The Acquisition of New Categories through Grounded Symbols: An Extended Connectionist Model

Alberto Greco[1], Thomas Riga[1], Angelo Cangelosi[2]

[1] Psychology Division, Department of Anthropological Sciences, University of Genoa,
vico S. Antonio 7, Genoa, Italy
greco@disa.unige.it   thomasriga@yahoo.com

[2] Centre for Neural and Adaptive Systems, School of Computing, University of Plymouth,
Drake Circus, PL4 8AA Plymouth, UK
acangelosi@plymouth.ac.uk

Abstract. Solutions to the symbol grounding problem, in psychologically plausible cognitive models, have been based on hybrid connectionist/symbolic architectures, on robotic approaches and on connectionist only systems. This paper presents new simulations on the use of neural network architectures for the grounding of symbols on categories. In particular, the connectivity patterns between layers of the networks will be manipulated to scale up the performance of current connectionist models for the acquisition of higher-order categories via grounding transfer.

1 The Grounding of Symbols in Categories

Cognitive models dealing with linguistic and symbol-manipulation tasks can use symbols that are either grounded or ungrounded (i.e. self-referential). Grounded symbols are those inherently significant to the cognitive system, such as an agent, and not mediated by the interpretation of an external user. Self-referential symbolic systems are those that use symbols that have no grounding in any other module of the cognitive agent. It has been claimed [5] that the cognitive relevance and psychological plausibility of a self-referential symbolic system is diminished as a result of the symbol grounding problem. To solve the problem, Harnad [5] suggested that symbols should be intrinsically linked to the agent's ability of acquiring categories from everyday experience it has of its environment. In particular, it is necessary that some basic symbols are directly grounded on sensorimotor categories. Subsequently, new (grounded) categories can be formed through the combination of previously grounded basic symbols.

Hybrid symbolic-connectionist models were originally proposed as ideal candidates for solving the symbol grounding problem [6]. More recently, alternative approaches have been introduced. Robotics approaches to symbol grounding focus on social learning and interaction between agents (including robots, internet agents and humans) to ground shared symbol communication systems. This has been implemented, for example, in experiments on robotic language games [9]. Fully connectionist models have also been proposed to deal with the symbol grounding problem [1,7,8]. For example, in [1] the ability of neural networks to acquire a small set of

basic categories through direct sensorimotor grounding was tested. The same networks were subsequently trained to acquire new higher-order categories solely through combination of the name of basic categories (symbolic theft). These networks were able to transfer the grounding from sensorimotor categories to higher-order categories learnt via symbol combination. Such an approach has also been used in evolutionary simulations of language origins [2].

Research on the connectionist implementation of grounded symbolic cognitive agents is still in progress. In particular, effort has focused on the design of modular connectionist architectures and its contribution in dealing with the nature/nurture debate (e.g. [3]). This paper presents new simulations based on the manipulation of the connectivity pattern of multi-layer perceptrons for the grounding of symbols on categories. In addition, it will deal with some problems of current connectionist architectures, such as the scaling up of categories and symbols.

2 Simulation One

In the first model, Cangelosi, Greco and Harnad's [1] model (CGH, thereafter) will be expanded to deal with larger category sets, and to look at different aspects of the transfer of grounding. In previous studies [10], the same fully-connected architecture from CGH was used with larger category sets. These included extra entry-level categories (e.g. 3 basic categories, each constituted by 2 exemplars), larger entry-level categories (2 basic categories, each constituted by 3 exemplars), and more high-order levels of categories (27 basic order categories which form 9 high-order categories, which then form 3 higher-order categories). Fully connected multi-layer perceptrons failed to transfer the grounding in any of the three levels of extension of the model. This indicated a significant shortcome of the proposed neural network model [1] for the symbol grounding. This paper presents a series of new simulations in which some of these limitations have been overcome by manipulating the pattern of connections between groups of units.

The goal of the first simulation is to use a fully connectionist architecture to scale up the performance of CGH with more categories (4x4 basic and 4 higher-order categories).

