frontiers in
NEUROROBOTICS

ORIGINAL RESEARCH ARTICLE
published: 11 November 2010
doi: 10.3389/fnbot.2010.00111

Compositional symbol grounding for motor patterns

Alberto Greco* and Claudio Caneva

Laboratory of Psychology and Cognitive Sciences, Department of Anthropological Sciences, University of Genova, Genova, Italy

Edited by:
Angelo Cangelosi, University of Plymouth, UK

Reviewed by:
Claudia Scorolli, University of Bologna, Italy

****Correspondence:***
Alberto Greco, Department of Anthropological Sciences, University of Genova, Via Balbi 4, 16126 Genova, Italy.
e-mail: greco@unige.it

We developed a new experimental and simulative paradigm to study the establishing of compositional grounded representations for motor patterns. Participants learned to associate non-sense arm motor patterns, performed in three different hand postures, with non-sense words. There were two group conditions: in the first (compositional), each pattern was associated with a two-word (verb–adverb) sentence; in the second (holistic), each same pattern was associated with a unique word. Two experiments were performed. In the first, motor pattern recognition and naming were tested in the two conditions. Results showed that verbal compositionality had no role in recognition and that the main source of confusability in this task came from discriminating hand postures. As the naming task resulted too difficult, some changes in the learning procedure were implemented in the second experiment. In this experiment, the compositional group achieved better results in naming motor patterns especially for patterns where hand postures discrimination was relevant. In order to ascertain the differential effect, upon this result, of memory load and of systematic grounding, neural network simulations were also made. After a basic simulation that worked as a good model of subjects performance, in following simulations the number of stimuli (motor patterns and words) was increased and the systematic association between words and patterns was disrupted, while keeping the same number of words and syntax. Results showed that in both conditions the advantage for the compositional condition significantly increased. These simulations showed that the advantage for this condition may be more related to the systematicity rather than to the mere informational gain. All results are discussed in connection to the possible support of the hypothesis of a compositional motor representation and toward a more precise explanation of the factors that make compositional representations working.

Keywords: motor representation, symbol grounding, compositionality, embodiment

INTRODUCTION

Compositionality and symbol grounding are two fundamental questions that have gained considerable theoretical attention in the last decades. Compositionality consists in the possibility of drawing the meaning of a complex linguistic expression from the systematic combination of meaningful components according to syntactical rules. It is considered one of the key features of human language, differently from animal communication or human ancestor protolanguage, fundamentally holistic and conveying meaning only through single gestaltic expressions (Jeannerod, 1988; Arbib, 2005). Compositionality has been called into play for explaining the ability of producing an indefinite number of linguistic expressions (what is known as productivity), and is relevant in formal languages of mathematics, logic, and computer science. The principle of compositionality, in fact, is a general key concept in all the cognitive sciences, since it has gained interest in philosophy, linguistics, artificial intelligence, robotics, psychology, and neuroscience.

As is well known, compositionality was an essential part of the traditional cognitivist "language of thought" hypothesis (Fodor and Pylyshyn, 1988), positing that human representations acquire their structure by the combination of distinct symbolic parts according to formal rules. This view was first challenged by connectionist theories (Smolensky, 1988; van Gelder, 1990) and more recently by new approaches that accept the idea of non-symbolic representations. These approaches stress the point that cognition cannot be explained solely by abstract symbolic processing, because human beings have a body interacting with environment (*embodiment*: e.g., Glenberg and Kaschak, 2002), and because a sensorimotor ground is needed for symbols. This is the symbol grounding issue (Harnad, 1990; Cangelosi et al., 2000).

Such new stances have influenced also the way of considering language. The question of how actions are internally represented is of general importance because words for action (predicates or verbs) are the essential ingredients of propositions, and actions are also fundamental for understanding, like predicates are essential in logic. In addition, representation of actions and of words could be tightly linked since, according to some theories, linguistic comprehension would be a sort of internal simulation (*re-enactment*) of actions expressed by linguistic symbols (Barsalou, 1999; Pulvermueller, 2005). Many other recent approaches have made similar points, like the "experiential view of language comprehension" (Zwaan, 2004). In the same vein is the finding that motor verbs activate brain regions associated with action (Ruschemeyer et al., 2007). Barsalou comes to considering perceptual non-symbolic representations as a system having the same features of symbolic ones, including *compositionality*. In this sense, Barsalou's approach implies supposing analog representations working compositionally (Wu and Barsalou, 2009).

The motivation for the present study then comes from the fact that, although compositionality has been traditionally considered as concerning the abstract combination of symbols that already must have a grounded meaning, the possibility of an analogical compositionality, and in particular of a motor compositionality, is a still open empirical question.

The hypothesis of a motor compositionality has obtained a substantial interest in current cognitive neuroscience research (Bizzi and Mussa-Ivaldi, 2004, p.415). There are several reasons for hypothesizing compositional motor representations: human motor control has a hierarchical nature, complex motor programs result from motor subroutines, elementary operation of body parts (i.e., joints, muscles, etc.) for action can be identified (Allott, 2003). In robotics, such a system has also obtained significant attention (e.g., Thoroughman and Shadmehr, 2000; Amit and Mataric, 2002; or the "Human Activity Language" primitives for segmenting human motor patterns as a language: Guerra-Filho and Aloimonos, 2006). The theoretical relevance of this issue is clear also since a compositional motor representation would entail that motor primitive elements could be distinguished that keep the same meaning in different contexts, like their possible verbal counterparts.

Some additional clarification seems convenient here about the expression "motor representation." It is obviously possible to consider either symbolic (conceptual, verbal) or analog motor representations; grounding is, of course, just the establishing of an association between these two kinds of representations. But the notion of *analog* motor representations seems to oscillate between psychological and neural senses (Greco, 1995; Peschl, 1997), ambiguously referring to different processes such as: (a) preparing motor action: motor schemata or motor imagery (Jeannerod, 1994; see also the symposium "Mental representations of motor acts" of the European Neurosciences Association: Deecke, 1996); (b) kinesthetic self-perception of motor action during execution; (c) visuospatial perception of motor action executed by others. Such senses evidently refer to different motor tasks that may be related to a more basic distinction between visuospatial and motor aspects (respectively implying perception and execution of motor patterns). The strength of this distinction, however, seems weakened by the celebrated and well-established mirror neuron theory, showing that perception and execution of motor patterns activate the same brain areas (Gallese et al., 1996). The mirror neuron hypothesis is compatible with the assumption that, even if evidence can be found that motor tasks are controlled by different systems at lower levels, at some higher level they should converge into a unique representation. This unique representation is responsible for the uniqueness of meaning, the one that normally is expressed verbally (e.g., when we speak of "walking" we mean the same thing either referring to what we *see* when someone else is walking or what we ourselves *do* when walking).

In any case, whatever the exact nature of analog motor representations is (as a form of imagery, or of mental simulation, or re-enactment), the point is how structured these representations are. Do they include primitive "images" for components of motor performance, or codes for individual features, that are then somehow assembled, or do they work as a whole? The question is relevant also for motor concepts and words that are associated to motor memories.

FRAMEWORK

The present study was aimed at an empirical investigation about the nature, compositional or holistic, of motor representations that provide analog ground for meaningless verbal labels.

