From Action to Symbols and Back: Are There Action Symbol Systems?

Alberto Greco (greco@unige.it) Claudio Caneva (clac77@libero.it)
Laboratory of Psychology and Cognitive Science, v.Balbi 6
University of Genoa, Italy

Abstract

We present an experimental and simulative method for exploring the properties of symbolic action representation. This work is related to the current debate about the compositional vs. holistic nature of such representation. We designed a new experimental setup to be used with human participants and a neural network model as well, in the context of an artificial language learning whose referents are actions. In different conditions, (a) actions are systematically connected with sentences made of arbitrary words, following a simple syntax, and consistently expressing features of involved movements; (b) actions as a whole are consistently associated with arbitrary words; (c) actions are consistently associated with other actions so that the former become symbols for the latter. In such tasks, results show a clear advantage for a holistic representation; there are reasons, however, to suppose this might be a first step towards a compositional representation. This analysis is supported by results with nets.

Introduction

Representations for concepts and actions are increasingly being considered as tightly connected. The study of *embodied* cognition is assuming more and more theoretical significance and a recent trend has emerged that questions the idea of a mental concept representation made by the combination of abstract, arbitrary, and amodal symbols, as the first cognitivist stance had posited. Even if there are still models that insist on symbolic abstract coding (e.g. Landauer & Dumais, 1997; Burgess & Lund, 1997), many authors now claim that linguistic representation is modal and emphasise its analogical aspects. In some cases (Barsalou, 1999; Glenberg & Robertson, 2000; Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Glenberg & Kaschak, 2002) the somewhat radical claim is posed that the comprehension of sentences substantially comes from the possibility of mentally simulating (or "re-enacting") the actual performance of implied actions.

Such perspectives substantially stress the role that internal simulation of concrete action, with its nonarbitrary aspects, has for comprehension of abstract representations. Barsalou (1999), in particular, suggested that perceptual (and motor) representations, even though modal, may have properties of symbols like compositionality. This view (*perceptual symbol systems*) to some extent also implies, conversely, a sort of perceptual-motor language; it still assumes that language can only be understood by referring to actions but also that actions are represented by the representation of their features.

Action representation can be considered as componential if sequences of actions can systematically be combined and recombined according to represented rules. In this case it would be possible to speak of *action symbol systems* (ACSS), i.e. modal representations but with compositional properties. This componential view of action representation may be opposed to a holistic one, that considers representations for actions as global, procedural, implicit and not analytical, i.e. not based upon single features (cf. Vigliocco, Vinson, Lewis & Garrett, 2004, for a thorough discussion), whereas combinatorial symbolic language relies on the combination of feature representations. In the case of ACSS, we should assume a representation mapping that works bidirectionally, from action to language and the converse. This bidirectional relationship is plausible enough, and both ways of the relationship have received consideration also during the history of psychology (Piaget, 1952; Luria, 1960, 1981).

When referring to "action" at least three different kinds of processes, and presumably related representations, could be involved: action visual recognition, action verbal description, and action control (see Rizzolatti, Fogassi, and Gallese, 2001 for neural relationships between the first and the latter). Also, there is a difference between how such representations are *acquired* and how they are *used*. Different cognitive processes are involved in these cases. When *acquisition* is concerned, information must be extracted from a perceived action, associated with verbal labels. This is what children (or adults learning a second language) do when they learn the meaning of words for actions, while observing or performing the action and concurrently listening to a verbal description. In order to develop a systematic vocabulary for features, they must ground single words for features by constructing a consistent mapping with observed or performed actions. In the case of acquisition, then, to construct an ACSS means to associate perceptual-motor parts or components of action to separate representations. In contrast, according to a holistic view, representations for action as global patterns could be constructed; their symbolic nature can be questioned, but in any case they do not constitute a true symbol *system*.

In the case of *use* of representations for actions, there is a difference between *describing* or *executing* actions. In a verbal description task, it seems little efficient to have a symbol for every possible action, and more useful to have an ACSS; on the contrary, an automatic procedure that works holistically could be more suitable in order to execute an action.