2.1 The Stimulus Set

The total stimulus set consisted of 396 images, 216 for the training and 180 for the generalization test (cf. prototypical stimuli in Fig. 1). These images were derived from the animal picture set of the second experiment of CGH [1]. Each stimulus consists of a 50x50 pixel image. A single image can represent an isolated shape, a texture or an animal obtained by combining a specific shape and texture. Four different animal shapes (e.g. a horse shape), four textures (e.g. a striped pattern) and four animals (zebra = horse shape + stripes pattern) were used. The four shapes and the four textures constitute the (basic) entry-level categories. These are learned through direct sensorimotor grounding in categorization and naming learning stages. The

four animals constitute the higher-order categories and are learned through symbolic theft. These are also used for the grounding transfer test only.

The training stimulus set was augmented by placing each image of the 8 entry-level categories into 27 different positions on the retina. This resulted in 216 training images. The testing set was also augmented by placing the 4 animals in 45 different positions in the retina image (9 spatial translation of the shapes x 4 translation of the texture position).
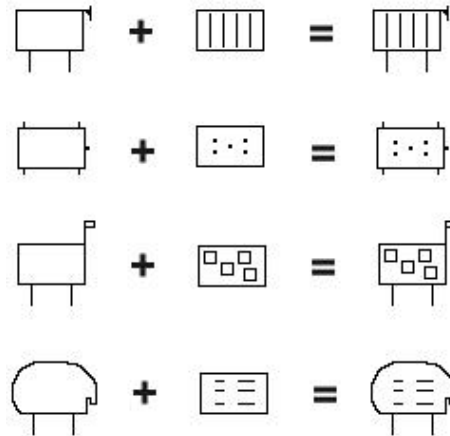

Fig. 1. Prototype of categories used in simulation 1. (Left) The shapes and textures of the Entry-Level categories. (Right) Higher-Level categories.

2.2 The Neural Network and Training Procedure

The architecture of the network has been significantly changed with respect to that of CGH. Each network has three layers of units, with connections to and from hidden units modularly organized. The network contains 61 input units, 49 for the retina and 12 for the category names (Fig. 2). The same type and number of units are used in the output layer. The 49 (7x7) retina input units consists of gaussian receptive field units. These process the 50x50 pixels of the original image, each using a square receptive field of 11x11 pixel [1,4]. Retina units are divided into two groups, the periphery and the center. The 6 hidden units are also divided into two groups of 3 units each, one specialized for shapes and one for textures. The periphery input units send connections only to the 3 shape hidden units. The retina central units send connections only to the texture hidden units. This is due to the fact that the units in the periphery of the retina encode the part of the image representing the various animal shapes. The central units encode the texture in the center of images. Twelve localist symbolic input and output units encode the category names (4 entry-level shapes, 4 entry-level textures, 4 higher-level animals).

The networks were trained using the error backpropagation algorithm. Training was similar to that of CGH and consisted of three stages: prototype sorting, entry-level learning and higher-level naming and imitation (Fig. 3a). During the prototype

sorting stage (i.e. entry-level categorization), networks learn the basic categories (4 animal shape and 4 textures) by receiving input exclusively from the retina images and responding with a retina representation of the prototype of the category (e.g. a fixed, centered shape of a horse). The entry-level learning stage consists of two network activation cycles, the naming and imitation cycles. In the naming cycle, the network sees the retina image and responds in output with the prototypical retinal image and the localist unit encoding the category name. In the imitation, only the symbolic units are used in both the input and the output layers. During the first two stages, learning occurs through direct trial and error experience supervised by corrective feedback ('sensorimotor toil'). Therefore, names acquired this way can be considered as symbols grounded in retinal input. In the higher-level stage the networks acquired new names defined solely on the basis of symbolic strings containing combinations of previously grounded names ('symbolic theft').