The most obvious and ecological way of analyzing the relation between language and motor behavior is considering when a *meaningful* association is established. This is obvious because motor activities are normally goal-directed, and meaningful words are used to describe them. We choose, however, to start from *meaningless* words and motor patterns, a rather extreme situation, because when studying the establishing of symbol grounding the interference of already-known motor patterns and words should be minimized. We needed to study how *new* symbols are associated and eventually combined for representing *new* motor patterns, eventually becoming meaningful. Thus we used non-sense words as arbitrary symbols that would acquire a meaning only (or as much as possible) from grounded sensory experience, namely in connection with perceived visuomotor stimuli. Similarly, we used non-sense motor patterns because if they already had a sense they would also had been already connected with a corresponding linguistic representation and the new word would only consist in a sort of "translation" or a synonym of this existing representation. We actually use the term "motor patterns" and not "gestures" just to stress that we are referring to meaningless motor behavior. We are obviously aware of limits of this perspective, since any stimulus (either verbal or not) is normally put in relation with semantic memory contents; this situation of artificial "semantic vacuum," however, seemed suitable as a starting condition for a study of symbol grounding establishment.

The present work continues a previous one (Greco and Caneva, 2005) where we already associated an artificial language with meaningless motor patterns in holistic and compositional conditions. In the experimental paradigm described in the present paper, there were two conditions. In the first condition one word acquired a grounding for an arm trajectory (irrespective of how it was executed) and a second word was grounded for denoting a particular way of executing it (how to put hands while executing it). In the second condition a single word was grounded for each motor pattern execution taken as a whole.

The main hypothesis tested was that when different verbal labels are learned in association with different aspects of visuomotor patterns in arm motor patterns (namely, in our case, arm trajectory and hand posture), a separate grounding is established for these symbols, based on compositional analog representations, that allows a facilitation in a subsequent naming task for the same patterns.

The rationale is that the ability of correctly naming visuomotor patterns, in our experimental conditions, is a true grounding test (Cangelosi et al., 2000), because this would reveal that labels, that were meaningless at the start, became meaningful symbols for these patterns as a result of an analog grounding. This kind of grounding may be ascribed an analog nature even if it does not necessarily involve really performed motor patterns. This idea is supported by the mirror neuron theory, that strengthens the idea that analogic patterns can be established on observed visuomotor patterns without a direct bodily execution.

If participants in the compositional condition were favored in this task, then, this outcome would show that a separate analog grounding was established for arm trajectory and hand posture, connected with the corresponding two labels. On the contrary, if patterns tended to be better represented by analog holistic codes, a naming task in the condition where each pattern as a whole was learned in association with a single word should be advantaged.

A further account for a possible advantage resulting in compositional condition is that memory load is reduced when the amount of information needed to name stimuli is smaller, as in the case when some words can be reused for recalling the same motor referents. However, not only informational load but also a reliable grounding system must be taken into account in this case: this involves a consistent association between symbols and their analog referents. We shall tackle this question with the help of neural network simulations.

We addressed also the question whether the visuospatial analog coding, on which recognition is based, might be affected by grounding as well. In fact, it is reasonable to suppose that naming implies first some pattern recognition process and after that – if grounding has been established – the retrieval of the corresponding label. We tested this possibility by introducing in our first experiment also a recognition test, in order to assess a possible difference between compositional and holistic groups.

Table 1 | Stimuli.

Basic motor patterns	Hands up	Hands down	Fist
	Nole	*Bote*	*Sove*
Baspi	Terpesova	*PD	Utrimosta
Gispi	Sertamina	Mutiralda	*PD
Respi	Tupifasta	*PD	Mertogala
Tispi	Volsicoda	Feltorana	*PD
Faspi	Patrasina	*PD	Luticanza
Cuspi	Rispaguna	Dortamana	*PD
(mov.#7)	*TD		
(mov.#8)	*TD		
(mov.#9)	*TD		
(mov.#10)	*TD		

**TD indicates trajectory distractors, *PD posture distractors.*

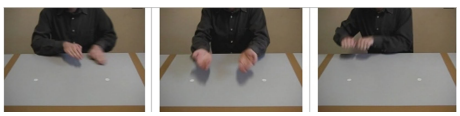

FIGURE 1 | Snapshots from some videoclips of different patterns in the three hand postures.

EXPERIMENT 1

METHOD

The task consisted in associating visuomotor patterns, presented as videoclips, with corresponding words, uttered aloud. There were two conditions: in the compositional condition (group A) motor patterns were associated with *two words*, whereas in the holistic condition (group B) with a *single word*. The two-word sentence presented in the compositional condition can be considered as a "verb–adverb" structure: what motor pattern is performed, how it is performed (i.e., using what posture). In this experiment a recognition test was performed prior to the naming test. The dependent variables were: (a) *recognition* of target motor patterns presented along with distractors; (b) *naming* (retrieving the name corresponding to each target motor pattern).

Stimuli

The structure of stimuli is shown in **Table 1**; some examples are given in **Figure 1**.

Motor stimuli. Consisted in arbitrary arm trajectories (as an example: moving arms toward oneself and then lifting them). Eighteen stimuli were constructed by combining six basic motor patterns, performed in three different hand postures (up, down, fist); four other motor patterns were added, performed in the hand up (called "nole") posture only. All motor patterns were performed by a sitting person, framed half-length, in front of the camera; only the chest and the arms were visible; in the starting position the hands (already in the palm, back or fist posture) rested on two reference circles marked on the table. Only 12 combinations (the ones with a bold name in **Table 1**) were presented during learning. The other 10, indicated with an asterisk, acted as distractors for recognition testing purpose; 4 of them (*TD) were arm trajectory distractors (corresponding to never seen trajectories), the other 6 (*PD) were hand posture distractors (corresponding to seen postures but performed in a different hand posture).

Linguistic stimuli. For group A, a two-word sentence was used to name patterns, resulting from the combination of the word for the trajectory and the word for hand posture (words for group A are in italic in **Table 1**). For group B, a single word (in bold in **Table 1**) was used to define each pattern as a whole. For example, the first pattern was named "baspi nole" for group A and "terpesova" for group B.

As in natural languages syntactical roles are marked by particular morphemes, some constraints were established for pseudowords that had to assume a syntactical role. The six pseudowords denoting verbs were 5-letter and bisyllabic, constructed by adding a consonant-vowel pattern to a fixed ending (–SPI). Pseudowords denoting adverbs were 4-letter and bisyllabic, constructed by the pattern consonant-O-consonant-E. Single pseudowords standing for full motor patterns had 9-letter and 4 syllables (resulting like the sum of the other two words) and all ended in -A.

Participants

Twenty students, volunteers, individually participated in the experiment for course credit. Informed consent was obtained prior to participation in the study. Half of them were randomly assigned to group A, half to group B.

Procedure

Participants seated in front of a 14″ computer monitor, in a different room than experimenter's; in the table in front of the screen a rectangular area measuring cm 77 × 53, including two reference

circles identical to ones shown in the videoclips, was traced; this allowed participants to repeat motor patterns when requested. Only a mouse (no keyboard) was available for responses. All instructions and stimuli were presented on the monitor screen. The procedure included the following stages.