The above hypotheses need empirical support. In this paper we present a new paradigm for testing such different aspects: acquisition, execution, description, and compositional use of representations for actions. In our paradigm the source of our empirical data is both experimental and simulative, since we used the same conditions with human subjects and neural networks. Neural networks are a good tool for investigating such questions, because they can naturally implement non-symbolic or analog representations.

In our research, we studied the acquisition and use of an artificial language denoting actions. We wanted to see what kind of representations are developed if a systematic association is established between meaningless motor and verbal stimuli. In

particular we intended to examine whether, if available, such representations may have a modal and compositional nature at the same time, as assumed by the ACSS hypothesis, and their effectiveness for different tasks (description or execution of actions).

We constructed a conceptual universe for actions made of simple movements, systematically associated with a sentence or a single word of an artificial language, or with a different action that acted as a cue. Human participants and neural networks were then trained to associate simple arbitrary actions with standard compositional symbols (words or other actions systematically connected with features) or noncompositional symbols (completely arbitrary words).

In the case of sentences, in establishing the connection between sentences and actions we strove to reproduce the natural process of language learning, where arbitrary symbols are combined following a syntax to systematically express featural regularities. In natural contexts, words are not normally presented in an ordered form, one at a time, each associated with a particular referent. This happens in symbol grounding by explicit teaching, like in our previous work (Cangelosi, Greco & Harnad, 2000; Greco, Riga & Cangelosi, 2003) where we simulated how the representation of abstract and arbitrary symbols can ground higher-level abstract concepts. Actually, language that describes what is happening is already syntactically structured, and the meaning of single words is inferred by abstraction of perceived regularities.

In the case of cue-actions, they were systematically associated to featural aspects of the target-actions set. In order to avoid a direct connection and to allow a possible symbolic mapping, correspondences between the two sets of movements were set up in such a way that target movements were associated with the hand position in the cue set (open, fist, pointing), the body part involved was associated with a movement (outwards or inwards), and the side with the forearm position (upwards, downwards, sideways). Here it is tried to establish an arbitrary connection between movements, in such a way that the ones of the first set may become systematic symbols for the second.

Experiment Method

Participants

36 students from the University of Genoa participated in this study for course credit. They were run individually, randomly and equally assigned to each condition: (a) actions-sentences; (b) actions-words; (c) only actions.

Stimuli

A conceptual universe was first defined including three possible actions (to tap, to raise, and to wave), two body parts (forearm, hand), and three side specifications (left, right, both). An artificial language was then defined, that included a set of 8 words arbitrarily paired to actions, body-parts, and side-specifications. Words were nonsense syllables, i.e. triplets made of a consonant, a vowel, and another consonant, avoiding phonetically confusable combinations. For group (b) bisyllables words were used in order to reduce the effect of associations with known words, more frequent with monosyllables. An arbitrary subset including a half of the 18 possible movements resulting from the combination of movements with body-parts and side-specifications was chosen as set 1, to be used in the learning phase as explained later; the remaining half was designated as set 2 (for test phases). A different set of 9 single words was also defined for denoting movements belonging to set 1. A simple syntax was defined for the artificial language, as the first word always denoted the action, the second the body-part, the third the side (see tab.1; for example, LOF DIN FIT indicates "wave the left forearm"; the same movement could also be expressed by the single word TANEG). Note that subjects were Italian speakers and that this sequence reflected the natural order of words in Italian. For group (c), a different set of actions was designed for cue-actions, systematically associated to features of target actions (the same for all groups). Only the right forearm-hand (for the subject) was used. The hand position could be open, clenched fist or pointing; the forearm could be (already rotated before the start of movement) showing the upper, lower or side part of the hand; the movement from the starting position (hands side by side) could be outwards or inwards the body. Features were denoted as shown in table 1: e.g., hand up, pointing outwards, means "wave the left forearm".

