dale, R., Fusaroli, R., Duran, N., & Richardson, D. (2013). The self-organization of human interaction. Psychology of Learning and Motivation: Advances in Research and Theory, 59, 43-95.
Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. Child Development, 79(1), 126-138.
Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. Nature Neuroscience, 15(4), 511-517.
Linell, P. (2005). Written language bias in linguistics. Routledge.
Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. Cognitive Science, 36(8), 1404-1426.
Lupyan, G.; Rakison, D. & McClelland, J. (2007). Language is not just for talking: Redundant labels facilitate Learning of novel Categories. Psychological Science, 18, 1077-1083.
Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. Proceedings of the National Academy of Sciences, 110(35), 14196-14201.
Misyak, J. B., Christiansen, M. H., & Bruce Tomblin, J. (2010). Sequential expectations: The role of prediction-based learning in language. Topics in Cognitive Science, 2(1), 138-153.
Pinker, S. (1999). Words and rules: The ingredients of language. Basic Books.
Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. Journal of Physiology-Paris, 102(1), 120-129.

Katharina J. Rohlfing
Universität Bielefeld
CITEC,
Bielefeld, Germany
kjr@uni-bielefeld.de



Katrin S. Lohan
Computer Science
Heriot-Watt University,
Edinburgh
k.lohan@hw.ac.uk

Language Does Not Only Enrich — It Meddles with Action Understanding

Katerina Pastra is reviewing many developmental studies with infants suggesting that language is playing a role in cognitive processes. In these studies, however, language is viewed as a spotlight or an inducer (Wolff & Homes, 2010).
A different perspective occurs when we view language not as augmenting ongoing processes but as meddling with them.

In a collaborative study (Sciutti et al., 2013), we have evaluated whether language in form of a verbal narrative concurrent with action observation can direct children's attention to the process of action execution rather than to its goal. Naturally, children select a goal of an action to attend to (Baldwin, Baird, Saylor, & Clark, 2001; Csibra & Gergely, 2007; Woodward, 1999). In our study, using a Tobii eye tracker, we analyzed gaze behaviour of a group of thirty-two 14 month-old infants, who were presented with movies, in which an actor reached for three balls on a table to transport them into a box. We considered three different conditions: the Base condition, in which no audio was associated to the movie; the Sound condition, in which a sinewave sound was added in correspondence of the reaching movements; and the Speech condition, in which a woman voice saying "Jag tar bollen" [I take the ball] emphasized the reaching movements. The results of our experiments show that at 14 months of age the presence of a narrative signal underlining the movement phase, rather than the action goal, reduced substantially the anticipatory shift of infants' gaze to the goal exhibited in other conditions. Our findings thus imply that acoustic packaging can be a meddler in the proactive behaviour typical of the observation of goal oriented actions and determine which of the action elements needs to receive more attention.

Benjamin Koch
Dept. of Psychology,
Uppsala University,
Sweden
benjamin.koch@psyk.uu.se

Alessandra Sciutti
Cognitive Humanoids Lab
IIT, Genova, Italy
alessandra.sciutti@iit.it

The fact that a concurrent narration can influence infants' perception of action units within a continuous action stream has been already demonstrated for 9.5-month-olds (see Brand & Tapscott, 2007). However, our findings extend the current evidence showing that in addition to an effect on action segmentation, verbal behaviour can meddle with non-linguistic action processing, even downgrading the importance of a particular chunk of the action (the goal), which is commonly considered as more relevant by the child (Meltzoff, 1995; Woodward, 1998). Our findings thus imply that language can determine which of the action elements needs to receive more attention — a process that might be basic for further more specific work of top-down attention as it is postulated within Cognitive Grammar (Langacker, 2007).

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A.(2001). Infants parse dynamic action. Child Development 72(3), 708-717.
Brand, R. J., & Tapscott, S. (2007). Acoustic packaging of action sequences by infants. Infancy 11(3), 321-332.
Csibra, G., & Gergely, G. (2007). 'Obsessed with goals': functions and mechanisms of teleological interpretation of actions in humans. Acta Psychol (Amst) 124(1), 60-78.
Langacker, R. W. (2007): Cognitive Grammar. In D. Geeraerts & H. Cuyckens (Eds.), Handbook of Cognitive Linguistics (pp. 1266–1294), Oxford: OUP.
Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. Developmental Psychology 31(5), 1-16.
Sciutti, A., Rohlfing, K. J., Lohan, K. S., Koch, B. & Gredebäck, G. (2013): Language modulate infants' perception of object manipulation task. Poster presented at the Budabest CEU Conference for Cognitive Development (BCCCD 2013), Budapest, 10.–12. Januar.
Wolff, P., & Holmes, K. J. (2010). Linguistic relativity. Wiley Interdisciplinary Reviews: Cognitive Science 2, 253-265.
Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. Cognition 69(1), 1-34.
Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. Infant Behavior and Development 22, 145-160.

Language Augments Cognition and Perception by Providing High-Level Hypotheses

Gary Lupyan

Department of Psychology,
University of Wisconsin-Madison,
USA
lupyan@wisc.edu

The role of language on cognition has been a subject of intense interest for centuries but only recently has it become the focus of directed empirical research in the cognitive sciences. As hinted at by Pastra's excellent opening, there is now considerable evidence that far more than being a highly flexible and efficient system for communicating our thoughts, language may be intimately involved in shaping those very thoughts, involved not only in the process of conceptual development in childhood, but continuing to augment ongoing cognitive/perceptual processing in adulthood. Normal human cognition may, to an important degree, turn out to be <i>language-augmented cognition</i> (Lupyan, 2012a, 2012b for reviews).

Why might language play such an important role? Consider vision. Because there are many potential sources for any pattern of retinal stimulation, the visual system has to make inferences as to the source—inferences that can be "tested" by making predictions about e.g., how a percept should change in time and space. Within this broad predictive framework, words and larger constructions can be thought as a source of high level hypotheses. Indeed, in a recent study, we have shown that simply hearing a word can boost an otherwise invisible image into awareness (Lupyan & Ward, 2013). The word provides a high-level hypothesis within which weak and fragmented visual input can be reassembled.