Verbal learning. The first stage was aimed at making participants familiar with words. All the words were presented in a panel with 9 (group A) or 12 (group B) buttons, where each single word was printed as a button label. Labels were disposed in alphabetical order. Participants were instructed to click with the mouse on each button to listen to a recorded male human voice that read the corresponding word aloud; the order of presentation was chosen by participants themselves. Only when all words had been listened, a closing button was enabled to proceed to the next step.

Associative learning. This was the main stage of the experiment. Twelve training clips were presented. For each clip, the voice uttering the sentence (gr. A) or word (gr. B) corresponding to the motor pattern was presented at the start, along with a blank screen; the videoclip was then shown immediately. Patterns were presented randomly but paired so that the same pattern was first presented in the "nole" (hands up) posture and then in one of the other two postures, like shown in **Table 1**. Participants were also instructed to repeat each pattern after having seen it while uttering its name aloud, in order to learn it better. It was stressed that the correctness of this performance would have not been assessed in any way. The full set of stimuli was repeated three times.

Integrated test. In the testing phase, recognition test and naming test were integrated. All 22 stimuli clips (12 target and 10 distractors) were presented in random order. For each stimulus, participants were first asked if they had already seen it; if they answered yes, then they were also asked to say the corresponding sentence/name. Motor performance was not requested. A final debriefing was conducted in order to assess possible task difficulties and hints for improvement.

Post-experimental debriefing. After completion of the experiment, a structured interview was conducted in order to assess task difficulty, the use of associations with known words or gestures, and above all to verify whether participants in group A had been able to identify the syntactic role of the two words. Almost all participants found the task difficult or very difficult, but the syntactical roles were identified without uncertainty by participants in group A, with the exception of only two subjects. Associations reported by participants were somewhat subjective and not consistently related to particular stimuli.

RESULTS AND DISCUSSION

Recognition test

Very high recognition scores resulted without any difference in both groups (condition A, $M = 0.81$, SD = 0.39; condition B, $M = 0.82$, SD = 0.38). This outcome shows that motor recognition, at least in our experimental conditions, is not related with the availability of a specific verbal label for components. Motor patterns were presumably not recognized using a verbal code but accessing to a specific visuomotor representation.

We also analyzed recognition scores for distractors only (**Table 2**, PD = posture distractors, TD = trajectory distractors). Recognition was almost fully correct ($M = 0.92$) for MD, i.e., different trajectories, but recognition scores were lower ($M = 0.71$) for PD, i.e., same trajectories with different hand postures. This difference is highly significant ($t = -4.41$, $p < 0.0001$) and depends on the fact that differences between motor patterns resulted very salient, whereas it was more difficult to distinguish hand postures. This result shows that, in a pure recognition test, motor stimuli were not processed at the hand posture detail level, characterized by more confusability, but only at the motor pattern level, more macroscopic, where a more immediate holistic representation seems sufficient for recognition. Retrieval in this case was based on perceptual similarities and not on the symbolic association with arbitrary labels.

Naming test

Naming task results were completely opposite to recognition ones, as very low scores resulted in both groups (condition A, $M = 0.16$, SD = 0.37; condition B, $M = 0.17$, SD = 0.37).

A difference between recognition and naming in our task is not surprising, because it is consistent with the well-established finding that performance is generally better in recognition memory than in retrieval memory, and that these are based on substantially different processes (Yonelinas, 2002). This difference holds in many areas of cognition, from words (Peynircioglu, 1990), to pictures (Langley et al., 2008), to faces (Cleary and Specker, 2007), to melodies (Kostic and Cleary, 2009). This effect was found also with pseudowords and even non-words (Arndt et al., 2008). Our result matches such theoretical premises, and seems to suggest that the recognition-retrieval difference could be extended also to motor memory. The dramatic extent of this difference in our task, however, suggests some caution in reaching this conclusion. Our outcome evidently indicates that name-pattern association was too a difficult learning task in these conditions and this could have amplified the recognition-naming difference. This issue would have deserved a deeper investigation in different learning conditions. We strived, in the course of our study, to remedy such learning difficulties, but, since the recognition-retrieval issue was not the main concern of our current research, this result was not further analyzed and the recognition task was abandoned.

EXPERIMENT 2

The main outstanding question from results of Experiment 1 was the floor effect we found for naming, clearly denoting that learning conditions were inadequate for grounding. This

Table 2 | Mean recognition proportion for distractors and target stimuli in Experiment 1.

	Distractor type							
	PD		TD		Target		Total	
Group	*M*	SD	*M*	SD	*M*	SD	*M*	SD
A	0.74	0.44	0.91	0.29	0.87	0.34	0.81	0.39
B	0.66	0.48	0.94	0.24	0.88	0.32	0.82	0.38
Total	0.71	0.46	0.92	0.27	0.88	0.32	0.82	0.39

motivated a revision of experimental setup in order to make learning easier. We must make clear that our interest is currently focused on differences between compositional and holistic conditions in comparable learning conditions, sufficiently adjusted as to difficulty, and not on learning conditions or mechanisms *per se*.

A new paradigm for Experiment 2 was then planned. In order to make learning easier, method and procedure were simplified. Instructions were improved by introducing an interactive example of task execution. A different stimuli presentation system was also adopted: in the first learning stage, all patterns were presented only in a single hand posture (upwards); in a second learning stage, after having tested that at least four of six stimuli had been learned, the same trajectories were paired with a second posture. As a further change, it was required that verbal stimuli be transformed into an infinitive verb, by adding the (Italian) ending "-are" (e.g., "baspi" into "baspare"). This helps categorizing such words as verbs reducing the cognitive load. An additional reason that motivated this change was that the task resulted rather passive, since names were still in echoic memory when repeated just after having being heard. This change was then aimed also at encouraging an active stimulus processing, so that echoic memory effect be removed or reduced, and participants be less passive and more attentive.

METHOD

The independent variables and the main task (i.e., learning to associate motor patterns with sentences or words) were the same as in Experiment 1. Naming was the only dependent variable.

Stimuli

The conceptual universe was the same as in Experiment 1 (**Table 1**), but only 12 target clips were used (no distractors were needed since no recognition test was performed).

Participants

28 students, volunteers, individually participated in the experiment for course credit; informed consent was obtained prior to participation in the study. Half of them were randomly assigned to group A, half to group B. As in Experiment 1, in group A each motor pattern was associated with a two-word sentence (one for trajectory, one for hand posture), in group B each motor pattern performance as a whole, i.e., regardless of hand posture, was associated with only one word.

Procedure

Instructions and stimuli were presented in the same conditions as in Experiment 1. Learning was split up into two stages. In the first phase (*target pattern learning*) only six patterns in the "hands up" posture were learned. This stage was followed by a first test (*target pattern test*, TPT). In the next learning stage (*posture learning*), each learned pattern was paired with the same pattern in a different hand posture. The procedure included the following stages.

Verbal learning. This stage was exactly as in Experiment 1. The same word panel was used; it included the full set of words for the group (9 A, 12 B), arranged in alphabetical order.