The final stage consisted of the grounding transfer test. New retina images exhibiting combinations of previously learned shapes and textures (e.g. images of zebras obtained by combining a horse shape and the striped pattern) were presented to the networks. The test aims to establish whether the networks, which have never seen these images before, are able to correctly categorize and name images with entry- and higher-level symbols.
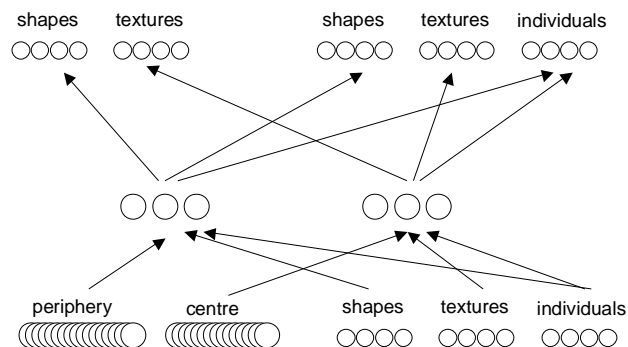

Fig. 2. Neural network architecture for simulation 1

2.3 Results

The simulation consisted of the training of 10 networks with different initial random weights. The momentum was 0.9 for all stages. The learning rate was 0.2 for the first stage and 0.5 for stages two and three. The stimuli were always presented in random order. All networks completed the three training tasks successfully. After training, all networks learnt the various entry- and higher-level categories. The percentage of images correctly categorized in all training stages is 100%. The percentage of correct responses (i.e. production of the correct output name) was computed by using the unit with highest activation to select the name of the input image.

The results on the grounding transfer test were also very positive. About 80% of higher-order animal images were correctly categorized and names. This clearly

shows that grounding is "transferred" from directly grounded names to higher-order ones (grounding transfer). Moreover, the networks were able to give the correct sensorimotor response when they received the name of a higher-level category in input (inverse grounding transfer).

This model dealt well with a scaled up stimulus set of 4x4 basic categories and 4 higher-order categories. However, the separation of input and output retina units into peripherical and central units was somewhat artificial. This separation was essential to achieve successful grounding transfer results. It is likely that this was due to the design of stimuli with no overlap in the retina for the position of the animal shapes and the texture pattern. In the next model, this problem will be dealt with by using a stimulus set with complete overlap of entry-level category features.
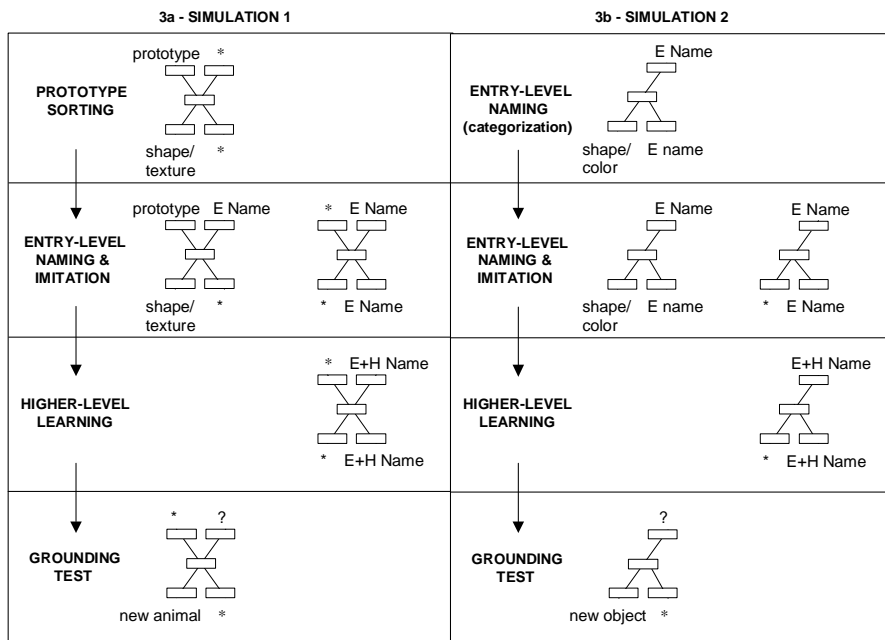

Fig. 3. Training and test stages for simulation one (3a) and two (3b). EL = Entry-Level, HL=Higher-level categories.

