Berthouze, L., Kaplan, F., Kozima, H., Yano, H., Konczak, J., Metta, G., Nadel, J., Sandini, G., Stojanov, G. and Balkenius, C. (Eds.)
Proceedings of the Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems
Lund University Cognitive Studies, 123. ISBN 91-974741-4-2

Scaffolding Cognition with Words

Robert Clowes
robertc@sussex.ac.uk

Anthony F. Morse
anthonfm@sussex.ac.uk

Centre for Research in Cognitive Science COGS
University of Sussex, Brighton UK

Abstract

We describe a set of experiments investigating the role of natural language symbols in scaffolding situated action. Agents are evolved to respond appropriately to commands in order to perform simple tasks. We explore three different conditions, which show a significant advantage to the re-use of a public symbol system, through self-cueing leading to qualitative changes in performance. This is modelled by looping spoken output via environment back to heard input. We argue this work can be linked to, and sheds new light on, the account of self-directed speech advanced by the developmental psychologist Vygotsky in his model of the development of higher cognitive function.

1. The Approach

We examine experiments designed to explore the way language directed toward self, which has previously been borrowed from an inter-agent communication, can come to play a role in intra-agent cognition. Our purpose is to try an elucidate the way that language becomes involved in behavioural tasks and how it can sculpt cognitive development; (cf. Clark, 1998).

Words here are not spontaneously developed by the agents, or self-organised between agents, but rather come available as ready-mades from their human creators. That said, human language is *appropriated* by the agents, in a way which is we argue analogous to how children appropriate the language of the society in which they grow up. Agents do not simply associate internal categorisations with particular 'words' but rather use words to accomplish and structure their cognitive processes.

We use the genetic algorithm to train the agents. However the language-using tasks are not about the evolution of language *per se* – e.g. (Christiansen & Kirby, 2003) but rather demonstrates how language elements can be appropriated to allow the reduction of an agent's cognitive workload by the re-organisation of its activities. By looking at this largely unexamined process in an evolutionary scenario we hope it is possible to build some generalisations of use to developmentalists.

Developing behaviours to respond to words provides new bases of stability – scaffolds - upon which more elaborate behaviours can be constructed. The work presented here provides a window into the beginnings of this process.

2. The Task

The agents in our experiments manipulate a window on a 2D world made up of geometric figures. Each agent begins a game tasked with a goal such as move an object down to a target area at the bottom of the screen.

Games are here always made up of one of four conditions: 'move objects to the top', 'move objects to the bottom', 'move objects to the left' or, 'move objects to the right'. At the beginning of each game the agent receives (as 'word' input) instructions telling which of these goals is in operation. This signal is intermittently repeated. Success requires the agent to find and move an object (objects can be carried in the agents view window) to the goal area specified by the signal sent to the agent's word inputs. By finding objects (moving the view window over an object), and then carrying it (moving the window while holding the object), agents are involved in an active restructuring of the world and are not simply - as is the case with much comparable work - labelling features or aspects of it.

We hypothesised that agents able to re-use public language (from the commands they received) to act as a scaffold adaptively regulating their own activity in order to achieve the goals of new or complex task situations with greater efficacy and less time taken. That is to say, the re-use of language should not only speed up the acquisition of normal behaviours, but further enable successful operation at tasks beyond the original capabilities of the agent.

3. The Network Architecture

Developing the architecture of (Floreano, Kato, Marocco, Sauser, & Suzuki, 2003) used in active vision research. We implement a simple recurrent neural network (SRNN) (see diagram) as an agent control system in the language games described. The network comprises of 50 input nodes and 14 output nodes with no hidden units. Visual input to the network comprises of 27 visual inputs (arranged in 3 by 3 grids of red, green, and blue pixel values), a border unit (active when the agent is at the edge of the game screen), 5 word units (intermittently indicating the current game goal), a 'holding' unit (active when an object is currently being held), and the current position (X & Y value). The input units were fully connected (feed-forward) to the output nodes, interpreted as the actions; 'up', 'down', 'left', 'right', 'widen', 'close', 'grab', 'drop', and 'sample'/ 'average' (see Floreano et al), plus an additional 5 output words. The entire output at any given time step is copied to directly form part of the input layer at the next time step.