Target pattern learning. The purpose of this stage was to make participants learn motor patterns irrespective of hand postures. As in Experiment 1, clips started with a blank screen while a male human voice uttered the corresponding word, then the motor pattern performance was shown on the screen. Only six clips were presented, in random order, and only one word, referring to the pattern, was used. The only difference between groups A and B was the word used (e.g., "baspi" for gr. A and "terpesova" for gr. B). Subsequently, the word panel was shown, where word labels appeared, transformed into the infinitive form (e.g., "baspare" or "terpesovare"), and the participant was requested to mouseclick the corresponding button. It was possible to correct mistaken choices before confirming. Then, the pattern was shown again without audio and the participant had the opportunity of performing it while uttering the verb aloud. In instructions it had been explained that the purpose of this procedure was to help participants "learn better" motor patterns; it had been also stated clearly the absolute irrelevance of correct performance. The series of six stimuli was repeated three times.

Target pattern test. At this stage, learning of six previously presented names was tested. A minimum learning threshold of 4/6 was required for passing this test, otherwise the first learning stage was repeated (up to two times, after that the protocol was discarded).

Posture learning. After a new warm-up example trial, at this stage all 12 stimuli, in different hand postures, were presented with the same procedure as in Target Pattern Learning. For participants in group A, motor patterns were described using a sentence where the first word was the same word for the trajectory previously learned, and the second was the word for the posture (e.g., "baspi nole," "baspi sove"); for group B the word uniquely denoting the motor pattern was used (e.g., "terpesova," "utrimosta"). Participants in group A could compose the corresponding sentence by clicking on two-word buttons (in group B just one button); all could correct mistakes before confirming. In the word panel all words denoting motor patterns were put into the infinitive Italian form (e.g., "baspare," "terpesovare") and this was the form that participants had to use when repeating aloud the verbal part. The presentation sequence was random, but, to make learning easier, motor patterns were paired so that each randomly selected trajectory was always followed by the same trajectory performed in the other scheduled posture. As in the previous Target Pattern Learning, the full set of stimuli was repeated three times, so that 36 stimuli were presented overall.

Final test. All 12 videoclips showing motor patterns without audio were randomly presented, each followed by the word panel. Participants in group A were requested to click on two words to compose the corresponding sentence; in group B they had just to click on the corresponding word. It was always possible to correct mistakes before confirming.

Post-experimental debriefing. A final debriefing was conducted following the same procedure used for Experiment 1. The task was still perceived as difficult but, as in the previous experiment, the syntactical role of words in group A was easily identified by

all participants (only one failed). Very few verbal or visuomotor associations were reported, that were not commonly shared but rather had a personal character. In any case, there is no reason to suppose that particular associations could favor one group over the other.

RESULTS AND DISCUSSION

We first analyzed learning progress in different experimental stages. **Figure 2** shows the learning curve (mean proportion of correct responses) from the first to the final phase. At the first TPT there were no significant differences between the two groups ($A = 0.47$, SD = 0.50; $B = 0.41$, SD = 0.49; $t = 0.82$, $p = 0.41$). This shows that there were no differences between subjects at the start and, importantly, that stimuli used for the two groups were equivalent. Mean values of correct responses at the final test (FT), instead, were significantly different ($A = 0.60$, SD = 0.49; $B = 0.46$, SD = 0.50; $t = 2.51$, $p = 0.01$).

As it is clear from our experimental set up, two kinds of stimuli were tested in the FT phase, i.e., motor patterns presented in both Target Pattern and Posture Learning phases (all with hands up posture) and motor patterns only presented in the latter, differing from previous ones because they had different hand postures. It seems obvious to expect that motor patterns seen in both learning phases (hands up posture) are considerably easier than others; in fact, there is no difference between groups for such stimuli learned during both training phases (see **Table 3**, 0.68 vs.0.62, $t = 0.83$, $p = 0.41$). If we consider other motor patterns, however, the difference between the two groups is dramatic ($A = 0.51$, $B = 0.29$) and statistically highly significant ($t = 2.83$, $p = 0.005$). Since new stimuli differed from previous ones only for hand posture, this supports the hypothesis that the compositional task was easier because a specific word denoting posture was available. We can say, then, that in the present experimental conditions verbal descriptions for motor patterns were better learned when a compositional verbal system was available.

In our experimental setup, in the FT, naming was influenced by having seen a pattern before. The effect of a verbal system could only be revealed by considering trajectories with a new hand posture. In fact, the compositional group (A) had better results than the holistic one (B) in naming new stimuli. If we compare the outcomes of the experiments 1 and 2, we find that there was no compositional representation in recognition (Experiment1) and a sort of compositional representation in naming (Experiment 2). The procedure in Experiment 2 presented two main differences from Experiment 1: (a) having splitted learning of trajectories and of postures; (b) having introduced the addition of the Italian suffix for verbal conjugation ("-are"). The first change may have helped participants identify more easily stimulus features. The second change may have helped group A (where an elementary syntactical system was needed) by giving a hint about the syntactical role of the first word.

Table 3 | Mean correct naming proportion in Experiment 2 (final test).

Group	Stimuli	Correct	
		M	**SD**
A		0.60	0.49
	Hands up posture	0.68	0.47
	Other posture	0.51	0.50
B		0.46	0.50
	Hands up posture	0.62	0.49
	Other posture	0.29	0.46
Total		0.53	0.50

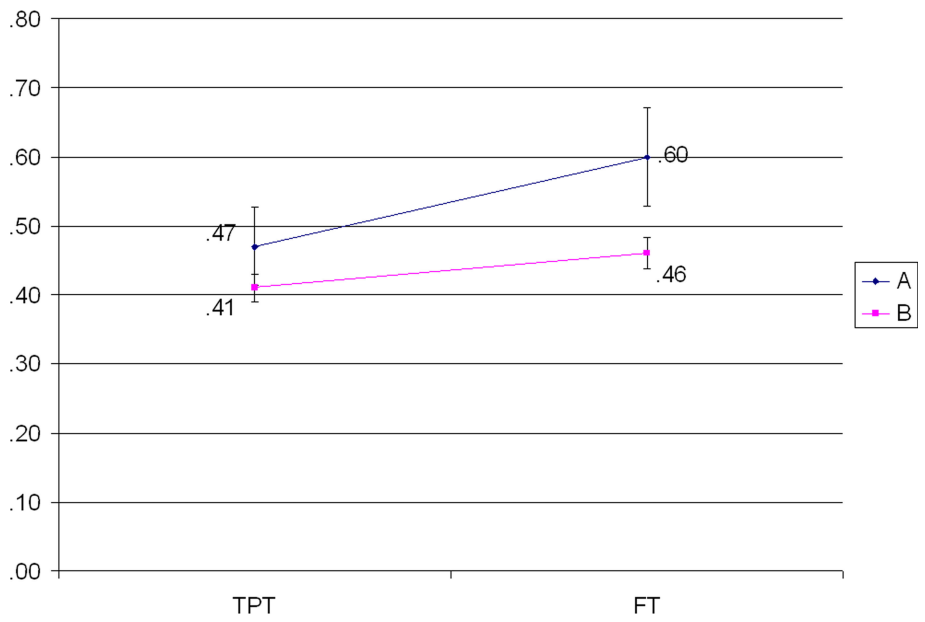

FIGURE 2 | Learning curve in Experiment 2.

