Published in *Cognitive Processing*, 2011, doi:10.1007/s10339-011-0427-7
The final publication is available at
http://www.springerlink.com/content/g205r8q3rk213822/fulltext.pdf

Grounding compositional symbols: no composition without discrimination

Alberto Greco, Elena Carrea

University of Genoa
Laboratory of Psychology and Cognitive Sciences
Via Balbi 6, 16126 Genoa (Italy)

greco@unige.it
elena.carrea@unige.it (corresponding author)

Abstract The classical computational conception of meaning has been challenged by the idea that symbols must be grounded on sensorimotor processes. A difficult question arises from the fact that grounding representations (GREPs) cannot be symbolic themselves but, in order to support compositionality, should work as primitives. This implies that they should be precisely identifiable and strictly connected with discriminable perceptual features. Ideally, each representation should correspond to a single discriminable feature. The present study was aimed at exploring whether feature discrimination is a fundamental requisite for grounding compositional symbols. We studied this problem by using Integral stimuli, composed of two interacting and not separable features. Such stimuli were selected in Experiment1 as pictures whose component features are easily or barely discriminable (Separable or Integral) on the basis of psychological distance metrics (City-block or Euclidean) computed from similarity judgments. In Experiment 2, either each feature was associated with one word of a two-word expression, or the whole stimulus with a single word. In Experiment 3 the procedure was reversed and words or expressions were associated with whole pictures or separate features. Results support the hypothesis that single words are best grounded by Integral stimuli and composite expressions by Separable stimuli, where a strict association of single words with discriminated features is possible.

Keywords compositionality, symbol grounding, representation, associative learning, integral stimuli, separable stimuli.

Introduction

One of the most important questions in cognitive science is how symbolic systems achieve their meaning, being connected with objects and events in the world. According to the classical computational account (Fodor 1980; Pylyshyn 1980), meaning emerges from a combination of arbitrary symbols, following formal rules. This view was based on the cognitivist conception of the language of thought (LOT, Fodor 1975) and inspired influential theories of meaning, such as Hyperspace Analog to Language (HAL, Lund and Burgess 1996), Latent Semantic Analysis (Landauer and Dumais 1997) or, in general, Distributional Models of Semantics (see e.g. Mitchell and Lapata 2010; Turner and Pantel 2010). The fundamental tenet of this view was considering meaning as based solely on word-to-word relationships, and having nothing to do with perceptual and motor aspects.

A very famous argument against this view was the Chinese Room thought experiment (Searle 1980), showing that simple manipulation of symbols is not equivalent to understanding. Harnad (1990) stressed that symbols cannot be defined only by other symbols. He named this puzzle "the symbol grounding problem", and asserted that linguistic labels (words) become meaningful symbols only when they are grounded, i.e. associated with sensorimotor processes. This issue has been widely discussed in many fields of the cognitive science literature (artificial intelligence, robotics, psychology, philosophy, linguistics) and is still particularly significant in cognitive modelling (see e.g. Cangelosi 2011).

One clear conclusion from the symbol grounding debate has been that grounding representations (GREPs) must ground symbols and cannot be symbolic themselves. A general problem that appears still unresolved concerns to what extent meaningful symbols are connected to perceptual mechanisms. More specifically, an issue that merits empirical investigation is how structured or modular GREPs are, i.e. how strict the correspondence between primitive GREPs and stimulus features is.

In fact, a strict correspondence between single representations and single perceived features was an important requirement for the classical LOT view. According to that hypothesis, conceptual primitives are combined when – to express meanings (referents) that have a complex structure – different symbols are needed. This view, by adopting the principle of compositionality (Frege 1884), was able to account for systematicity and productivity, that characterize both language and thought. This principle asserts that the meaning of composite expressions is determined by their syntactic structure and by the meaning of components.[1] The symbol grounding view may rely on the principle of compositionality as well, since new symbols can be grounded by "symbolic theft", namely by a combination of already grounded symbols, by means of propositional descriptions (Harnad 1996; Cangelosi et al. 2002).