The power of language becomes clearer still when we consider that language acts directly on mental states. A word like "red" activates neural representations that overlap with those evoked by actual experiences of redness. But while experiences of redness are always specific ones, the word "red" is a <i>categorical</i> cue. Thus, the phrase "tomatoes are red" abstracts over the details of tomatoes and redness irrelevant for the proposition, allowing for the activation of a mental state that would not be instantiated by sensorimotor experiences alone because they are always grounded by concrete instances.
Indeed, when we test how verbal cues (e.g., "dog") compare to unambiguous nonverbal cues (e.g., barking sound) in activating visual knowledge, we find that words activate more categorical states while other cues activate more specific instances (Lupyan & Thompson-Schill, 2012; Edmiston & Lupyan, 2013). In short, language not only captures the joints of nature, but carves joints in nature, allowing for efficient activation of categorical states.

The view that language augments cognition is not shared by all within the cognitive science community. Researchers working within a nativist/modularist tradition do not concede that experience in learning and using language contributes to the development of, e.g., abstract compositional thought, symbolic capacities, and ability for complex relational reasoning—which they view to be pre-requisites to language rather than the consequence of language (e.g., Gleitman & Papafragou, 2005). Given the impossibility of randomly assigning humans to linguistic and non-linguistic worlds, these debates are difficult to settle conclusively through conventional experimental manipulations. Studies of non-human animals trained with symbolic systems offer a source of provocative, but often overlooked evidence in favor of the causal role of language in cognition: Animals trained with symbols—the very sort that language provides human learners—become capable of certain operations such as simple relational reasoning that appear out of reach for their untrained brethren (Thompson, Oden, & Boysen, 1997). Of course this work is expensive, difficult to carry out, and often open to alternative interpretations. The possibility of using robotic agents learning to negotiate actual sensorimotor contingencies and augmenting this process by providing various kinds of linguistic labels strikes me as hugely attractive. Language overlays on top of the perceptual world a rich hierarchical system of categories, many of which are not derivable from the sensorimotor information alone. The process of constraining and manipulating lower-level representations using higher-level language operators may allow for representational capacities and behaviors that may prove impossible in the absence of language learning and use.

Edmiston, P., & Lupyan, G. (2013). Verbal and Nonverbal Cues Activate Concepts Differently, at Different Times. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), Proceedings of the 35th Annual Conference of the Cognitive Science Society (pp. 2243–2248). Austin, TX: Cognitive Science Society.
Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & B. Morrison (Eds.), Cambridge Handbook of thinking and Reasoning (pp. 633–661). Cambridge: Cambridge University Press.
Lupyan, G. (2012a). Linguistically modulated perception and cognition: the label-feedback hypothesis. Frontiers in Cognition, 3(54). doi:10.3389/fpsyg.2012.00054
Lupyan, G. (2012b). What do words do? Towards a theory of language-augmented thought. In B. H. Ross (Ed.), The Psychology of Learning and Motivation (Vol. 57, pp. 255–297). Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780123942937000078
Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. Journal of Experimental Psychology-General, 141(1), 170–186. doi:10.1037/a0024904
Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. Proceedings of the National Academy of Sciences, 110(35), 14196–14201. doi:10.1073/pnas.1303312110
Thompson, R. K., Oden, D. L., & Boysen, S. T. (1997). Language-naive chimpanzees (Pan troglodytes) judge relations between relations in a conceptual matching-to-sample task. Journal of Experimental Psychology. Animal Behavior Processes, 23(1), 31–43.

Functions of Language: Communication Is Basic, Categorisation Is Derived

Catriona Silvey

Language Evolution and Computation Research Unit, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, UK

lupyan@wisc.edu

Katerina Pastra presents an interesting hypothesis: the evolutionarily basic function of language is not communication, but imposing distinctiveness on sensorimotor experience. This suggestion is initially persuasive; previous work has shown that labels assist in category learning (Lupyan, 2007) and may indeed trigger the formation of categories in the first place (Waxman & Markow, 1995). Additionally, evidence suggests that some more abstract categories may only be learnable via language (Gentner et al., 2013; Ozçalışkan et al., 2009). However, this does not entail that the primary function of language is discretising sensorimotor input. Firstly, non-linguistic animals such as higher apes have rich conceptual repertoires (see Hurford, 2007 for review); furthermore, children who lack linguistic input, for example deaf children raised by non-signing parents, are still able to form categories at a high level of abstraction, such as generic kinds (Goldin-Meadow et al., 2005), despite being given no labels to assist them in acquiring these categories. This shows that linguistic input is unnecessary for the formation of structured categories, casting doubt on the hypothesis that this is language's basic function.

Secondly, the idea that words label sensorimotor experiences is a radical simplification of how language works. Many words do not correspond to regions of sensorimotor input, but rather serve discourse functions inseparable from communication: for example, deictic terms, morphemes encoding evidentiality, articles marking novelty or familiarity of referents, question markers, and active/passive distinctions. All of these 'serve to organize the flow of information in connected discourse' (Slobin, 1996), rather than referring directly to individuals' sensorimotor experience. It could be argued that these discourse-organising words emerged only after language was co-opted for communication. However, even seemingly basic object nouns and action verbs are polysemous, with the precise categories they delineate highly dependent on the communicative context of the utterance (Klein & Murphy, 2002). Experiments that show category learning advantages for labels abstract away from this real feature of language, giving a misleadingly simple picture of the linkage between label and category.