NEURAL NETWORK SIMULATIONS

As we have mentioned previously, a possible account for the advantage resulting in the compositional group is that in this condition more motor stimuli can be coded using a fewer number of words. This implies an informational gain, that would be maximally exploited with all theoretically expressible stimuli: using the nine available words in condition A, 18 motor patterns can be named combining six trajectories with three postures (type-token ratio 9/18 = 0.50). Even if in our actual condition only 12 patterns were learnt (type-token ratio = 9/12 = 0.75), anyway a consistent reduced memory load results. In this account, however, two aspects are not clearly distinguished, i.e., the informational-syntactic aspect (i.e., the mere number of alternatives and word positions) and the need for consistent and systematic semantic associations.

In order to test some different possible changes to our paradigm without having to engage a number of new human participants, we reproduced and modified the task using neural network models. Considering that grounding analog information in symbolic codes is tantamount to use more compact representations, we can expect a still greater advantage for the compositional condition respect to a condition where the number of words is equal to the number of stimuli to be distinguished and remembered (type-token ratio = 1.00). To take into account the role of systematic correspondence between words and analog patterns, upon which syntax and grounding are based, we also devised a simulation where such correspondence was disrupted, while maintaining an equivalent informational load.

We then performed three simulations:

- *basic* simulation, that faithfully reproduced the Experiment 2 conditions;
- *extended* simulation, where the basic simulation was augmented using an increased number of inputs;
- *non-systematic* simulation, similar to the extended simulation, where verbal inputs were modified in order to be informationally equivalent to original ones, but without any systematicity.

GENERAL METHOD AND BASIC SIMULATION

General neural network architecture and I/O encoding

For all our simulations we used a set of 50 neural networks that implemented three modules of hidden units, with the function of processing motor and verbal information, and of establishing an associative grounding between the two kinds of data.

The architecture (shown in **Figure 3**) included an input layer, divided into two distinct modules (12 motor units and 9 verbal units), two hidden layers, and only one (verbal) output layer. The first hidden layer was divided into two distinct (not interconnected) modules, each including six units, with the purpose of independently processing motor and verbal inputs. The second hidden layer (including 15 units) had an "associative" function, that is to relate the two kinds of inputs and to generate the output. Each layer was connected in a recurrent way with its lower layer.

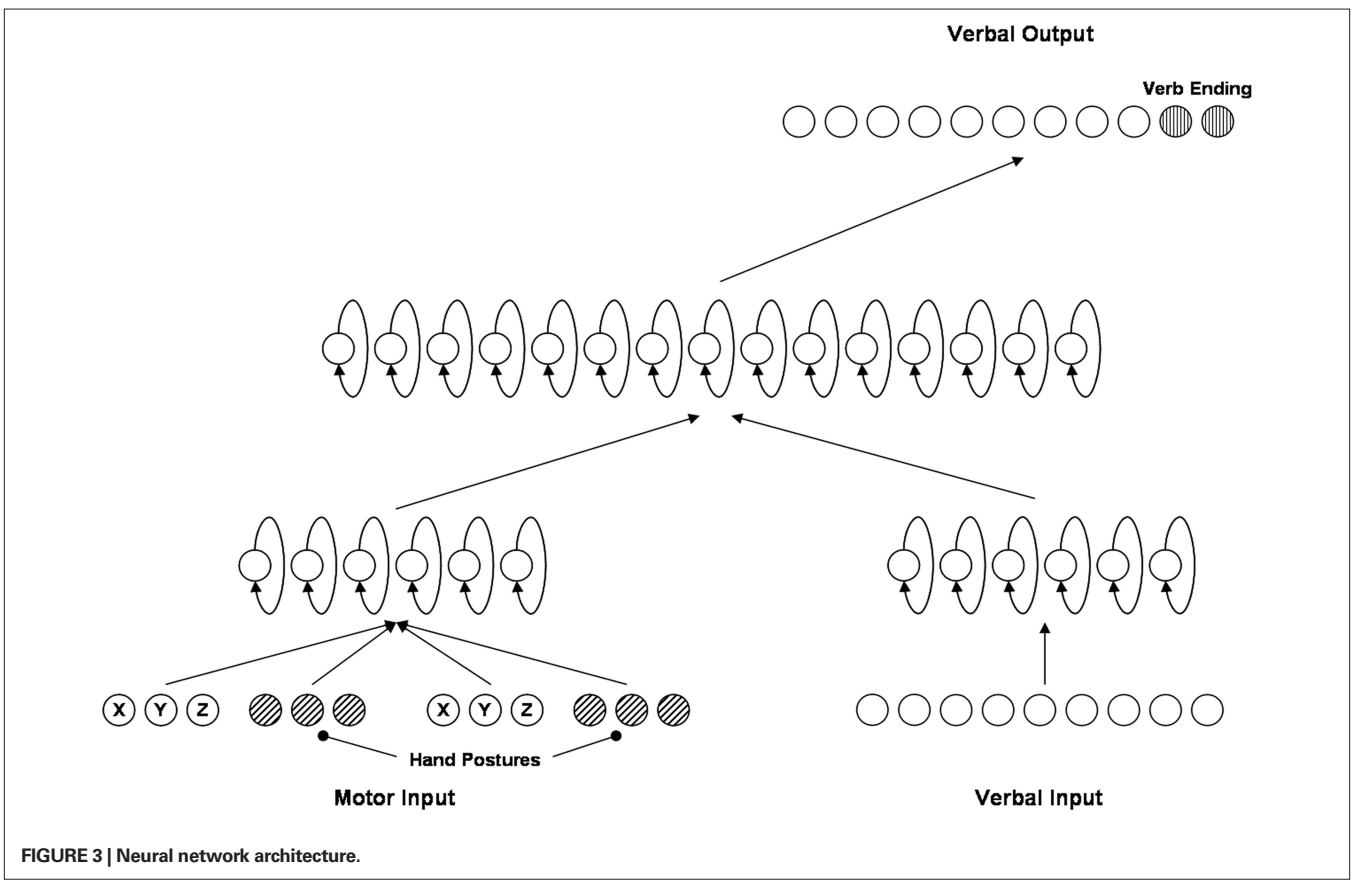

FIGURE 3 | Neural network architecture.

Since motor and verbal data flows had a different length, we introduced a parametric bias to synchronize the data flow. This bias was computed during a pre-training stage by a set of four "timing" units (not shown in figure) supervisioned by a back-propagation algorithm. In this pre-training, networks were given 10 motor pseudo-inputs that had the same streaming structure of real inputs, but that did not represent points fitting on the same curve. During this stage, the supervision algorithm acted uniquely on timing units, while weights of all other connections were not modified. At the simulation start, all timing units were set up the same way in all networks; weights of other units were generated randomly. Parameters of timing units were never modified during simulations. Thus our networks can be considered as discrete-time RNNPB (Recurrent Neural Networks with Parametric Bias) with a dynamic input. During simulation learning was achieved by a supervised Bayesian algorithm using Gibbs sampling.

The basic conceptual universe consisted of a number of inputs equal to the number of stimuli used for the Experiment 2. Verbal and motor input were given as data streamings. Words or sentences were input as a flow of four consecutive strings (corresponding to the 4-syllable verbal inputs, e.g., ba-spi-no-le or ter-pe-so-va). Motor patterns were encoded in a pseudo-analog way, i.e., input as 25 consecutive sets of spatial coordinates of the two hand postures in different time moments (frames) during pattern execution; such coordinates were obtained through the analysis of motor patterns of a virtual dummy (Poser 7.0). Each stream also included information about the hand posture, encoded using three binary units.