In additional modifications to the architecture, the output words can be copied directly (as additional input) to the input words (thus hearing what is said). Variations in this additional mechanism form three experimental conditions investigating self-directed speech:

Condition 1: The network has an additional output unit (or 'gating' neuron), switching the word re-entrance loop on or off. This allows agents to control whether or not they can hear themselves. (Agents act on instructions and gate whether or not they instruct themselves).

Condition 2: The word re-entrance loop is permanently on. In this condition an agent can always here the words it is saying. (Robots act on instructions and trigger their own instruction nodes).

Condition 3: A condition where words are not re-entrant, i.e. a control group. (Robots act on external instructions alone).

Agent architectures (as described) were evolved to maximise fitness defined as the product of 4 game scores (one in each direction, see 'The Task') where for each game the score is incremented once for each time-step only if the agent has successfully moved a shape into the appropriate scoring area (the shape must stay in the score area to continue scoring).

In one of the simulations reported by Floreano et al's a Genetic Algorithm was used to select neural architectures that could perform the task of discriminating between circles and squares. The architectures of the agents in our experiments have a crucial difference to that of Floreano. Whilst Floreano's agents are essentially world labellers and evolved just to achieve this task, our agents are evolved to be world manipulators moving objects in the scene to follow instructions.

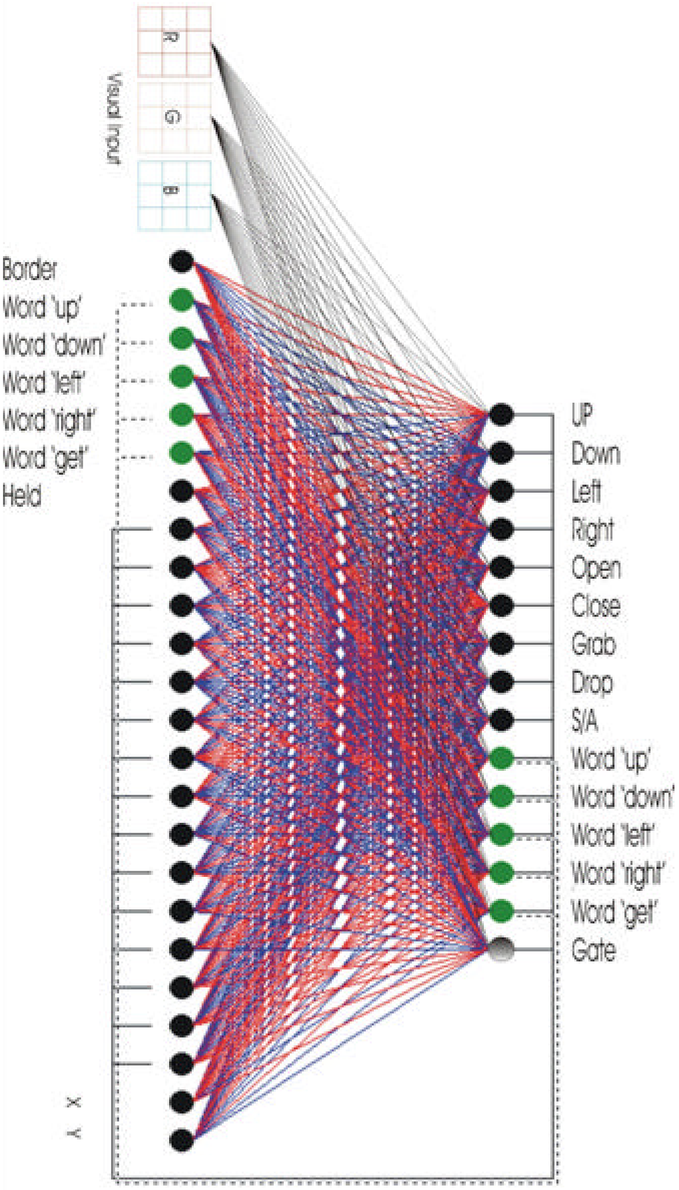
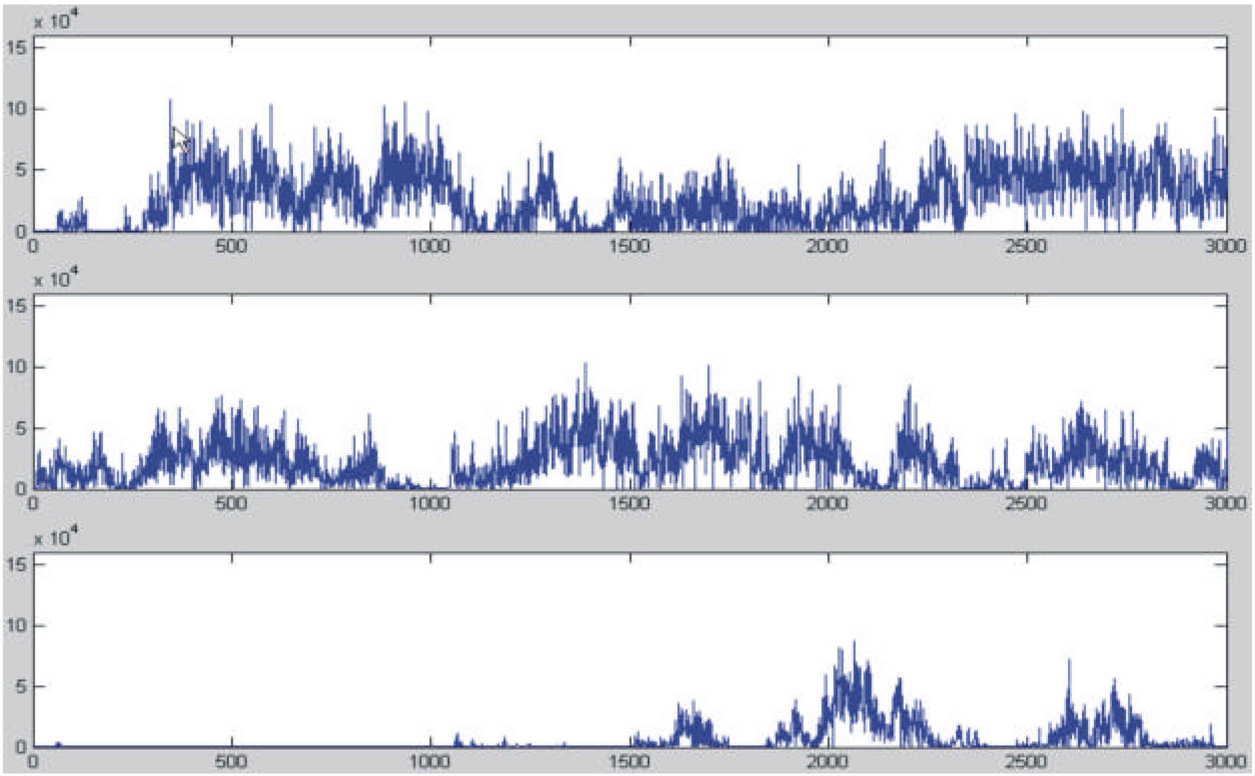


Figure 1: showing the Network Architecture.