Thirdly, the hypothesis poses a fundamental chicken-and-egg problem for how language evolved. If the basic function of language is to aid learners in categorisation, and language plays a causal role in the acquisition of these categories, how would the first language users have known to apply their labels appropriately? It seems more reasonable to posit the categories labelled by words as initially the product of a pressure for communication. Once established and repeatedly learned by subsequent generations, these labels in turn affect individual cognition, giving rise to the category learning effects Pastra cites. Simply put, words help us to learn and form categories because communication encourages the conventionalisation of labels for referring to aspects of our sensorimotor experience (Voiklis & Corter, 2012). Of course, it remains unlikely that every feature of language can be explained by an appeal to communicative function: language must also be learned, and learning biases shape the structure of language over cultural transmission (Brighton et al., 2005). However, since this learning takes place within the arena of use, category learning advantages from language are likely to be derivative of a more basic communicative function, rather than the other way around.

Brighton, H., Kirby, S., & Smith, K. (2005). Cultural selection for learnability : Three principles underlying the view that language adapts to be learnable. In M. Tallerman (Ed.), Language Origins: Perspectives on evolution. Oxford: Oxford University Press.
Gentner, D., Ozyürek, A., Gürcanli, O., & Goldin-Meadow, S. (2013). Spatial language facilitates spatial cognition: evidence from children who lack language input. Cognition, 127(3), 318–30.
Goldin-Meadow, S., Gelman, S. A, & Mylander, C. (2005). Expressing generic concepts with and without a language model. Cognition, 96(2), 109–26.
Hurford, J. (2007). The origins of meaning: Language in the light of evolution. Oxford: Oxford University Press.
Klein, D. E., & Murphy, G. L. (2002). Paper has been my ruin: Conceptual relations of polysemous senses. Journal of Memory and Language, 47, 548–570.
Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. Psychological Science, 18(12), 1077–1083.
Özçalışkan, S., Goldin-Meadow, S., Gentner, D., & Mylander, C. (2009). Does language about similarity play a role in fostering similarity comparison in children? Cognition, 112(2), 217–28.
Slobin, D. I. (1996). From “Thought and language” to “Thinking for speaking.” In J. J. Gumperz & S. C. Levinson (Eds.),Rethinking linguistic relativity (pp. 70–96). Cambridge: Cambridge University Press.
Voiklis, J., & Corter, J. E. (2012). Conventional wisdom: Negotiating conventions of reference enhances category learning. Cognitive Science, 36(4), 607–634.
Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. Cognitive Psychology, 29, 257–302.

The Many Functions of Linguistic Labels

Kerstin Fischer

University of Southern Denmark

fischer@informatik.uni-hamburg.de

Katerina Pastra argues that linguistic labels help the developing human to form categories and thus that they fulfill important cognitive functions. The question she raises on that basis is then, if language has these central cognitive functions, are its communicative functions possibly a by-product?

The fact that linguistic categories have cognitive functions does not necessarily mean that cognitive functions are prior, though. If language contributes to making sense of the world by allowing the learner to classify sensorimotor experience as instances of a more general category and thus to connect current data with knowledge about similar instances, then this is as much a social as a cognitive achievement. Moreover, the world to make sense of for a developing human is a social world, and the categories developed are socially relevant – which is crucial for developing humans since they cannot survive on their own. This social nature of categories becomes apparent in the fact that categories are language-specific. For instance, languages differ concerning the color distinctions they make (e.g. Kay et al. 2009), and also objects and events are categorized differently by different languages (e.g. Lakoff 1987, Croft 2001). For example, English distinguishes between *swim* and *float*, whereas in German no distinction is made between self-propelled, active swimming and passive floating. The categories languages encode are thus specific to the culturally relevant distinctions of the respective social group (cf., for instance, Pullum (1989) on Eskimo vocabulary for different types of snow).

In child language acquisition, linguistic labels furthermore serve purposes other than cognitive. For instance, they are used by caregivers to provide the child access to segmentation and initial sense making. Caregivers often use linguistic labels isolatedly in child-directed speech, where they help the child to identify the units of the language and thus to segment the speech stream (cf. Brent & Siskind 2001). Here, labels do not have cognitive functions apart from enhancing understanding well enough to serve as an access point into linguistic structure.

Labels may also function as social starting points for joint attention. Initially, with young infants, caregivers have been found not to talk about objects much and instead to focus on issues related to the direct relationship between caregiver and infant (e.g. Sylvester-Bradley and Trevarthen 1978). After a few months, caregivers begin to initiate triadic relationships in which the child's attention is drawn to an object, which is then in the joint focus of infant and caregiver. In this phase, the child may also practice establishing joint attention by pointing to or demonstrating an object for the caregiver to attend to (cf. Filipi 2009). Thus, the labeling of an object may serve merely social functions, namely establishing joint attention, which is independent of cognitive categorization tasks.

When we look beyond linguistic labels, we furthermore need to acknowledge that the social functions of language reach far beyond information transmission, and that the distinction between the cognitive and the social cannot be easily upheld. Already Bühler (1934) assumes appellative and expressive functions together with representative, and Jakobson (1960) extends this list by meta-linguistic and poetic functions of language. In a cross-linguistic study of speech to 12-months-olds, for instance, Fernald (1992) finds distinct intonation contours for the communicative functions approval, prohibition, comforting and attention getting in all languages she investigates. Considering these sets of functions, it becomes obvious that language plays many roles and involves many facets of human existence.

We can conclude that linguistic labels themselves are the result and reflection of social processes, and that the act of labeling serves many other, social functions apart from the cognitive functions. Labels are consequently only foregrounded in some phases of the child's development, and they may be more related to social learning than to categorization. Labels consequently do introduce a cognitive structure into the world, as Katerina Pastra convincingly argues, but this world is a social world in which language plays multiple roles.