3 Simulation Two

This simulation will use a new neural network architecture and a new set of categorization stimuli. The objective is to avoid the artificial division of retina units into peripheral and central units. This division did not have any plausible justification, but was simply introduced to facilitate the network in the classification of the specific set of animal picture stimuli. Only the modular organization of hidden-to-output connections will be preserved.

3.1 The Stimulus Set

The stimuli consisted of 81 abstract images, 54 for training purposes and 27 reserved for testing. Each image, constituted by a 5x5 pixel drawing, was obtained by combining 3 different shapes (square, cross, dots) with 3 different colors (red, green, blue) in 9 different positions (Fig. 4). Every pixel of the image is presented to the network with three input units, coding the primary color components (red, green, blue: RGB).
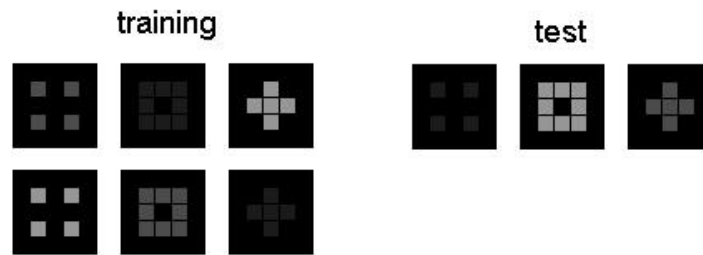

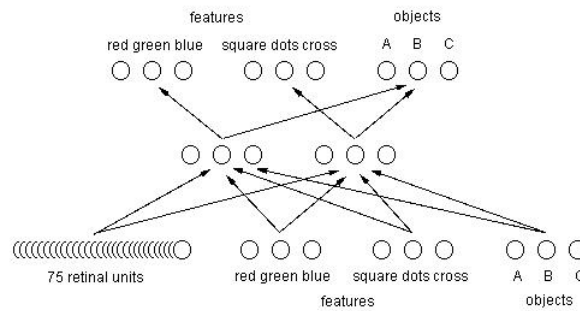**Fig. 4**. Sample training and testing stimuli for simulation 2

Fig. 5. The neural network architecture of simulation 2

The names of categories are encoded with localist symbolic input and output units. These may contain the names of the categories of different visual features (e.g. blue, square), the name of the object as a whole, or a full description of the objects (e.g. blue + square = object A).

This new stimulus set is intended to be more flexible than the one used in Simulation one. Specifically, in this simulation the distinction between the periphery and the center of the retina image becomes irrelevant since the color and shape completely overlap.

3.2 The Neural Network and Training Procedure

A three-layer, feedforward neural network was used (Fig. 5). The input layer consisted of a retina with 75 units and a symbolic input group containing nine units. The retina had three units for each of the 25 pixels, measuring its RGB component. The

symbolic input group consisted of nine units receiving names for the colors, shapes and objects perceived. The hidden layer had six units, organized into two separated groups of three units each. The output layer was structured in the same way as the symbolic input group, with nine units representing the symbolic output.

The input layer was fully connected to the hidden layer. The symbolic output units representing the names for the objects were fully connected to all hidden units. The output units indicating shapes were only connected to the first three hidden units, whilst those indicating colors had connections solely to the last three hidden units. This modular connectivity forces the functional division of the hidden layer into a group dedicated to categorizing shapes and a group to classify colors. Note that the retina units are not artificially divided into two groups as in simulation one.