All actions were videotaped and transformed into digital clips. A sitting person was framed half-length, in front of the camera, only the chest and the arms resting on a table were visible. The starting position for each action was: elbows rested on the table and arms still in extended position, open hands side by side resting downwards (fig. 1). For actions to be presented along with the corresponding verbal label, a male voice (realised by using a vocal synthesis program) uttered the corresponding sentence/word in the videoclip soundtrack. The voice started at the same time when the movement started. For the first language training, as explained later, the voice uttered the sentence/word twice, while the corresponding written form was presented in the centre of a white screen. In subsequent action videoclips, the verbal stimulus was only in aural form in order to ensure that visual attention was directed to the movements. Each action in videoclips ended back at the starting position.

Procedure

1. Pre-training. In this phase, the 9 target actions belonging to set 1 were used. Before engaging in the main tasks, participants in groups (a) and (b) had to be familiarised with the artificial

Table 1: The conceptual universe (stimulus set inside brackets, *cue-movements in italic*).

	forearm DIN*outwards*			hand SOD*inwards*		
	left FIT *up*	right NUV *down*	both POC *side*	left FIT *up*	right NUV *down*	both POC *side*
tap GAB *open*	(1) LIBAC	(1) SODEB	(2)	(1) CARUM	(2)	(2)
raise REC *fist*	(2)	(1) BIREN	(2)	(1) GAZEC	(2)	(1) DENAL
wave LOF *point*	(1) TANEG	(2)	(1) MAFIR	(2)	(2)	(1) NIDAP

Figure 1: The starting position of movements

language as a pure sequence of sounds associated, for more clarity, with the written form. They were told they had to learn some sentences/words from an artificial language. They had to look at each sentence/word, and had to repeat it loud. Each stimulus duration was 5 sec, the set was presented in alphabetical order and repeated three times for group (b). This pre-training had the goal of avoiding that much attention be spent just in phonetic decoding in subsequent tasks.

In the pre-training second part subjects were made familiar also with movements; after presentation of each movement subjects were asked to simply repeat it; it was made clear that they should exactly imitate the complete movement specularly, as in front of a mirror. Participants in group (c) were orderly shown and had to repeat the entire sets of cue and target actions; no break separated the two sets. Movements were always repeated in case of error at this stage, until a perfect execution.

2. Basic learning and test cycles. This phase consisted of a series of cycles of Serial Learning, Interactive Learning, Basic Test. In Serial Learning, the 9 target actions belonging to set 1 were presented, in such a order that each movement differed from the previous just for one aspect. Participants in groups (a) and (b) were instructed to watch movie clips and to repeat movements and words, in group (c) they had to repeat the sequence of cue and target movements. In case of mistake, subjects were corrected and the stimulus was repeated. A bell sound highlighted good or bad performance in order to enforce learning. In the Interactive Learning, a screen with 9 buttons in the upper part and a movie window in the lower part was shown. Buttons were arranged in 3 x 3 array labelled with sentences (a), words (b), or icons depicting the start ing position of the cue movements (c). The position of buttons was fixed and the three movements were sorted by columns. Participants could observe target actions in the movie window by clicking with the mouse the corresponding cue-buttons, as many times they wished and in any order. There was no time limit, they were instructed to click the end button when they thought they had learnt. This step had the purpose of allowing participants learning at their pace and following their personal strategies. Upon exit, a random test was performed (Basic Test). If, at the test that ended a cycle, a participant did not reach a level of performance success of 67 % (6 correct answers out of 9), a new cycle was started, up to 3 cycles. In the first Serial Learning actions were presented sorted by movements, in the second were sorted by body-part, in the third by body-side.

3. Inverse test. Participants of all three groups were presented, in a new order, the target actions originally learned (set 1) and had to produce the corresponding sentence, word, or cue-movement. The purpose was to test whether a bidirectional connection had been established. Given the relative difficulty of task and for ensuring a good base for proceeding with subsequent tasks, feedback was given also at this stage.

4. Transfer test. Only participants in group (a) and (c) had this task. It was similar to the previous, but sentences never heard before, or cue-movements never seen before, describing actions of set 2, never seen, were presented. No feedback was given. The purpose was to test whether subjects had represented cues as single features and were able to use them combinatorially.