Brent, M. R. and J. M. Siskind (2001). The role of exposure to isolated words in early vocabulary development. Cognition 81, B33–B44.
Bühler, K. (1934/1999). Sprachtheorie: Die Darstellungsfunktion der Sprache. 3rd edition, Stuttgart: Fischer, pp. 24–33.
Croft, W. (2001). Radical Construction Grammar. Oxford University Press.
Fernald, A. (1992). Maternal vocalisations to infants as biologically relevant signals: An evolutionary perspective. In J.H. Barkow, L. Cosmides, & J. Tooby (Eds.), The adapted mind: Evolutionary psychology and the generation of culture. Oxford: Oxford University Press.
Filipi, A. (2009). Toddler and Parent Interaction. Amsterdam/Philadelphia: John Benjamins.
Jakobson, R. (1960). Linguistics and Poetics: Closing Statement. In: Sebeok, T. (ed.): Style in Language, The MIT Press & John Wiley.
Kay, P., Berlin, B., Maffi, L., Merrifield, W.R. and Cook, R. (2009). The World Color Survey. Stanford: Center for the Study of Language and Information.
Lakoff, G. (1987). Women, Fire and Dangerous Things. What Categories Reveal about the Mind. The University of Chicago Press.
Pullum, Geoffrey K. (1989). The great Eskimo vocabulary hoax. Natural Language and Linguistic Theory 7, 275-281.
Sylvester-Bradley, B. and C. Trevarthen (1978). Baby talk as an adaptation to the infant's communication. In N. Waterson and C. E. Snow (Eds.), The Development of Communication, pp. 75–92. Chichester, New York, Brisbane, Toronto: JohnWiley and Sons.

Towards Quantitative Studies of Early Cognitive Development

Emmanuel Dupoux

Ecole des Hautes Etudes en Sciences Sociales, France

emmanuel.dupoux@gmail.com

Katerina Pastra asks fascinating questions regarding the relationship between language and the visuo-motor system during development. Among them, whether language could *affect* or *facilitate* motor exploration and whether verbal labels could have a *causal role* on sensorimotor experiences beyond the role already documented in perceptual category formation. Questions regarding causal roles of various factors during development are notoriously difficult to address (Morton, & Frith, 1995). One reason is that it is impossible to use the gold standard of scientific investigation, i.e., experimental manipulation. One could not, for obvious ethical reasons, deprive a child from part of language or sensorimotor experience in order to test hypotheses about causal relationships. Experimental micro-interventions using miniature artificial languages learning (Saffran et al., 1996) or sensory-motor learning (Sommerville et al., 2005) only allow to identify *possible* causal factors, but one never knows whether such factors actually play a role in practice. Correlational studies allow to *quantify* the relationships between language and sensorimotor capacities, capitalizing on the natural variations that spontaneously arise in the "ecological" environment of a child. Yet, in order to fully address Pastra's questions with this approach, three major breakthroughs are needed.

1. The availability of dense, labeled databases of infant development. Infant research has made major advances thanks to the availability of open source transcribed parent-infant interactions (the CHILDES database, MacWinney, 2000). However, the available databases seldom contain information beyond verbal transcriptions. They typically lack the audio and video data which would be necessary to establish correlations between language and motor development. In addition, they only sample the infant's environment and behavior in a very sparse fashion (a few hours or a few days). The availability of cheap audio and video sensors makes it possible to envision the capture of every instant of a baby's life. Such dense databases would require a

monumental annotation effort, and raise privacy issues in relation to requirement of open data access. The Speechome project (dkroy.media.mit.edu, Roy et al., 2012) is a pioneering effort in that direction (see also studies using the Lena devices: www.lenafoundation.org).

2. The availability of longitudinal experimental data. Much of infants' early acquisitions are not directly observable in their behavior. Experimental methods reveal that during the first year of life (before they can talk), infants develop sophisticated representations of the physics of inanimate objects (spatio-temporal permanence, solidity, gravitation and support), assign hidden states to animated entities (goals, beliefs, emotions) and acquire the basic structure of language (phoneme categories, elements of syntax and semantics). Just relying on observable behaviors would result in serious underestimation of infant's competences (see Carey, 2009; Werker et al 2012). Yet, our knowledge of this period is still fragmentary: most studies measure a single dependant variable, are too noisy for individual measurements, and are conducted in too few languages and cultures. In order to understand the *causal* relationships between language, cognitive and motor development, it is necessary to measure variations in individual infant's cognitive development using longitudinal multivariate studies, and across cultures. Significant methodological developments are necessary in order to increase the resolution and portability of preverbal cognitive measurements in order to achieve this goal. See Bornstein et al (2013), Aslin (2012) and the rise of meta-analyses repositories (sites.google.com/site/InVarInf) for efforts in this direction.

3. The availability of computational models at scale. Much of what we know about the *mechanisms* underlying early cognitive development is speculative. Computational/modeling approaches often simplify the learning problem and use toy corpora. For instance, the development of phonetic categories has been modeled using non-supervised clustering of acoustic data. However, most studies have simplified the input using synthetic data and reduced the phoneme inventory to a few segments (e.g. Vallabha et al., 2007). Confronted with real size corpora, however, these algorithms fail to scale up and a completely different architecture is needed (e.g. Martin et al, 2013). While toy problems are useful to establish proof of principles, it is not clear what they teach us about realistic learning situations, such as the ones faced by infants. Even though recent advances in machine learning make it possible to address more realistic problems, we are still far from implementing a complete end-to-end system that would learn cognitive skills in a naturalistic environment (see cocosci.mit.edu and syntheticlearner.net).

In order to escape sparse experimental data and vague theorization, an outstanding collaboration effort is needed between engineering and developmental psychology in order to meet the three above-mentioned challenges. When this happens, it will be possible to give quantitative answers to Pastra's questions.