As is well known, the LOT view has been challenged by other perspectives that have loosened the rigid coupling between symbolic representations and particular features. The symbol grounding problem is not solved by simply stating that symbols must be grounded on sensorimotor processes. The next question concerns the nature of such sensorimotor processes. If these processes are not considered as simple neural patterns, but as a psychological construct, they can be considered as intermediate representations between sensorimotor states and verbally expressed knowledge. We refer to these representations as to *grounding representations*. According to such perspectives, the question about the nature of GREPs can be answered variously, considering them as mental images, perceptual anticipations (Chambers and San 2008), structural coupling (Riegler 2002), internal emulations (Grush 2004), pre-linguistic conceptual features (Howell et al. 2005). In the connectionist view, sub-symbolic features have been posited (Smolensky 1988) and forms of non-concatenative compositionality have been suggested (Van Gelder 1990). Barsalou (1999) proposed a sort of compromise, by keeping the LOT idea of a combination of primitive features but considering such primitives as non-symbolic (modal or analogical) internal reproductions or simulations that can work compositionally (Wu and Barsalou 2009). So the aforementioned question remains open, concerning whether a single GREP is strictly linked to a particular feature so that there is a one-to-one correspondence between single perceptually discriminated features, corresponding GREPs, and corresponding symbols. In this case, a true symbolic composition takes place. A possible different view is that GREPs themselves have a composite nature, where features already perceptually interact giving place to emergent global properties. In this case we could speak of a perceptual composition.

In a fully computational language-like account, such as the classic one, there would be no room for perceptual composition. For all accounts that suppose a strict correspondence between symbols and GREPs, and then a symbolic composition, identification of perceptual or motor primitives is essential. Such identification must rely on sensorimotor *discrimination*, that hence plays a central role in classical views of cognitive composition. On the contrary, from other perspectives, such as the connectionist, a sort of composition without identification of primitives is supposed.[2] Furthermore, the compositional account relies on the presupposition that GREPs remain perfectly stable if they work as primitives.

[1] Compositionality, like symbol grounding, is also a fundamental issue in cognitive science. Formal languages used in logic and computer science are typically compositional, and the philosophical discussion on how the parts of an utterance contribute to the overall significance is still in the foreground (Horwich 2005; Werning 2005; Pagin 2003).

[2] Also in Barsalou's perspective, primitives do not coincide with static features identified once for all and then combined (Schyns et al. 1998, endorse a similar position for this respect). Views contrasting the classical account, and supporting perceptual composition, then, don't need to suppose perceptual discrimination as a starting step for identifying elements to be composed. A sort of "composition without discrimination" is assumed. A functional composition (Van Gelder 1990), for example, would be based on interacting features, or micro-features, that need not to be represented separately.

The purpose of the present study was to investigate the role of feature discriminability in grounding and in composition. We wanted, in particular, to assess if perceptual discrimination is a fundamental requisite for grounding compositional symbols. We can define "compositional symbols" as symbols used compositionally, i.e. to act as primitives in constructing complex expressions (e.g. words in sentences). We also aimed to evaluate if there is a relation between perceptual stableness of GREPs and their usability as primitives in a compositional manner.

The candidates for grounding considered in our study were meaningless words. We assumed that if a consistent and systematic learning of the association of such words with some sensorimotor features is established, this is evidence that they are transformed from simple labels to grounded symbols.

Our study rests on the following rationale. When grounding is established, a formerly arbitrary and meaningless label is associated with sensorimotor GREPs. If grounding is based on different combined GREPs, it should be better established by the association of a stimulus with separate labels for different features rather than with a single label for the whole stimulus. On the contrary, if grounding is based on a holistic GREP for the stimulus, regardless its component features, it would be better established by the association with a single label.