5. Inverse transfer test. It was similar to the previous, but actions of set 2, never seen, were presented and subjects were asked to give the corresponding cue. No feedback was given. The purpose was again to test the combinatorial nature of their representation.

6. Final test. This task required both a direct and inverse association. In the direct task, participants were given a pair of cues (a sentence, two words, or two cue-movements) and had to execute the corresponding two target movements; in the inverse task, they had to produce the cues from a pair of target movements. For group (a) the sentence included 4 words that expressed the two target movements in a more compact form (e.g. GAB LOF DIN FIT for GAB DIN FIT and LOF DIN FIT), in order to test whether subjects were able to construct higher-order combinations using the basic elements previously learnt. Given that for group (a) this task implied a slight syntactical modification, the first pair was always direct (so that the verbal form or the cue was presented first) and was correspondingly the same in all groups. 6 direct and 6 inverse pairs were presented in total, in random order.

7. Final debriefing. In order to have also a qualitative source of data, upon completion of the procedure, participants were fully debriefed. They were asked what strategies they had adopted in learning and what difficulties they thought had encountered; participants in groups (a) and (b) were also asked to describe in Italian what they thought the meaning of each word was and the movements they had learnt.

Experiment Results

At the final debriefing emerged that, in general, subjects found the task very difficult; they often tried to resort to associations with words, images, and whatever could help from daily life. This normally happens when using meaningless material. A number of participants in groups (a) and (c) had some insight about certain associations, but almost none of them was aware of a clear and systematic framework.

The mean number of learning cycles was 4,55 for group (a), 3,51 for group (b) , and 5,71 for group (c), showing that learning resulted easier for group (b). The mean number of repetitions in the Interactive Learning phase was: (a) 76,69; (b) 133,27; (c) 144,14. This may have happened because in (b) condition subjects presumably spent much more time in search for a rule, or because they exercised more being aware that the best strategy was rote memorisation.

Table 2 shows the proportion of correct responses for each test condition. The overall best performance has been obtained by group (b) subjects. In the direct test (Basic Test, BT) (a) and (b) condition results were comparable (a .70; b .75), while in (c) condition subjects found the task more difficult (.51). In the Inverse Test (IT) subjects in group (b) shown better results than the other groups; the result obtained at this test in (a) and (c) conditions were worse than BT ones. Results in transfer conditions (Transfer Test, TT; Inverse Transfer Test, ITT) were not significant. The final direct test (FT) mirrored the trend shown by the BT.

Experiment Discussion

In the hypothesis that the compositionality of symbols be transferred to the internal representation, a more analytical representation and the best overall performance was expected in condition (a). In this case, the worst performance was expected in condition (b), because one single word for a relatively complex action should be more ambiguous than a featural description (e.g. it could be referring to the movements, or the body part, etc.). In particular, subjects in group (a) should have revealed above chance success in transfer tasks, where a productive operation is involved of re-assembling symbols for describing or executing new actions. In condition (c), if the systematic symbolic association of cue actions worked as a compositional system, we should have obtained results comparable to condition (a).

These predictions have not been confirmed by our results. The main effect seems to support the hypothesis that when a label is symbolically associated to an action as a whole (as in condition b), it can predict better the associated movements than when the symbol was a more complex code. At the first direct execution test (BT) the performance in groups (a) and (b) was comparable (Student's $t=.21$ $p < 0.05$), hence the result for condition (a) cannot be interpreted as being effect of the treatment.

The poorer result for group (c) leads us to suppose that representations corresponding to motor stimuli elicited both by cue and target do not succeeded in connecting well into a symbolic system, but suffered from a sort of reciprocal interference. A testable hypothesis is that if cue and target movements were more obviously related (i.e. target movements had been predicted by other movements, positions by other positions, and the target side by the cue side), then the task would have been much more easier, but in this case because of a motor, not symbolic, connection.