Aslin, R.N. (2012). Infant eyes: A window on cognitive development. Infancy, 17(1), 126-140.
Bornstein, M. H., Hahn, C.-S. and Wolke, D. (2013), Systems and Cascades in Cognitive Development and Academic Achievement. Child Development, 84: 154–162
Carey, S. (2009). The origin of concepts. Oxford University Press.
MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
Martin, A., Peperkamp, S. & Dupoux, E. (2013). Learning Phonemes with a Proto-lexicon. Cognitive Science, 37, 103-124.
Morton, J., & Frith, U. (1995). Causal modeling: a structural approach to developmental psychopathology. Manual of developmental psychopathology, 1, 357-390.
Roy, B.C., Frank, M.C., and Roy, D. (2012). Relating Activity Contexts to Early Word Learning in Dense Longitudinal Data. Proceedings of the Cognitive Science Society. Sapporo, Japan.
Saffran, J.R., Aslin, R.N. & Newport, E.L. (1996) Statistical learning by 8-month-old infants. Science 274, 1926 – 1928.
Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. Cognition, 96, B1-B11.
Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. & Amano, S. (2007) Unsupervised learning of vowel categories from infant-directed speech, PNAS, 104:33.
Werker, J.F., Yeung, H.H., & Yoshida, K. (2012). How do infants become native speech perception experts? Current Directions in Psychological Science, 21(4), 221-226.

Beware... of the Label

Katerina Pastra

It is with great pleasure that I received the replies of a number of esteemed colleagues on thought provoking questions and points regarding the role of language in sensorimotor experience acquisition. Of course, this is not a new question at all. For example, some aspects of it have been largely addressed in the long-standing debate on language evolution and what it is that makes human language unique (cf. for example, Hauser et al. 2002, Fitch et al. 2005, Chomsky 2005, Pinker and Jackendoff 2005).

In this dialogue, we focus more on the developmental perspective of this question and in particular, on the impact of language over the

Katerina Pastra

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

kpastra@csri.gr

primary cognitive process of categorization: *Why do we verbally communicate the categories into which our sensorimotor experiences fall?* In what way do humans in their very early months of life benefit – if they do – from verbal interaction during sensorimotor exploration of the world?

There is a growing literature on verbal (and non-verbal) social interaction and its significance in sensorimotor experience acquisition (cf. for example, Rohlfing and Tani 2011). The focus in this strand of research is mainly on *how such* interaction takes place (i.e. *the cognitive mechanisms that enable such interaction, such as shared attention*). Thus, along these lines of research, Rohlifing et al. bring to this discussion findings from experimental literature that point to language as a mechanism that modulates attention, which may affect – for example – visual action segmentation. They report on experimental research in which language is shown to modulate infants' attention to specific – non-otherwise similarly attended – elements of a visually observed action. This supports what Dale also emphasizes: that language has a multifaceted role in cognition, which goes beyond categorization to processes such as attention, segmentation, memory encoding and retrieval and other cognitive processes. Lupyan builds on top of a similar line of thinking and adds also the role of language as a prior-knowledge system that is used for prediction.

However, research on the mapping of natural language to sensorimotor experiences misses this perspective of the dynamic, multifaceted role of language in human cognition. The philosophical debate on symbol grounding and the related computational modeling research, has primarily focussed on what the language system gains from such integration, overlooking the fact that integration is a bidirectional process in which all involved systems contribute (cf. criticism and the notion of Double Grounding in Pastra 2004). Verbal labels have been largely treated as monolithic indexes in this process; their dynamic nature has been completely overlooked.

Both Silvey and Fischer comment directly on that dynamic nature of labels. Silvey picks up an underlying negative connotation of the term and argues that words/language *does not just label sensorimotor experiences*. She mainly argues along two directions:

a) *Not all labels correspond to sensorimotor experiences*; Silvey provides examples from verbal units/morphemes/words that do not correspond to sensorimotor experiences, such as functional words (articles, deictics etc.). However, even among such cases, one can find direct links to sensorimotor experiences: deictic words are a case par excellence where language comments on its link to context (either linguistic or sensorimotor) (a deictic word has no content – it is just a pointer to something mentioned previously in discourse or shared in the sensorimotor space by the interlocutors). Abstract words (e.g. poverty), also do not correspond directly to sensorimotor experiences, but they do indirectly (e.g. one may talk about poverty in a region and show shacks, beggars and so on to illustrate aspects of such evaluative category). One may follow the Cognitive Linguistics argumentation in that respect, that such verbal units are still – even indirectly – rooted to sensorimotor experiences.

b) *Labels are polysemous*; however, not in a sensorimotor context with shared attention, not in situated communication. Actually, the linking of verbal units to their denoted entities, movements, perceptual features or abstract referrents (Pastra et al. 2011), can be a robust criterion for deciding on polysemy avoiding false alarms.

Fischer brings to this discussion two more aspects of verbal labels:

c) *Labels are language-specific*, i.e. they are the product of social interaction, and thus verbalize culturally relevant distinctions drawn by different social groups. Indeed, diversity is an important aspect of human cognition, without which progress would be seriously hampered. Language reflects this diversity; however, it also reflects commonalities: commonalities in physiology, environment, and needs that human species share by definition. Evidence that language and the sensorimotor system share common syntactic mechanisms has recently led to the formalisation of a generative grammar of action, providing a common space where language and action syntactico-semantic analysis may meet (Pastra and Aloimonos 2012).

d) *Labels have a social function too*, for example, they are used for establishing shared attention. This is also a correct observation, which takes us back to the multifaceted role of language in cognition, and the modulation of attention.

There is agreement that language expresses intention explicitly, it goes beyond here and now, it can modify the status of facts. However, we argue that there is something more in labels. Going back to language that refers directly to sensorimotor experiences, *labels name sensorimotor experiences. And naming is not indexing.* The label/name *does a lot more than indexing. It's a generator*. It generates new labels. Word formation mechanisms (e.g. compounding and derivation) are key in the linguistic system and hold a large part of its dynamic nature. The label itself, usually incorporates cues that reveal its association (categorical or thematic) to other labels. We believe that this is a missing point in current computational and experimental research; a point that distinguishes natural language labels from other symbolic labels.

As mentioned by Dale, seeing cognition from an embodied and enactive perspective, increasingly supported by findings in neuroscience, it should be of no surprise that language affects cognitive processes and systems and is also affected by them. However, though not surprising, our knowledge of the involved multisensory integration mechanisms and the functionality of such interaction between a symbolic system and the sensorimotor ones is still limited.