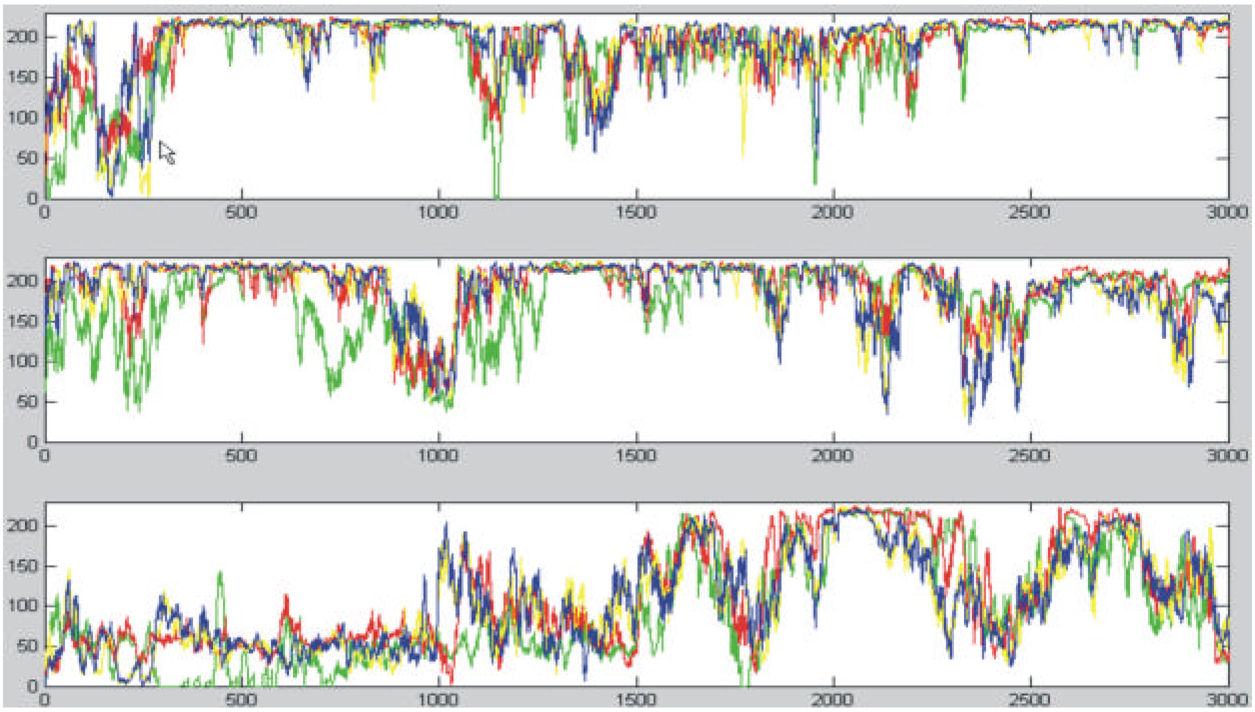
Dotted line shows the re-entrance of conditions 1 and 2.
The gating neuron –only used in condition 1 – which gates word re-entrance when on is shown as being shaded.

4. Results

The three conditions showed at least two significant effects in simulation. Most obvious was that when language is re-entrant, as in conditions 1 and 2, high-fitness is achieved in far fewer generations than in the control condition where language is not re-entrant (see graphs 1 and 2).



Graph1: Showing the fitness across all four game types (dot product) averaged per generation. **Top** shows condition 1 (selective re-entrant), **Mid** shows condition 2 (re-entrant), **Bot** shows condition 3.



Graph2: Showing the scores for each game type of the fittest individual per generation averaged over four sequential generations[1]. **Red** – 'up', **Green** – 'down', **Yellow** – 'left', **Blue** – 'right'. **Top** shows condition 1 (selective re-entrant), **Mid** shows condition 2 (re-entrant), **Bot** shows condition 3.

[1] Because of the contingencies of the game, scoring can be quite variable and averaging over 4 consecutive generations highlights agents that consistently score highly in a particular game types.

Learning to succeed in each direction task, places a considerable burden on the evolution of the agent architecture. In conditions 1 and 2, agents typically achieve high levels of performance in at least three and often all four game types (i.e. 'up', 'down', 'left' and 'right' see graph 2), while high performance cannot be robustly demonstrated by agents in condition 3. Despite the prolonged evolution necessary to achieve high fitness in this condition, improvement in any condition, at any particular game type usually incurs cost to one or more of the other game types (see graph 2 e.g. condition 2 600-800 generations). It is here that agents in with word re-entrance (conditions 1 and 2) demonstrate a qualitative increase in their performance over those without. To investigate the possibility that the additional recurrent connections (although not providing new units) may increase the 'memory' of the network, condition 3 was re-run with an additional 5 output & context-input units, bestowing it with significantly more memory than the networks in conditions 1 and 2. Although not shown here for space reasons, agents in this condition 3+ show only slight increases in performance over the standard condition 3 agents, remaining significantly below the performance of agents in conditions 1 and 2.