The advantage for group (b) reveals itself both in the direct (execution) and inverse (description) tasks. A possible explanation is that in the other groups the greater number of combinations of cues, compared to a single one, make error more likely just because they did not organise themselves into a true system. The hypothesis may be formulated that, at least in the (a) condition, two mechanisms might have worked concurrently, i.e. (1) a holistic representation in the first stance - similar to the one effective in the group (b), that may have lead to some rote memorisation of composite sentences (or of parts of them) as single arbitrary words, and (2) an initial organisation towards a compositional system. In doing so, a negative factor may have been the greater confusability between different sentences (e.g. GAB DIN FIT vs. GAB DIN NUV) whereas single words were of course well distinctive. Only when a true and full systematisation has been reached the benefits of compositionality may start to be effective.

In the Inverse Test the proportion of correct verbal descriptions of the seen movement has been sensibly poorer than the BT result in condition (a) ($t=.00000071$ $p < 0.05$) but not in condition (b) ($t=.11$). This may be due to a well-known effect in language learning, i.e. that description is more difficult than execution (competence comes before production). If actions were represented as a whole also in group (a), it is also possible that this hindered the composition of a sentence, whereas this did not happen in condition (b), where it's a matter of a simple bidirectional global link. This happened also in condition (c), where the cue-movement should have been retrieved given the target.

Results for The Transfer Test and the Inverse Transfer Test do not reach statistical significance and are very poor for group (c). This makes clear that, at least in the context of the present research, subjects did not construct a coherent representation system for component features of presented actions, the only case that might have helped in transferring to new stimuli the acquired knowledge. However, this happened in part in condition (a) and it probably could have been reached with a more intensive training.

The Final Task confirms the trend of previous tasks. It is more difficult because it includes both direct and inverse tests, and in the case of group (a) subjects had also to learn a new syntax. This task had been originally aimed at testing the capability of scaling-up, assembling symbols at a higher level, but in many cases our subjects rather took it as a simple juxtaposing of previous responses; in a future revision of the present work, we plan to change the target movents for this task and use true composite actions.

Table 2: Mean proportion of correct answers in subjects and neural networks (italic)

	Condition A		Condition B		Condition C	
BT	.70***	*.69****	.75***	*.71****	.51**	*.70****
IT	.39*	*.48***	.68***	*.61***	.34*	*.54***
TT	.24	*.38**			.05	*.51***
ITT	.24	*.28*			.07	*.40**
FT	.25**	*.30**	.42**	*.38**	.22	*.34**

Significance values (binomial distribution)
***p < .0001, **p< .01, *p < .05

Simulation Method

Network architecture

In order to execute the simulations of conditions carried out with the subjects we chose to use a single feed-forward architecture built on 3 layers. An overall scheme of neural networks architecture is displayed in figure 2, where arrows represents a full connection between linked sets. Nets are equipped with 3 different types of input set: 1) a visual one for not-

compositional stimuli consisting of 21 binary units; 2) a second visual input for compositional stimuli composed of 3 sets, each consisting of 2 binary units (each set represents one of the compositional elements that constitute the cue stimuli of group (c)); 3) a verbal input composed of 3 sets each containing 3 binary units (every set is representing one of the syllabic words of group A; moreover the first 2 sets are able to represent the bisyllabic words of group (b). Input 2 and 3 have been duplicated to emulate the sequence demanded for the development of the simulation of final test. Verbal output is a specular reproduction of verbal input, while the motor output is composed of 20 binary units so distributed: 12 units reproduce single target movements (the 6 movements executed with both sides are represented by the activation of the respective nodes of both sides) while the other 8 units reproduce the 8 cue movements necessary for the descriptive purpose of the inverse tests. Likewise input units, it was necessary to duplicate all output units in order to emulate the sequence required for the final test.