The reasons for this may be related to an extent to what Dupoux argued: the methodological difficulties in exploring such basic research questions and the need of both experimental and computational modelling research in naturalistic settings and with longitudinal data involving even prelanguage infants. One could add to this, a need for *informed syntheses* of findings and achievements in the involved disciplines that would point to aspects of the problem for which there is common agreement, unexplored directions, and remaining challenges. Interdisciplinary research reminds me of these integration processes in the human brain that we so much need to explore: language (the linguists' perspective), perception (e.g. the visual processing perspective), motor control (the engineering perspective), learning (the developmental perspective) and so on. Interdisciplinarity is very hard, because it requires integration of perspectives; integration requires understanding; understanding requires knowing; knowing requires interaction; interaction needs some form of dialogue.

In this dialogue, it has been evident that we agree on the multifaceted nature of language and its important contribution in cognitive processes. However, it is also evident that the nature of language when addressed from an embodied and enactive perspective is still largely unexplored. Let's look at labels again, keeping in mind that they are governed by basic, generative mechanisms. This may open new directions in our experiments and models.

Chomsky N. (2005). Three Factors in Language Design. Linguistic Inquiry. 36:1–22.
Fitch W, Hauser M, and N. Chomsky (2005). The evolution of the language faculty: Clarifications and Implications. Cognition. 97:179–210.
Hauser M, Chomsky N, and W. Fitch (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? Science. 298:1569–1579.
Pastra K. (2004). Viewing Vision-Language Integration as a Double-Grounding Case. In Proceedings of the American Association of Artificial Intelligence (AAAI) Fall Symposium on "Achieving Human-Level Intelligence through Integrated Systems and Research", pp. 62-67, Washington D.C. USA.
Pastra K., E. Balta, P. Dimitrakis, G. Karakatsiotis (2011). Embodied Language Processing: a new generation of language technology. In Proceedings of the AAAI 2011 International Workshop on "Language-Action Tools for Cognitive Artificial Agents: Integrating Vision, Action and Language", San Francisco, USA.
Pastra K. and Y. Aloimonos (2012). The Minimalist Grammar of Action. Philosophical Transactions of the Royal Society B, 367(1585):103.
Pinker S, and R. Jackendoff (2005). The Faculty of Language: What's special about it? Cognition. 95:201–236.
Rohlfing K. and J. Tani (2011). Grounding Language in Action. IEEE Transactions on Autonomous Mental Development. Vol. 3:2, pp. 109-112.

New Dialogue Initiation

Trained on Everything

Katharina J. Rohlfing
kjr@uni-bielefeld.de

Britta Wrede
bwrede@techfak.uni-bielefeld.de

Gerhard Sagerer
sagerer@techfak.uni-bielefeld.de

Universität Bielefeld
CITEC,
Bielefeld, Germany

Cognitive Developmental Robotics as a theme has been proposed 13 years ago (Weng 2001), as formulated later in Asada and colleagues (2009). By that time, many researchers had recognized the need to cross disciplinary boundaries in order to push the progress towards developing systems that can learn and act flexibly in physical as well as social environments. Today, our everyday work draws on the input from many different disciplines: at the ICDL-Epirob conference, we welcome contributions from developmental psychology, linguistics, and neuroscience in addition to developmental robotics; conversely, more and more symposia and workshops on modeling learning and development are organized within the SRCD, Infancy or IASCL conferences. Thanks to this trans- and interdisciplinary research, we can approach complex phenomena. Consider, for example, the role of contingency in language acquisition. Developmental studies have shown that contingent interaction is important for infant development. For example, infants prefer contingent face movements to still faces. Furthermore, contingency plays an important role in learning as an ostensive cue as it signals to the infant that (1) interaction is going on and, even more, (2) a teaching situation is taking place (Csibra, 2010). These different contingent features could be operationalized on a robot to model an interaction that achieves a new interactive quality (although no understanding on the side of the robot takes place) (Lohan et al., 2012). Such rich interactive capability can now be applied in a real teaching and learning scenario leading to new questions, namely, if there are different levels of contingency, e.g. online feedback signals, that are applied immediately in an interaction when something is happening (going wrong, or right, etc.), and how these signals can be used to enhance the underlying learning model in an incremental and online fashion.

This example illustrates that a topic of investigation – such as contingency – can and should not be treated as an isolated phenomenon that can be modeled in a modular fashion. Rather, it is embedded in a complex developmental and interactional process as it interacts with other phenomena (e.g. learning or acting) and triggers specific forms of interaction. Many of us certainly enjoy such a comprehensive scientific view.

But do our students enjoy it as well? Or are they instead "lost in the complexity" of the topic? A non-trivial question for our community is therefore how we can pass our knowledge on to our students, so that they become interdisciplinary thinkers able to formulate questions about complex systems?

Without providing the perfect formula, we would like to discuss two options which could be the good ways to provide interdisciplinary training for students.

Option 1: Interdisciplinarity at the PhD-level
There is certainly a non-exhaustive list (s. Figure 1) that students have to check during their PhD-period. This workload forces a successful student to be very focused because, in Europe, this list needs to be accomplished within three or four years. We do not want students to focus too much on specific disciplinary methods. We would like to train our students to not only formulate questions about complex systems, but also to appreciate the methods with which other disciplines approach relevant and exciting questions.

- [x] conduct studies and research
- [ ] write dissertation
- [ ] publish work in journals and proceedings
- [x] attend conferences
- [ ] work abroad
- [ ] attend courses at home
- [x] make first experience with teaching
- [ ] socialize with key persons from your field

Figure 1

To become an interdisciplinary thinker, a student has to learn about the topic and the methods. Thus, the complexity of the chosen topic might be huge at the beginning, but the reward could be a comprehensive contribution. Certainly, students starting with this load need to talk to people from different fields. Optimally, they will also be supervised by persons who understand the problems of "getting lost in complexity". Such solutions are implemented at Bielefeld University within CITEC Graduate School. Weekly student meetings and regular retreats allow the students to get to know and exchange different perspectives. An interdisciplinary dialogue is also practiced at the level of Master's students, who attend classes taught by two teachers from different fields. Such classes benefit from lively discussions.