General procedure

Each simulation was composed of two sets of 50 nets and they followed the same steps considered in the corresponding experiment. Simulations followed the same steps as in the Experiment 2. In the Basic simulation, the conceptual universe was exactly the same used in Experiment 2 (referred as "standard" in **Figure 5** that summarizes all simulations), including 12 stimuli.

Basic simulation results

Networks learning in both conditions was very close to human participants performance. The learning curve from the TPT and the FT is shown in **Figure 4**. As for human subjects, there were no significant differences between the two conditions at the TPT ($A = 0.41$, SD = 0.26; $B = 0.38$, SD = 0.31; $t = 0.41$, $p = 0.98$), while significant differences were found at the FT ($A = 0.53$, SD = 0.34; $B = 0.40$, SD = 0.31; $t = 1.80$, $p < 0.05$).

ADDITIONAL SIMULATIONS

Extended simulation procedure

This simulation differed from the Basic simulation only because a larger number of input (motor and verbal) stimuli was used. 48 new motor patterns were created using a custom program for generating random 3-D trajectories, using a procedure based on random point generation and spline interpolation; some constraints were included in this procedure to avoid trajectories impossible to be performed by human-like arms; each trajectory was also planned to be performed in 3 + / − 1 s by virtual hands moving at constant velocity. When the final spline system defining a new trajectory was completed, a new set of 25 coordinates was calculated getting points at regular intervals.

Correspondingly, 24 new words were introduced for denoting patterns in condition A and 48 for condition B. New words were also generated using a custom software that reproduced the structure of original words. The constraint was established that each new word be different from previously generated ones, by computing the number of repeated letters (for condition A) or syllables (for condition B). The three original motor and verbal codes were kept for hand postures. The total number of stimuli was then augmented to 60.

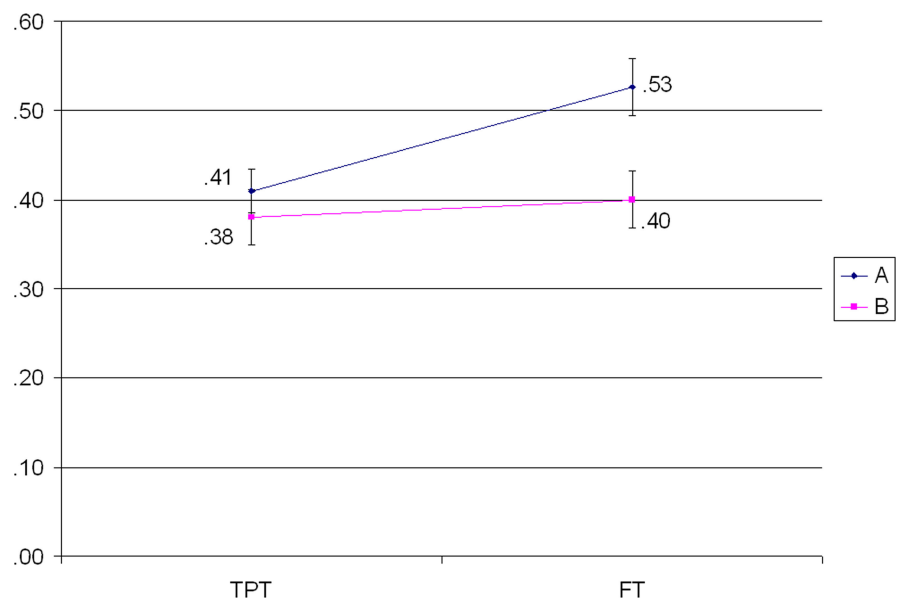

FIGURE 4 | Learning curve in basic simulation.

Extended simulation results

Even if the simulation was run using all 60 stimuli, only 30 can be considered in results since this number is already sufficient for significant differences between conditions A and B. As shown in **Figure 5**, when extending the simulation with an increasing number of stimuli the advantage for condition A persisted and was even more robust. When 30 stimuli were used, correct performance was 0.30 in condition A and 0.07 in condition B ($t = 2.04$; $p < 0.05$).

Non-systematic simulation procedure

In this simulation the same data and procedure of previous ones were replicated, with the exception that a new condition was introduced. In this condition C, the set of verbal inputs included the same bisyllable words used for condition A. The original syntactic structure was kept (5-letter words first and 4-letter words following), but words were associated with randomly selected trajectories and postures. The association was arbitrary when sentences were generated: the first word was chosen randomly in the list of words used for trajectories (e.g., baspi), the second word similarly chosen randomly in the list of words used for hand posture (e.g., nole). For example, "baspi bote" and "baspi nole" in condition A were referred to the same trajectory performed in two different hand postures, while in condition C these word combinations were referred to different trajectories and hand postures. So there was no consistent association between single words and single components of motor patterns: there was only a formal compositional-like structure but without any true compositional meaning.

Once generated, such composed sentences were obviously used consistently throughout the experiment. Since word combinations predicted the same referent as if they were single words, condition C was somewhat similar to condition B, but from the informational quantity aspect (number of different words, type-token ratio) it was equivalent to condition A (0.75).

Non-systematic simulation results

Performance in condition C (**Figure 5**) was always very scarce and smaller than other conditions, and comparable to condition B (even if the type-token ratio in this case was more favorable than in condition B). Already at the standard 12 stimuli level, there was no significant difference between B and C conditions performance ($B = 0.40$, $C = 0.22$; $t = 1.78$, $p = 0.10$) and, as the curve shows, the distance between the two conditions becomes shorter and shorter when the number of stimuli increases.

Additional simulations discussion

Results of additional (Extended and Non-systematic) simulations, taken together, suggest that the better performance in condition A may be explained by an informational advantage only when this is joined with a systematic and consistent association between words and their referents.

As we have seen, the main advantage of symbol grounding is its ability to offer more compact representations than analog ones, but even if representations exhibiting one-to-one correspondences between symbols and referents are still more compact than original