Observing the activity of the additional 'gating' unit in condition 1, shows that in early generations the unit is typically inactive, while in later generations it is generally active. We suppose that until the agent has knowledge of the input word meanings (functionally), the re-entrant loop simply provides noise, degrading performance (and is hence switched off). Once the words are known, they can be triggered to self-cue behaviour (and hence this loop is switched back on). In order for there to be functional advantages to the word re-entrance loop, the output words must be organized (involving an evolutionary load). Every condition 1 agent evolved (passed 1000 generations), used its re-entrance loop, this is evidence of the advantages such a loop supports. Although small, this effect was robust and can be seen in graph 2 by comparing conditions 1 and 2 at around 200 generations. The drop in performance here in condition 1 is attributed to the activation of the word re-entrance loop.

These results clearly demonstrate a qualitative difference between the control group (condition 3) and the remaining conditions, despite the internal re-entrance of SRNN architectures present in all three conditions.

Analysis of the weights of several successful agents in each condition has identified numerous architectural solutions; however there are clear structural (implying functional) differences between those agents in condition 3 and those in conditions 1 and 2.

5. Scaffolding and Language for Thought

Some reflection might be in order at this point into the possible reasons why the robots whose architecture allows the use of re-entrant commands might have an advantage over non language re-entrant agents. In recent years there has been a great deal of interdisciplinary interest, in the way that language might play a role in cognition (Carruthers & Boucher, 1998; Clark, 1996; Dennett, 1994). The traditional picture in non-connectionist circles has been that an internal language of thought is entirely separate from natural languages. What's sometimes known as the received view (Fodor, 1987) postulates that thinking happens in an internal realm structured in a quasi-linguaform manner, a mentalese, or language of thought. Natural language on the other hand is considered as merely a communicational system, handled by a cognitive "sensory" system (Fodor, 1983). We might sloganise this view as 'LOT for computation, Natural language for communication[2]'.

An alternative picture comes from the early connectionist literature, present in a *radical* version in the work of O'Brien & Opie (2002), taking the ability of language to operate as a conventional system of signs (Saussure, 1959) as the basis for understanding the cognitive capacities it confers. A version of this idea is present in Deacon's (1997) as well as work on adaptive language games developed by Steels and his co-workers. Although language can co-ordinate cognitive abilities between agents, what this work fails to show is how language could come to play the same role within agents[3].

Theoretical models of how inter-agent communication, might become intra-agent cognitive structure, has antecedents stretching back to the pioneering work of Vygotsky in the 1930s (Vygotsky, 1986) and extended by Bruner (Bruner & Sherwood, 1975) with the notion of scaffolding. These accounts focused on the role that language plays in intimately structuring the learning environment, allowing the construction of more complex cognitive activities. This notion of scaffolding refers to the

[2] A more recent concession toward the idea that natural language might play some role in doing cognitive work can be found in Carruthers (2002) paper the cognitive function of language. This view still relies on intra-modular mentalese as indeed do other "mixed" theories such as that of (Devitt & Sterelny, 1987) which argue for a reduced basic mentalese with internalised natural language playing a subsidiary role.

[3] There is an important sequence of work on related themes showing that categorical perception can be developed over generations (Cangelosi, 1999, 2001; Cangelosi, Greco, & Harnad, 2000). However this work shows the development of categories which are of use to agents in recognising relevant aspects of their environment, rather than how signs might be incorporated in a behavioural repertoire to restructure activities.

use of external structures in both material, and social culture, supporting the development of higher forms of cognition. Scaffolding like effects are demonstrated in these experiments where self-cuing enhances the development of robust (and behaviourally distinct) action systems in conditions one and two. As shown in Graph 2, agents with re-entrant language learn to use a system of commands much more quickly than those without it.