The error backpropagation algorithm was used to train the network. Training was organized in three stages: Entry-Level naming, Entry-Level imitation and naming, and Higher-Level learning (Fig. 3b). In the Entry-Level naming stage the neural nets are initially trained to categorize (through naming) the color and shape of objects perceived on the retina. The retinal stimuli and names for colors and shapes are presented simultaneously in input. The networks learn to respond to the symbolic and retinal stimuli indicating the corresponding names for the color and shape in output. In the Entry-level imitation and naming stage an extra imitation learning cycle is executed in addition to the repetition of the naming cycles. Imitation consists on the use of only the symbolic units in both the input and the output layers. During these first two stages the networks learn through direct trial and error experience supervised by corrective feedback. Visual stimuli are categorized and linked to arbitrary grounded names.

In the third training phase (Higher-Level learning), networks acquire new higher-order categories through symbolic descriptions only. New categories are built by combining grounded names. Each description contains the name of a shape, a color and the name of a object that is new to the network. The grounding test is performed at the end of training.

3.3 Results

The training procedure was replicated with 30 networks having different initial random seeds. The momentum was 0.9 for all stages. The learning rate was 0.2 for the first stage and 0.5 for stages two and three. The stimuli were presented in random order during EL categorization and in sequential order afterwards.

All 30 networks completed the three training tasks successfully. The networks categorized the colors and shapes of the training stimuli correctly, with a success rate of 100%. The percentage of correct responses (i.e. production of the correct output name for the shape/color/object input) was computed using a winner-takes-it-all approach in which the unit with highest activation determines the name of the input image.

After the networks had completed the final stage, 27 retinal stimuli depicting new objects were presented to the networks for the first time, in order to check if grounding had been "transferred" from directly grounded names to higher-order categories. The rate of correct test responses for the 30 nets was 85%. Even when the networks

had never seen the test images before, they were able to categorize most of them correctly.

4 Conclusion

In this paper we have presented two simulations that model autonomous cognitive systems, immune to the symbol grounding problem: the connections between symbols and their meanings are direct and intrinsic to the system, without need for mediation by an external interpreter.

Our results reinforce the approach to symbol grounding based on fully connectionist models. The same network processes both the sensorimotor grounding and the generation of new categories through symbolic learning. The modular organization of the hidden units suggests that it is important that sensorimotor grounding be separated for different classification features. In fact, when a fully distributed network was used [10], the grounding transfer was difficult to achieve.

In order to improve the psychological plausibility and scalabilty of connectionist approaches to the symbol grounding, various extensions of the models presented here are being studied. For example, alternative learning algorithms like Kohonen's self organizing map and hebbian learning are being tested for the basic categorization stage.

Acknowledgements

References

1. Cangelosi A, Greco A, Harnad S (2000). From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level Categories. *Connection Science*, 12: 143-162
2. Cangelosi A, Harnad S (2000). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication*, 4: 117-142
3. Elman JL, Bates EA, Johnson MH, Karmiloff-Smith A, Parisi D, and Plunkett K (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge: MIT
4. Jacobs RA, Kosslyn SM (1994). Encoding shape and spatial relations: The role of receptive field size in coordinating complementary representations. *Cognitive Science*, 18: 361-386.
5. Harnad S (1990). The symbol grounding problem. *Physica D*, 42: 335-346
6. Harnad S (1993). Grounding symbols in the analog world with neural nets. *Think*, 2: 12-78
7. Harnad S, Hanson SJ, Lubin J (1995). Learned categorical perception in neural nets: Implications for symbol grounding. In Honavar V, Uhr L (Eds) *Symbol Processors and Connectionist Network Models in Artificial Intelligence and Cognitive Modeling: Steps toward principled integration*. Academic Press (p. 191-206)
8. Plunkett K, Sinha C, Moller MF, Strandsry O (1992). Symbol grounding or the emergence of symbols? Vocabulary grouth in children and a connectionist net. *Connection Science*, 4(3-4): 293-312
9. Steels L (2002). Grounding symbols through evolutionary language games. In Cangelosi A, Parisi D (Eds) *Simulating the Evolution of Language*, (p. 211-226), London: Springer
10. Stuart EJ, Cangelosi A (1999). Unpublished data. University of Plymouth