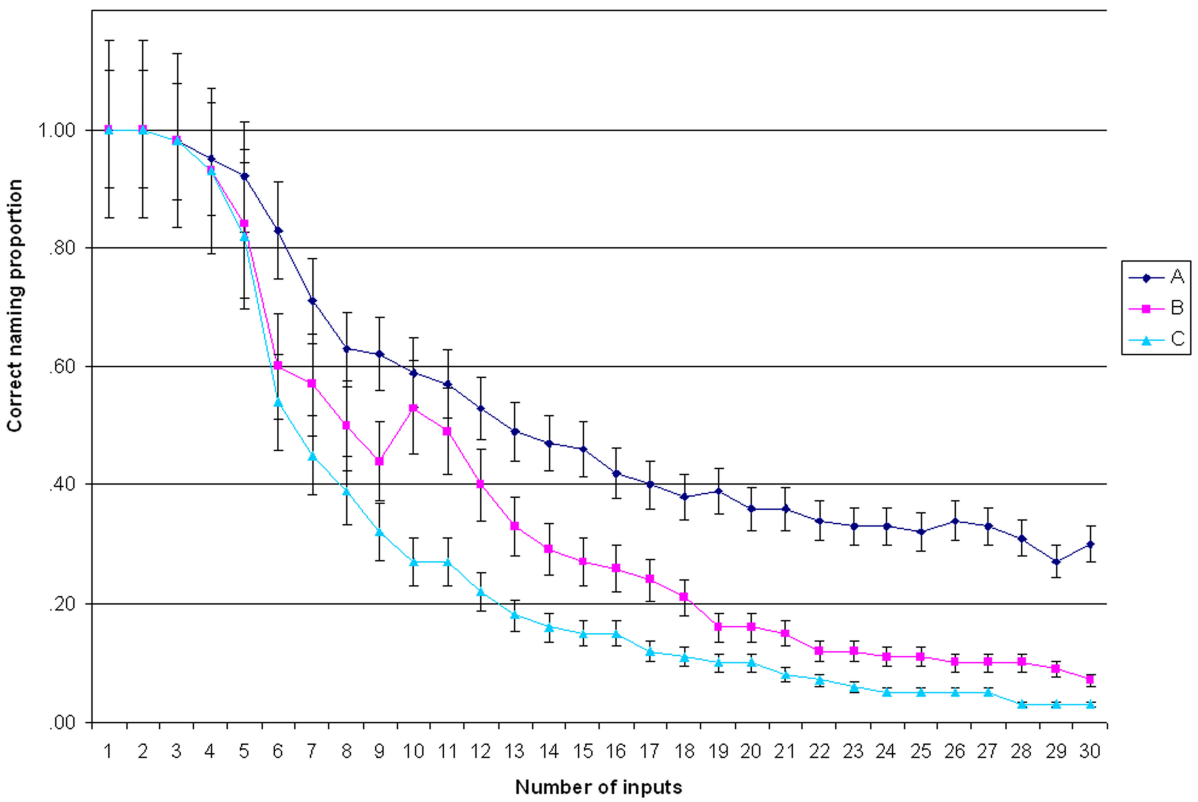

FIGURE 5 | Summary of three simulations.

analog representations, they do not work efficiently in a world where there are regularities and redundancies. We can speculate that the most important reason why compositional systems work better is not their ability of reducing cognitive load but, instead, their ability of making possible a systematic reusing of corresponding grounded analog representations.

GENERAL DISCUSSION

In this paper we described an empirical paradigm aimed at studying the possible compositional nature of grounded analog motor representations. The question asked was whether a compositional internal representation, for arm trajectories that could be executed in different hand postures, can help recognition and naming of such patterns when associated with symbolic (verbal) labels. We performed two experiments and simulations with neural networks, using meaningless stimuli, in two conditions, i.e., when labels were single words, corresponding to motor patterns regardless of hand postures (holistic condition), and when two-word sentences (the first word for the arm trajectory, the second for the hand posture) acted as labels (compositional condition).

In the first experiment, a good performance in pattern recognition was generally achieved regardless of verbal compositionality, but was poorer when distractors differed from targets just in hand postures. This showed that verbal labels did not help reducing the main source of confusability in this task, concerning hand postures, because recognition was only based on perceptual, not symbolic, cues. The mediating representation in this case was a purely analog and holistic code. Nothing could be drawn from this experiment as to the naming task, because of difficulties of the learning procedure.

In the second experiment, as a result of substantial changes to the learning procedure, we obtained acceptable learning performances in the naming task for all participants. In this task, that as we have noted above is a true grounding test, we found a significant difference between the compositional and holistic groups after having introduced a condition where the hand posture was relevant for differentiating between stimuli. Since in the compositional condition the second word had been consistently associated with the hand posture aspect, we can say that a separate grounding representation was established for it, different from the one acting as a ground for the word denoting the arm trajectory. This means that different analog (visuomotor) representations worked compositionally as a ground for the corresponding symbols, similarly to what happens with symbolic composition. The full representation of each new concept that we have tried to construct, on this view, includes both verbal and sensorimotor information corresponding, in different conditions, either to the whole pattern or to aspects of it.

In both our experiments, but notably in the first one, where no grammatical cues were present, almost all (99.9%) responses of participants in the compositional condition were syntactically correct: the first word denoted a trajectory, the second a posture. This happened even when participants, during the post-experimental debriefing, did not show to be aware of the syntactical functions of words. This is not surprising, because the automatic emerging of syntax is a well known fact, evident also from natural language acquisition in infancy. However, we would like to stress here that the correct binding between perceived patterns and appropriate grammatical word categories was established without explicit teaching, showing that syntax and semantics acquisition cannot be clear-cut separated, much like in the experiments of Sugita and Tani (2004) where a robot learned from scratch the compositional meaning of simple sentences from correspondences between sentences and sensory-motor patterns.

Several studies have stressed the role of verbal labels in motor learning. Helstrup (2000) findings support the hypotheses that motor sequences are coded as verbal strings rather than motorically or visuospatially; Frencham et al. (2004) found a better recall of hand movement sequences associated with verbal labels congruent with hand postures, still supporting the hypothesis that motor sequences are coded as verbal strings. These authors explain such results with a greater familiarity of verbal codes, easier to rehearse than actions. In our conditions, where two kinds of unfamiliar stimuli (verbal and motor) were associated, we can hypothesize that, assuming independent symbolic and analog representations, this coupling may rather reinforce a sort of mutual grounding. In fact, our findings support the idea that when an association is established between meaningless analog patterns and verbal symbols, grounding may work in a two-way direction: symbols become meaningful on the sensorimotor grounds, but also analog representations aspects (e.g., in our case, trajectories and postures) become more distinguishable when a specific label is available for them.

These remarks address also a possible issue stating that the use of two linguistic labels (words) in the compositional condition could have been a hint to look for two different components of the shown motor patterns[1]. Even if this turned out to be true, however, it could only be a demonstration that grounding can work bidirectionally, since in this case words had the power of facilitating the perceptual discriminations that in turn must necessarily be considered part of the grounding representations for the same words. This would also be evidence that grounding representations do not depend only on the visuomotor information, but language is fundamental. In any case, this does not cancel the fact that a compositional grounding was established but rather provide a further explanation of how it was obtained.

This somewhat Whorfian hypothesis, obviously, would deserve some deeper investigation, but is compatible with an interpretation of compositionality as a function of cognitive economy. If we take, as a baseline condition, that a single word for each motor pattern (group B, holistic condition) is matched with one fixed corresponding composite sentence (group A, compositional condition), then if group A performs *worst* than B, this indicates the *cost* of compositionality. On the other hand, if group A performs *better* than B, this indicates the *gain* of compositionality. Our results indicate that condition A led to a gain especially for patterns where hand postures *discrimination* was relevant. Some computational studies on language evolution (Kirby, 2002; Vogt, 2005; Smith et al., 2003) have claimed that compositional language has emerged in the cultural evolution as a consequence of the fact that examples actually encountered during verbal learning are necessarily limited (what has been called a *bottleneck* in cultural transmission); in this view, the advantage of compositionality is maximized in more structured

[1]We thank an anonymous referee for pointing out this issue.

environments, and depends on the structure of the meaning space, i.e., the number of distinctions that can be made. When distinct words are available for different aspects, attention can be best focused on such aspects, and this mutual grounding can further explain the cognitive gain of the compositional condition.