A suitable experimental condition for testing these two alternatives is using stimuli that have clearly separable features versus stimuli whose features are known to be interacting. The latter is the case of Integral stimuli, i.e. stimuli that are unanalyzable and are perceived as a unitary whole, having dimensions that interact and "fuse" together, where it is difficult to place selective attention to just one of them (e.g. Shepard 1964; Handel and Imai 1980; Nosofsky 1985). These are opposed to Separable stimuli, having dimensions that can be identified and processed separately.

In our experiments we controlled feature discriminability by using Integral and Separable stimuli, and compositionality by setting up two conditions: (i) compositional, using two-word expressions, with different words associated with single features, (ii) holistic, where a single word was associated with the whole stimulus.[3]

Overview

We performed 3 experiments. Experiment 1 had the purpose of selecting two types of stimuli which can be defined Separable or Integral, i.e. that respectively can or cannot permit an easy discrimination of basic component features. We replicated a classical study of Handel and Imai (1972) that distinguished between such two kinds of stimuli on the basis of their fit to different measures of similarity.

The aim of Experiments 2 and 3 was to examine the establishing of grounding with nonsense words, associated – singly or as part of two-word expressions – with stimuli previously defined as Separable or Integral. We expected that this association is best learned in the Integral stimuli condition with single-word labels and in the Separable stimuli condition with two-word expressions. Stimuli used in our experiment were pictures containing geometrical figures. Manipulated features were shape and color for Separable stimuli, brightness and saturation for Integral stimuli. In Experiment 2, pictures were presented first and then associated with words. In Experiment 3, the attentional focus was shifted by presenting words first and then pictures. The latter setting follows the more natural grounding direction (from words to perceptual features) and both conditions together test bidirectionality of grounding regardless of learning conditions.

[3] In a previous study (Greco and Caneva 2010) we used a similar procedure associating nonsense words to motor patterns. In this study, in the compositional condition the first words were associated with arm trajectories and the second words with hand postures; in the holistic condition single words were associated to whole patterns. We found that the group in the compositional condition performed better than the holistic one, only when the feature relevant for composition (hand posture) was clearly discriminable and critical for distinguishing between motor patterns. This result appears to support the idea that sensorimotor discrimination is a preliminary step for composition.

Experiment 1

The main purpose of Experiment 1 was to select two types of stimuli that can be defined Integral or Separable. Another aim of this experiment was to evaluate the stability of feature identification with such stimuli.

We replicated a task created by Handel and Imai (1972), based on the judgment of similarities between pairs of pictures that varied on two dimensions. It was expected that, with Separable stimuli, attention would be focused on each of the two dimensions of the stimulus and then, in the case of a change in both dimensions simultaneously, the perceived distance would be higher. In the same case, but with Integral stimuli, the judgments are based on a distance that is perceived directly without considering the two variations. Based on judgments on the variation of a single dimension, the expected values were computed for all logical pairs in which there was a change in two dimensions simultaneously. This range represents a metric of Euclidean type at one end and a City-block metric at the other. Comparing the actual data with the range of expected data, the fit to one of the two metrics can be computed and the integrality (Euclidean) or separability (City-block) of the type of stimuli can be inferred.

Method

Participants

Thirty students (23 female and 7 male participants) of the University of Genoa took part in this experiment for course credit. Their age ranged between 18 and 59 (*M* = 24.17, *SD* = 11,67). All participants were native Italian speakers with normal or corrected-to-normal vision; they passed a preliminary Ishihara test (Ishihara 1986) for color blindness. Informed consent was obtained. Participants were randomly assigned to one of the two experimental groups, that had one of the two different types of stimuli, as explained below.

Apparatus

Instructions, stimuli, response recordings and data collection were controlled by a PC running custom software. A 14’’ CRT monitor (Nek MultiSync V720 with 800x600 screen resolution) was used for displaying stimuli. It was initially adjusted with the built-in degauss function, at 9300K color temperature, and not changed at any point during experimental sessions described in this paper. To ensure a stable level of luminance, the monitor was allowed to warm-up for at least 5 minutes before starting a session. Participants sat approx. 60 cm away from the display, in a separate room. The room was normally lit, the monitor was positioned at 90° angle to window and other light sources were controlled to minimize glare and reflections. Only a mouse (no keyboard) was available for responses.