The models discussed in this paper also provide a new way of approaching the question of languages role in the development of cognition. They explicitly acknowledge the role of natural language in the development of the coordination of behaviour. Although recent work in developmental robotics has hinted that it should be possible to build models of such processes, this has been hampered by the theoretical baggage inherited from an earlier era of cognitive research. By re-conceptualising crucial elements of these theoretical models we hope to have shown how existing techniques can be used to look at processes like scaffolding, and start to illuminate some of its internal dimension. Hints from this current work are that the dynamical properties of agents which can auto-stimulate with words can develop in quite different ways from those lacking such capabilities.

One way of framing this theoretically is to consider the *semiotic potential intelligence* available in the commands presented to the agents. In our experiments we use a genetic algorithm to train agents to respond to commands that are based on the semantic categories, which it seems natural for human beings to apply to the scene. Using labels such as “up”, “down”, “left” and “right” - as well as colour and shape terms that we are currently investigating in related experiments - does not merely rely on meanings which are objectively given in the scene, but on socially derived categorical forms which are present in the socially created meaning systems on which human language relies. The Agents with re-entrant connections to the linguistic command nodes develop the ability to take advantage of at least some of the semiotic potential intelligence in these signs and thus develop more robust, and differently structured, solutions. Future work will attempt to analyse this process in much greater depth and embed the language using behaviour in a more explicitly developmental architecture.

References

Bruner, J. S., & Sherwood, V. (1975). Peekaboo and the learning of rule structures. In J. S. Bruner & K. Sylva (Eds.), *Play: Its role in development and evolution*. Harmondsworth, England: Penguin Books.

Cangelosi, A. (1999). Modelling the evolution of communication: From stimulus associations to grounded symbolic associations. In D. Floreano (Ed.), *Proceedings of ECAL99 European Conference on Artificial Life* (pp. 654-663). Berlin: Springer-Verlag.

Cangelosi, A. (2001). Evolution of Communication and Language Using Signals, Symbols, and Words.

Cangelosi, A., Greco, A., & Harnad, S. (2000). From Robotic Toil to Symbolic Theft: Grounding Transfer from Entry-Level to Higher-Level Categories. *Cognitive Science, 12*(2), 143 - 162.

Carruthers, P. (2002). The Cognitive Function of Language. *Behavioral and Brain Sciences, 25*(6).

Carruthers, P., & Boucher, J. (1998). *Language and Thought: Interdisciplinary Themes*: Cambridge Univeristy Press.

Christiansen, M. H., & Kirby, S. (Eds.). (2003). *Language Evolution: The States of the Art*.

Clark, A. (1996). Linguistic Anchors in the Sea of Thought? *Pragmatics And Cognition, 4*(1), 93-103.

Clark, A. (1998). Magic Words: How Language Augments Human Computation. In P. Carruthers & J. Boucher (Eds.), *Language and Thought. Interdisciplary Themes* (pp. 162 - 183). Oxford: Oxford University Press.

Deacon, T. W. (1997). *The Symbolic Species: The Co-Evolution of Language and the human brain*: The Penguin Press, Penguin Book Ltd.

Dennett, D. C. (1994). The Role of Language in Intelligence. In D. C. Dennett (Ed.), *What is Intelligence*. Cambridge: Cambridge University Press.

Devitt, M., & Sterelny, K. (1987). *Language and Reality: An Introduction to Philosophy of Language*.

Floreano, D., Kato, T., Marocco, D., Sauser, E., & Suzuki, M. (2003). *Active Vision & Feature Selection: Co-development of active vision control and receptive field formation. Complex visual performance with simple neural structures.* Retrieved 30 June 2004

Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J. (1987). *Psychosemantics*: MIT Press.

O'Brien, G., & Opie, J. (2002). Radical Connectionism: Thinking With (Not in) Language. *Language and Communication, 22*, 313-329.

Saussure, F. d. (1959). Course in General Linguistics. In. New York: McGraw_Hill.

Vygotsky, L. S. (1986). *Thought and Language* (Seventh Printing ed.): MIT Press.