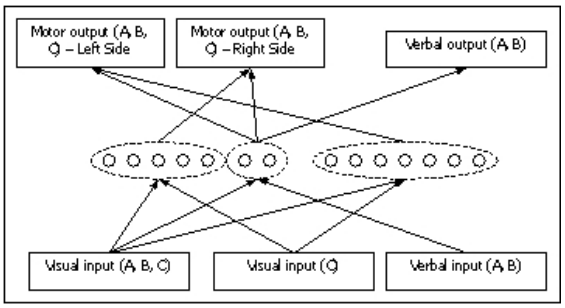

Figure 2: Network architecture

The hidden layer is composed of 14 nodes subdivided in 3 groups. A group is equipped of only 2 nodes, its task is to correlate visual inputs with verbal ones and it is employed in the simulations for groups (a) and (b). The other two groups are composed respectively of 5 and 7 units, which have to process visual stimuli for imitation tasks. The group composed of 5 units controls the reproduction of movements to be carried out with the right side, the other one controls those to be executed with the left side. This distinction between the nodes that govern the motor stimuli is necessary because of the choice to represent cue motor stimuli (group c) by only right hand and forearm. The difference between the set that associates verbal stimuli with motor stimuli and the one that associates motor stimuli to other motor stimuli is due to the assumption that this last task would have to be more difficult for subjects than other ones.

Procedure

The learning procedure strictly followed the same steps considered in the experimental section. Nets were supervisioned by a back-propagation algorithm with learning rate fixed to 0.60 and a momentum of 0.05. Networks used in the same condition only differ by initial synaptic weights, randomly generated in order to make each net evolve autonomously.

The same stimuli proposed to human participants were opportunely encoded and input to nets of the corresponding condition. Verbal stimuli and responses of group (a) were encoded by 4 bits strings while verbal stimuli of group (b) were encoded by 8 bits strings (for example "DIN" is encoded as "1011" and "LIBAC" as "10110010"). Not-compositional visual stimuli were encoded by 21 bits strings while every element of compositional ones (only group (c)) were encoded by 2 bits strings. Motor output did not use any code as explained in the previous paragraph. The only difference between experiment with subjects and simulation concerns the final task, because it was necessary a short final training (only 3 composed stimuli) in order to teach nets the correlation between stimuli required by the final test.

Simulation Results

As for human participants, we expected also in the model that if the compositionality of symbols is transferred to the internal representation, a more analytical representation should be formed in condition (a), respect to other conditions, with separate representations for semantic component features like forearm, to tap, etc. As in human subjects, results (Table 2) show a better performance in the (b) condition; this obviously happened because the pattern was not composite so a linear association could be established. We can hypothesise that this was exactly the reason why we had a similar result in human participants, namely a sort of cognitive economy, versus the advantage of a more costly analytic system that perhaps can reveal itself only after a longer training. A cluster analysis, performed on nets in condition (a), made clear that a structured representation of verbal elements emerged.

Nets showed a clear-cut difference from subjects only in condition (c), especially in transfer tests. This group of nets showed both descriptive and execution capabilities because more resources were available that could allow them to extract features. This makes us suppose that presumably a more prolonged pre-training with human subjects could allow a better feature extraction, which is a fundamental basis for a possible subsequent construction of manipulable representations.

General Discussion

The aim of this paper was to explore the possibility of action symbol systems (ACSS) with the property of compositionality. In two experimental conditions such symbols had a verbal nature, in another condition they were just other actions used as symbols. Verbal symbols were established by means of an artificial language, in one case referring to features and in the other case completely arbitrary. Neural networks may give the opportunity to analyse more closely the kind of representation involved. The process here may be somewhat similar to the one involved in previous simulations of symbol grounding (Cangelosi, Greco & Harnad, 2000; Greco, Riga & Cangelosi, 2003), where the "verbal" part of the architecture assumes a symbolic function because it can transfer its grounding to new perceptual stimuli.

The main questions our paradigm tried to answer were: when actions are learnt by definition through a componential language, capable of expressing featural aspects of component movements, are componential representations correspondingly created? In this case, are such representations used and helpful

in retrieving information about movements to be executed when the same verbal code is used to evoke them? If actions were defined during learning as a whole by a single, arbitrary word, would this code be more efficient? Or, if just movements were associated to establish a modal symbolic system, could it work as an ACSS, that is as a modal system but with compositional properties?