Stimuli

The first type of stimuli consisted of isosceles triangles (10 x 8 cm) with the same *Hue* (170), but four different degrees of brightness, corresponding to the *Value* parameter in the Münsell system[4], and four of saturation or *Chroma*. The second type of stimuli consisted of four polygons (triangle, square, hexagon, circle) each in four colors (blue, green, yellow, red). Putting these values into a matrix we get 16 possible stimuli (Table 1; all pictures are included as Electronic Supplementary Material).

[4] The Münsell system (Münsell 1905) classifies all the colors on the basis of three dimensions: *Hue*, *Value* (brightness) and *Chroma* (saturation). The figures here indicated are referred to the system used in the Microsoft Office PowerPoint ® program, ranging from 0 to 255. For example, stimulus 1 had hue=170 (fixed), brightness=80 and saturation=100, corresponding to R=49,G=49,B=111 in the RGB color code, and, in Münsell system, *Hue*=6.86, *Value*=1.49, *Chroma*=9.28.

	I Bright.: 80 II Color: Blue	I Bright.: 120 II Color: Green	I Bright.: 160 II Color: Yellow	I Bright.: 200 II Color: Red
I Satur.: 100 II Shape: Triangle	1 (R49, G49, B111)	2 (R73, G73, B167)	3 (R123, G123, B197)	4 (R178, G178, B222)
I Satur.: 150 II Shape: Square	5 (R33, G33, B127)	6 (R49, G49, B191)	7 (R104, G104, B216)	8 (R168, G168, B232)
I Satur.: 200 II Shape: Hexag.	9 (R17, G17, B143)	10 (R26, G26, B214)	11 (R85, G85, B235)	12 (R157, G157, B243)
I Satur.: 250 II Shape: Circle	13 (R2, G2, B158)	14 (R2, G2, B238)	15 (R67, G67, B253)	16 (R146, G146, B254)

Table 1 The matrix used to construct the two types of pictures. For Type I stimuli, each attribute had four different grades of constant interval (see text). For Type II stimuli, each of two attributes had four different values. Arabic numbers denote possible stimuli. RGB values for each stimulus are also shown in parentheses

Design and procedure

The design was between subjects. To make the task less abstract, pictures were presented as if they were decorations on ancient amphorae. Participants were given the role of archaeologists contacted by a museum to make assessments.The instructions explained that the task consisted of assessing the similarity between pairs of amphorae, keeping in mind that the difference was only in the decoration, i.e. the manipulated picture.

Fifteen logical pairs were formed according to the diagram of Handel and Imai (1972). A logical pair is a kind of distance between two stimuli. For example the distance between the stimulus 1 and the stimulus 5 (Table 1) is the same that there is between 11 and 15 or 4 and 8: they are all instances of a logical pair type (moving a step in vertical).

Pairs of pictures were presented one after another using a custom software. Stimuli were balanced so that they were presented an equal number of times among subjects. The first figure remained on the screen for 3 seconds, then a black mask appeared for 1 second and the second figure for 3 seconds. Participants were asked to assess the similarity of the pair just seen, by clicking on one of 11 buttons that expressed the slightest resemblance (0, left) to the maximum (10, right). Judgments of similarity were preferred to judgments of dissimilarity (used by Handel and Imai 1972) because they appeared more natural.

At the beginning, to enable a stabilization of judgments, a screen was shown for 4 seconds with all the amphorae used in the task. For additional stabilization, six pairs (identical for all participants) were then presented, chosen to show the full range of variation (1-11, 4-10, 7-13, 6-16, 9-14, 3-8). These judgments were not taken into account.