Results from our simulations clarify that such cognitive gain is not just the effect of a reduced memory load in a compositional symbol system. In fact, the environmental structure of meaning space is not important just because when the number of stimuli increases more information can be tackled with a smaller number of symbols, but because some symbols, by virtue of their grounding, can be reused as far as they are able to reinstantiate the same analog representations.

Our research can be continued in several directions. Our tasks only required that participants recognized or named visually presented (in videoclips) motor patterns. Interesting additional information about the compositional nature of motor representation could be provided by requiring the inverse performance, i.e., performance of corresponding motor patterns when being told their name. We did not implement this task for practical reasons, because we found it difficult to assess the correctness of motor performance. We are planning to use a robotic arm, programmed both to directly "teach" motor patterns associated with words/sentences and to assess how much the participant's subsequent performance matches with the original motor pattern.

Another significant improvement is trying different forms of composition. In tasks that we have considered, composition comes out from the combination of a motor pattern and a hand posture while executing it. There are theoretical reasons for supposing that other forms of composition could be considered: for example, considering other ways of executing the same pattern (slowly, firmly…), or as a concatenation of different motor patterns (e.g., left push + right push = grab). These two different composition kinds can be considered examples of the two different kinds of compositionality (concatenative and functional) described by van Gelder (1990). They can also be regarded as different ontological "wholes," as Meirav (2003) defined them: either as a sum (juxtaposition) or as a unit (according to a perspective or function).

It would be also of interest manipulating task informational complexity (e.g., the number of different words/features) and motor stimuli complexity. Using the motor coding system adopted in our networks, we can devise new simulations to obtain similarity and complexity scores by the analysis of different motor patterns, scores to be used for assessing the complexity of motor stimuli used in experimental tasks.

ACKNOWLEDGMENTS

The authors wish to thank Art Glenberg and Ivan Vankov for valuable remarks. This study was supported by European Union grant FP7-ICT-2007-1, as part of the research program I-TALK.

REFERENCES

Allott, R. (2003). *Language and Speech as Motor Activities*. Language Origins Society, Nijmegen 4–5 July 2003.

Amit, R., and Mataric, M. J. (2002). "Parametric primitives for motor representation and control," in *Proceedings of International Conference on Robotics and Automation (ICRA), May 11–15, 2002*, Washington, DC.

Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav. Brain Sci.* 28, 105–124.

Arndt, J., Lee, K., and Flora, D. B. (2008). Recognition without identification for words, pseudowords and non-words. *J. Mem. Lang.* 59, 346–360.

Barsalou, L. W. (1999). Perceptual symbols systems. *Behav. Brain Sci.* 22, 577–660.

Bizzi, E., and Mussa-Ivaldi, F. A. (2004). "Toward a neurobiology of coordinate transformations," in *Cognitive Neurosciences*, ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 413–425.

Cangelosi, A., Greco, A., and Harnad, S. (2000). From robotic toil to symbolic theft: grounding transfer from entry-level to higher-level categories. *Conn. Sci.* 12, 143–162.

Cleary, A. M, and Specker, L. E. (2007). Recognition without face identification. *Mem. Cognit.* 35, 1610–1619.

Deecke, L. (1996). Planning, preparation, execution, and imagery of volitional action (Introduction to the Special Issue "Mental representations of motor acts"). *Cogn. Brain Res.* 3, 59–64.

Fodor, J. A., and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition* 28, 3–71.

Frencham, K. A., Fox, A. M., and Maybery, M. T. (2004). Effects of verbal labeling on memory for hand movements. *J. Int. Neuropsychol. Soc.* 10, 355–361.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Greco, A. (1995). The concept of representation in psychology. *Cogn. Syst.* 4-2, 247–256.

Greco, A. and Caneva, C. (2005). "From actions to symbols and back: are there action symbol systems? COGSCI 2005," in *Proceedings of XXVII Annual Conference of the Cognitive Science Society, 2005 July 21–23,* Stresa.

Guerra-Filho, G., and Aloimonos, Y. (2006). Understanding visuo-motor primitives for motion synthesis and analysis. *J. Vis. Comput. Animat.* 17, 207–217.

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346.

Helstrup, T. (2000). The effect of strategies and contexts on memory for movement patterns. *Scand. J. Psychol.* 41, 209–215.

Jeannerod, M. (1988). *The Neural and Behavioural Organization of Goal-Directed Movements*, Vol. 15. Oxford: Clarendon Press.

Jeannerod, M. (1994). The representing brain: neural correlates of motor intention and imagery. *Behav. Brain Sci.* 17, 187–245.

Kirby, S. (2002). "Learning, bottlenecks and the evolution of recursive syntax," in *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, ed. E. Briscoe (Cambridge: Cambridge University Press), 173–203.

Kostic, B, and Cleary, A. M. (2009). Song recognition without identification: when people cannot "name that tune" but can recognize it as familiar. *J. Exp. Psychol. Gen.* 138, 146–159.

Langley, M. M., Cleary, A. M., Kostic, B. N., and Woods, J. A. (2008). Picture recognition without picture identification: a method for assessing the role of perceptual information in familiarity-based picture recognition. *Acta Psychol.* 127, 103–113.

Meirav, A. (2003). *Wholes, Sums and Unities*. Dordrecht: Kluwer Academic Publishers.

Peschl, M. (1997). The representational relation between environmental structures and neural systems: autonomy and environmental dependency in neural knowledge representation. *Non-linear Dynamics Psychol. Life Sci.* 1, 99–121.

Peynircioglu, Z. F. (1990). A feeling-of-recognition without identification. *J. Mem. Lang.* 29, 493–500.

Pulvermueller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582.

Ruschemeyer, S. A., Brass, M., and Friederici, A. D. (2007). Comprehending prehending: neural correlates of processing verbs with motor systems. *J. Cogn. Neurosci.* 19, 855–865.

Smith, K., Kirby, S., and Brighton, H. (2003). Iterated learning: a framework for the emergence of language, *Artificial Life*, 9, 371–386.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behav. Brain Sci.* 11, 1–74.

Sugita, Y. and Tani, J. (2004). "A holistic approach to compositional semantics: a connectionist model and robot experiments," in *Advances in Neural Information Processing Systems*, D. S. Touretzky, S. Thrun, L. K. Saul, and

B. Schölkopf (Cambridge, MA: MIT Press),969–976.

Thoroughman, K. A., and Shadmehr, R. (2000). Learning of action through combination of motor primitives. *Nature* 407, 742–747.

van Gelder, T. (1990). Compositionality: a connectionist variation on a classical theme. *Cogn. Sci.* 14, 355–384.

Vogt, P. (2005). The emergence of compositional structures in perceptually grounded language games. *Artif. Intell.* 167, 206–242.

Wu, L.-l., and Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: evidence from property generation. *Acta Psychol. (Amst.)* 132, 173–189.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Lang.* 46, 441–517.

Zwaan, R. A. (2004). "The immersed experiencer: toward an embodied theory of language comprehension," in *The Psychology of Learning and Motivation*, Vol. 44, ed. B. H. Ross (New York: Academic Press), 35–62.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 March 2010; paper pending published: 07 September 2010; accepted: 16 October 2010; published online: 11 November 2010.

*Citation: Greco A and Caneva C (2010) Compositional symbol grounding for motor patterns. Front. Neurorobot. **4**:111. doi: 10.3389/fnbot.2010.00111*