Option 2: Interdisciplinarity at the postdoctoral-level
Another option for students is to not give them the impression to be trained on everything, but to provide a solid education in one field, focusing on very specific methods. After finishing their dissertation, a postdoctoral project can

be targeted, in which students could focus on a different field without legacy from the PhD-period. One would expect that a postdoctoral student can be more resilient to getting lost in complexity.

Time constraints are essential for both solutions. The scientific experiences that the students will make within a few years, while still having time to process everything, is limited. As supervisors we prioritize experiences that the students should make, but our judgments are guided by impact and success. However, we think that if we want to educate an interdisciplinary community, we should prioritize the dialogue with people from other fields and allow room for students to speak up and develop novel ideas.

Dialogue

Talking to each other is also a developmental process.

The first time that e.g. a linguist by training speaks to a "guy from Computer Science" can be peculiar. First, there is the matter of terminology, which differs from discipline to discipline. Interestingly, a conversation can be even more difficult if e.g. a psychologist and a linguist talk about "social interaction" because they assign different phenomena to this key word.

Second, there is the matter of the complexity of the topic that one would like to convey, but that the other would not necessarily like to hear. For example, while a linguist may be interested in how children learn to use words flexibly, a computer scientist might be more interested in how a robot can show rapid learning capabilities.

Third, a successful dialogue will rely on bi-directional appreciation of the scientific methods. Ethnographic studies – qualitative in their nature – can open up new exciting questions, which can then be followed up quantitatively and eventually result in capabilities of an implemented system (Pitsch et al., 2014).

Fourth, and related to the bi-directional appreciation, there is the matter of constructive thinking. Any interdisciplinary dialogue is a construction of a novel topic and one needs time to actually work on it.

The education of an interdisciplinary researcher needs to foster mediator capabilities. These capabilities will enable the researcher to constructively find the relations between methods from different disciplines and synthesize the insights into a new structure. In this new structure or map, new research questions will arise.

Room for dialogue

We should think of room where such dialogues (and dialogical skills) can be developed. It is difficult to create such an exchange in a virtual environment (internet) because it seems that every student accesses this field in an individual way. One possibility would be to foster small projects that students could work on in tandem. On the one hand, they can reflect upon the applied methods from one or the other field. On the other hand, this reflection should result in critical awareness and assertiveness about one's own methods, i.e. its potential and limitations.

Interdisciplinary mentoring would be another possibility to broaden students' perspectives. Maybe we can think of giving such exchanges room in the context of the ICDL-Epirob conference, where students can offer topics to exchange methods and ideas.

J. Weng and J. McClelland and A. Pentland and O. Sporns and I. Stockman and M. Sur and E. Thelen. (2001) Autonomous Mental Development by Robots and Animals, Science, 291 (599-600)

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M. & Yoshida, C. (2009), Cognitive developmental robotics: a survey. In: IEEE Transactions on Autonomous Mental Development, 1: 12–34.

Csibra, G. (2010), Recognizing communicative intentions in infancy. In: Mind & Language 25: 141–168.

Lohan, K. S., Rohlfing, K. J., Pitsch, K., Saunders, J., Lehmann, H., Nehaniv, C. L., Fischer, K. & Wrede, B. (2012), Tutor Spotter: Proposing a feature set and evaluating it in a robotic system. In: International Journal of Social Robotics 4: 131–146.

Pitsch, K., Vollmer, A.-L., Rohlfing, K. J., Fritsch, J. & Wrede, B. (2014), Tutoring in adult-child interaction. On the loop of the tutor's action modification and the recipient's gaze. In: Interaction Studies 15: 55–98.

IEEE TAMD Table of Contents

Volume 5, Issue 4, December 2013

Computational Audiovisual Scene Analysis in Online Adaptation of Audio-Motor Maps

Rujiao Yan, T. Rodemann, B. Wrede

For sound localization, the binaural auditory system of a robot needs audio-motor maps, which represent the relationship between certain audio features and the position of the sound source. This mapping is normally learned during an offline calibration in controlled environments, but we show that using computational audiovisual scene analysis (CAVSA), it can be adapted online in free interaction with a number of a priori unknown speakers. CAVSA enables a robot to understand dynamic dialog scenarios, such as the number and position of speakers, as well as who is the current speaker. Our system does not require specific robot motions and thus can work during other tasks. The performance of online-adapted maps is continuously monitored by computing the difference between online-adapted and offline-calibrated maps and also comparing sound localization results with ground truth data (if available). We show that our approach is more robust in multiperson scenarios than the state of the art in terms of learning progress. We also show that our system is able to bootstrap with a randomized audio-motor map and adapt to hardware modifications that induce a change in audio-motor maps.

Modeling Cross-Modal Interactions in Early Word Learning

N. Althaus, D. Mareschal

Infancy research demonstrating a facilitation of visual category formation in the presence of verbal labels suggests that infants' object categories and words develop interactively. This contrasts with the notion that words are simply mapped "onto" previously existing categories. To investigate the computational foundations of a system in which word and object categories develop simultaneously and in an interactive fashion, we present a model of word learning based on interacting self-organizing maps that represent the auditory and visual modalities, respectively. While other models of lexical development have employed similar dual-map architectures, our model uses active Hebbian connections to propagate activation between the visual and auditory maps during learning. Our results show that categorical perception emerges from these early audio-visual interactions in both domains. We argue that the learning mechanism introduced in our model could play a role in the facilitation of infants' categorization through verbal labeling.