In the test phase 35 pairs of pictures were then presented as follows: one pair for each of the 15 logical pairs, each repeated twice but reversing the order of presentation of stimuli. To limit the influence of individual differences in working memory, participants could see the sequence of pictures as many times as they wanted by clicking on the appropriate button. The number of repetitions was recorded, as a further index of uncertainty in discrimination.

For each set 5 additional pairs, randomly selected, were presented. These pairs were formed by the same stimulus repeated, to determine if they correctly received the highest rating.

Separability or integrality of the stimuli was calculated on the basis of the metric that best fitted the data: if it was closer to the City-block type, the stimulus was considered Separable, if closer to the Euclidean metric it was considered Integral.

Judgments on pairs that vary on only one attribute were taken as accurate and those values were used to calculate the expected distance value for the two metrics about the pairs that vary on two attributes at once (e.g. logical pairs 1 and 2 for the pair 7, see Table 1). The computed distance *d* was

$$d = \sqrt[n]{x^n + y^n} \text{ where } 1.0 \leq n \leq 2.0 \qquad (1)$$

This formula was applied to logical pairs in which two dimensions simultaneously change (from 7 to 15, Table 1). The values for *x* and *y* were computed as distance judgments[5] using values of logical pairs denoting one-dimensional changes (from 1 to 6). A iterated calculation was done by replacing the exponent (*n* in the formula) with different values between 1.0 and 2.0. For example, the expected value for the logical pair number 7 (a diagonal shift), was computed by taking as *x* and *y*, respectively, the values of pairs 1 (vertical shift) and 2 (horizontal shift). The fit of actual data empirically obtained in the experiment to the expected values of the two metrics was then checked.

Results and discussion

The mean proportion of repetitions during judgment phase was .10 (*SD* = .30) for Type I pictures and .04 (*SD* = .19) for Type II (*t* = 4.37, *df* 1299, *P* < .0001) This shows that uncertainty in judgments was much higher for Type I pictures.

Average ratings of similarity for each logical pair are shown in Table 2. Results confirmed that Type I pictures are better suited to a metric of Euclidean type, and can be therefore defined as Integral, while Type II are suited to a City-block metric, and are defined as Separable (Table 3 and Fig. 1). Our results are consistent with literature (Handel and Imai 1972, 1980).

Average ratings for pairs formed by the same stimulus were 9.89 (*SD* = .41) for Integral and 9.99 (*SD* = .03) for Separable stimuli; this can give us confidence that participants' ratings were reliable.

The consistency of ratings was also assessed by computing the correlation between ratings of the same pairs presented in reversed order (e.g. picture 1 compared with 2, and picture 2 compared with 1). Pearson's *r* was .875 (*P* < .01) for Integral pictures and .947 (*P* < .01) for Separable pictures. This result shows that such judgments were internally consistent considering each kind of stimuli separately.

As illustrated in Introduction, we can assume that stability of perceptual features is an important requirement for GREPs in order to act as primitives for subsequent composition. We were therefore interested in evaluating which of the two types of pictures had more stable judgments. Mean differences, in absolute value, between ratings of the same pairs in different order were then computed. Higher mean differences resulted in ratings with Integral pictures (*M* = .92, *SD* = .98) compared to Separable pictures (*M* = .54, *SD* = .88). Standard deviation resulted higher for Integral pictures and the difference between means was significant (*t* = 4.41, *df* = 454, *P* < .001), showing that judgments for Integral pictures were much less stable.

Of course, integrality and separability do not denote absolute properties but may be arranged on a continuum. It has been shown that, with appropriate training, color saturation and brightness can be psychologically differentiated (Goldstone 1994; Burns and Shepp 1988). This result, however, helps us understand how the two types of pictures were perceived. In particular, pictures with perceptually separable dimensions had a distinct representation for each of them, while this was not the case for pictures whose dimensions interact.

[5] Such distance (dissimilarity) values were calculated as the complement to 10 of similarity judgments actually obtained in the experiment on the scale from 1 to 10.