Our results do not corroborate the hypothesis of an immediate creation of componential representations during the acquisition of an artificial language, even if it was componentially structured. The fact that such representations were not available to our subjects in groups (a) and (c) is revealed by the their poor performance in the inverse tests. The best results were obtained only when actions were defined by a single word.

There are at least two possible explanations for this result. One could be that the hypothesis that actions are represented componentially is simply to reject. There is, however, another possibility. This result can probably also be due to the fact that the productive aspect of language has a categorical basis, namely it is strictly connected to the discovery of relevant features as an essential premise for any subsequent featural and not holistic representation, and this may have been a supplementary burden that made the task more difficult than expected, at least in condition (a). We had been confident that the systematic organisation of our stimuli, whose structure followed the natural language syntax, and that also could be inferred from the button arrangement on the screen during the Interactive Learning phase, could allow such discovery. Evidently, this did not happen to a full extent, even if some more salient or distinctive movements, like to wave hands, were quickly associated with corresponding cues; the true problem was to detect more subtle or less evident differences, like the one between hand and forearm movements.

We must also take into account that holistic association might have been the most economical strategy in the context of our experiment, where tasks were not goal-directed. We are aware of the limits (also from an ecological point of view) of trying to reproduce the early stages of symbol acquisition using meaningless material with adults subjects, because, among many differences with infants, they may resort to strategies for connecting meaningless stimuli with meaningful representations (images, associations, etc.). We know from the final interviews that our subjects did it. The use of nets is paradigmatic then to let us see the difference between what a cognitive system can do "in the vacuum": paradoxically, nets can achieve more easily a clearly structured representation by extracting featural regularities without interference from other associations.

As to the issue of modal nature of action representation, we believe that symbols help in making distinctions; they are used when they are needed to accomplish this function. The issue is not to state definitively whether they are modal or not, analogue or not, but if they are consistently mapped to what they are assumed to represent.

Future developments. The paradigm we have presented is only at an early stage of development. Many improvements may increase its usefulness for more accurate theoretical accounts. A more extensive and accurate pre-training could reduce the cognitive overload of feature extraction, by making categorization and naming tasks more separate. Truly composite actions should be used in the final task, so that combination and not simple module juxtaposition be required. A fourth group with a treatment similar to condition (c) but where target movements be associated with single, not compositional cue-movements could mirror the condition of group (b). As to neural nets, the task would be more realistic (and less straightforward in some cases) if a stimulus categorization were required (but this would be possible only by adopting more sophisticated visual inputs, like from a simplified retina). The temporal sequence should also be better controlled, by introducing different forms of learning, also by self-organization.

Acknowledgments

The authors wish to thank Angelo Cangelosi and Thomas Riga for discussing some aspects of this work, and Marcello Mistrangelo, Monica Corte, Ramona Rivano for help in performing the experiments.

References

Barsalou, L.W. (1999). Perceptual symbols systems. *Behavioral & Brain Sciences, 22,* 577-660.

Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes, 12,* 177–210.

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,* 2, 143-162.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47,* 30-49.

Glenberg, A. M. & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin, & Review, 9,* 558-565.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language, 43,* 379-401.

Greco, A., Riga, T., Cangelosi, A. (2003). The acquisition of new categories through grounded symbols: An extended connectionist model. In O. Kaynak, E. Alpaydin, E. Oja & L. Xu (Eds.). *Artificial Neural Networks and Neural Information Processing* - ICANN/ICONIP 2003. Berlin: Springer.

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

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Luria, A.R. (1960). *The Role of Speech in the Regulation of Normal and Abnormal Behavior*. New York: Pergamon Press.

Luria, A. R. (1981). *Language and Cognition*. Ed. and translated by J. V. Wertsch. New York: Wiley.

Piaget, J. (1952). *The origins of Intelligence in Children*. New York: International Universities Press.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2,* 661-670.

Vigliocco, G., Vinson, D.P., Lewis, W. & Garrett, M.F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology, 48,* 422–488.