Learning to Reproduce Fluctuating Time Series by Inferring Their Time-Dependent Stochastic Properties: Application in Robot Learning Via Tutoring

S. Murata, J. Namikawa, H. Arie, S. Sugano, J. Tani

This study proposes a novel type of dynamic neural network model that can learn to extract stochastic or fluctuating structures hidden in time series data. The network learns to predict not only the mean of the next input state, but also its time-dependent variance. The training method is based on maximum likelihood estimation by using the gradient descent method and the likelihood function is expressed as a function of the estimated variance. Regarding the model evaluation, we present numerical experiments in which training data were generated in different ways utilizing Gaussian noise. Our analysis showed that the network can predict the time-dependent variance and the mean and it can also reproduce the target stochastic sequence data by utilizing the estimated variance. Furthermore, it was shown that a humanoid robot using the proposed network can learn to reproduce latent stochastic structures hidden in fluctuating tutoring trajectories. This learning scheme is essential for the acquisition of sensory-guided skilled behavior.

Conceptual Imitation Learning Based on Perceptual and Functional Characteristics of Action

H. Hajimirsadeghi, M. N. Ahmadabadi, B. N. Araabi

This paper presents a conceptual model for imitation learning to abstract spatio-temporal demonstrations based on their perceptual and functional characteristics. To this end, the concepts are represented by prototypes irregularly scattered in the perceptual space but sharing the same functionality. Functional similarity between demonstrations is understood by reinforcements of the teacher or recognizing the effects of actions. Abstraction, concept acquisition, and self-organization of prototypes are performed through incremental and gradual learning algorithms. In these algorithms, hidden Markov models are used to prototype perceptually similar demonstrations. In addition, a mechanism is introduced to integrate perceptions of different modalities for multimodal concept recognition. Performance of the proposed model is evaluated in two different tasks. The first one is imitation learning of some hand gestures through interaction with the teachers. In this task, the perceptions from different modalities, including vision, motor, and audition, are used in a variety of experiments. The second task is to learn a set of actions by recognizing their emotional effects. Results of the experiments on a humanoid robot show the efficacy of our model for conceptual imitation learning.

A Robotic Model of Reaching and Grasping Development

P. Savastano, S. Nolfi

We present a neurorobotic model that develops reaching and grasping skills analogous to those displayed by infants during their early developmental stages. The learning process is realized in an incremental manner, taking into account the reflex behaviors initially possessed by infants and the neurophysiological and cognitive maturation occurring during the relevant developmental period. The behavioral skills acquired by the robots closely match those displayed by children. The comparison between incremental and nonincremental experiments demonstrates how some of the limitations characterizing the initial developmental phase channel the learning process toward better solutions.

Volume 6, Issue 1, March 2014

An Approach to Subjective Computing: A Robot That Learns From Interaction With Humans

P. Gruneberg, K. Suzuki

We present an approach to subjective computing for the design of future robots that exhibit more adaptive and flexible behavior in terms of subjective intelligence. Instead of encapsulating subjectivity into higher order states, we show by means of a relational approach how subjective intelligence can be implemented in terms of the reciprocity of autonomous self-referentiality and direct world-coupling. Subjectivity concerns the relational arrangement of an agent's cognitive space. This theoretical concept is narrowed down to the problem of coaching a reinforcement learning agent by means of binary feedback. Algorithms are presented that implement subjective computing. The relational characteristic of subjectivity is further confirmed by a questionnaire on human perception of the robot's behavior. The results imply that subjective intelligence cannot be externally observed. In sum, we conclude that subjective intelligence in relational terms is fully tractable and therefore implementable in artificial agents.

LIDA: A Systems-level Architecture for Cognition, Emotion, and Learning

S. Franklin, T. Madl, S. D'mello, J. Snaider

We describe a cognitive architecture learning intelligent distribution agent (LIDA) that affords attention, action selection and human-like learning intended for use in controlling cognitive agents that replicate human experiments as well as performing real-world tasks. LIDA combines sophisticated action selection, motivation via emotions, a centrally important attention mechanism, and multimodal instructionalist and selectionist learning. Empirically grounded in cognitive science and cognitive neuroscience, the LIDA architecture employs a variety of modules and processes, each with its own effective representations and algorithms. LIDA has much to say about motivation, emotion, attention, and autonomous learning in cognitive agents. In this paper, we summarize the LIDA model together with its resulting agent architecture, describe its computational implementation, and discuss results of simulations that replicate known experimental data. We also discuss some of LIDA's conceptual modules, propose nonlinear dynamics as a bridge between LIDA's modules and processes and the underlying neuroscience, and point out some of the differences between LIDA and other cognitive architectures. Finally, we discuss how LIDA addresses some of the open issues in cognitive architecture research.

Development of First Social Referencing Skills: Emotional Interaction as a Way to Regulate Robot Behavior

S. Boucenna, P. Gaussier, L. Hafemeister

In this paper, we study how emotional interactions with a social partner can bootstrap increasingly complex behaviors such as social referencing. Our idea is that social referencing as well as facial expression recognition can emerge from a simple sensory-motor system involving emotional stimuli. Without knowing that the other is an agent, the robot is able to learn some complex tasks if the human partner has some "empathy" or at least "resonate" with the robot head (low level emotional resonance). Hence, we advocate the idea that social referencing can be bootstrapped from a simple sensory-motor system not dedicated to social interactions.

Object Learning Through Active Exploration

S. Ivaldi, S. M. Nguyen, N. Lyubova, A. Droniou, V. Padois, D. Filliat, P.-Y. Oudeyer, O. Sigaud

This paper addresses the problem of active object learning by a humanoid child-like robot, using a developmental approach. We propose a cognitive architecture where the visual representation of the objects is built incrementally through active exploration. We present the design guidelines of the cognitive architecture, its main functionalities, and we outline the cognitive process of the robot by showing how it learns to recognize objects in a human-robot interaction scenario inspired by social parenting. The robot actively explores the objects through manipulation, driven by a combination of social guidance and intrinsic motivation. Besides the robotics and engineering achievements, our experiments replicate some observations about the coupling of vision and manipulation in infants, particularly how they focus on the most informative objects. We discuss the further benefits of our architecture, particularly how it can be improved and used to ground concepts.

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