<table>
<tr><th>Logical pairs
(frequencies)</th><th>Graphical representa-tions</th><th>Average rating of similarity
Type I (<i>SD</i>)</th><th>Average rating of similarity
Type II (<i>SD</i>)</th></tr>
<tr><td>1,2
(12,12)</td><td><table><tr><td>x</td><td></td><td>y</td><td>y</td></tr><tr><td>x</td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table></td><td>8.28
(<i>1.69</i>)</td><td>6.33
(<i>2.15</i>)</td></tr>
<tr><td>3,4
(8,8)</td><td><table><tr><td>x</td><td>y</td><td></td><td>y</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td>x</td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table></td><td>6.29
(<i>2.28</i>)</td><td>5.94
(<i>2.37</i>)</td></tr>
<tr><td>5,6
(4,4)</td><td><table><tr><td>x</td><td></td><td></td><td></td></tr><tr><td>y</td><td></td><td></td><td>y</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td>x</td><td></td><td></td><td></td></tr></table></td><td>4.99
(<i>3.00</i>)</td><td>5.73
(<i>2.50</i>)</td></tr>
<tr><td>7
(18)</td><td><table><tr><td>x</td><td></td><td></td><td></td></tr><tr><td></td><td>x</td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table></td><td>7.96
(<i>1.73</i>)</td><td>1.46
(<i>1.62</i>)</td></tr>
<tr><td>8
(8)</td><td><table><tr><td>x</td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td>x</td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table></td><td>5.10
(<i>2.68</i>)</td><td>1.20
(<i>1.44</i>)</td></tr>
<tr><td>9
(2)</td><td><table><tr><td>x</td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td>x</td></tr></table></td><td>3.49
(<i>3.18</i>)</td><td>.81
(<i>1.02</i>)</td></tr>
<tr><td>10,11
(12,12)</td><td><table><tr><td>x</td><td>y</td><td></td><td></td></tr><tr><td></td><td></td><td></td><td>y</td></tr><tr><td></td><td>x</td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table></td><td>5.95
(<i>2.58</i>)</td><td>1.26
(<i>1.45</i>)</td></tr>
<tr><td>12,13
(6,6)</td><td><table><tr><td>x</td><td></td><td></td><td></td></tr><tr><td>y</td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td>y</td></tr><tr><td></td><td>x</td><td></td><td></td></tr></table></td><td>4.67
(<i>2.84</i>)</td><td>1.09
(<i>1.28</i>)</td></tr>
<tr><td>14,15
(4,4)</td><td><table><tr><td>x</td><td></td><td></td><td></td></tr><tr><td>y</td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td>x</td><td>y</td></tr></table></td><td>3.92
(<i>3.10</i>)</td><td>.93
(<i>1.25</i>)</td></tr>
</table>

Table 2 Average ratings of similarity for each logical pair

	Type I			Type II		
Logical pairs	Expected value *n* = 1.0	Expected value *n* = 2.0	Actual data	Expected value *n* = 1.0	Expected value *n* = 2.0	Actual data
7	3.43	2.95	**2.18**	7.41	5.24	**8.54**
8	7.42	5.91	**5.06**	8.31	5.90	**8.80**
9	10.03	7.39	**6.81**	8.69	6.18	**9.19**
10	4.68	3.40	**3.01**	8.08	5.75	**8.80**
11	6.17	5.66	**5.10**	7.64	5.40	**8.72**
12	6.43	4.57	**3.87**	8.36	5.98	**9.30**
13	7.03	6.52	**6.78**	7.74	5.47	**8.54**
14	9.17	6.65	**5.38**	8.60	6.12	**9.04**
15	8.28	6.74	**6.78**	8.40	5.96	**9.08**

Table 3 Expected values for City-block and Euclidean metrics compared with actual data of Experiment 1 for both types of stimuli, where "n" is the exponent in the formula (1)

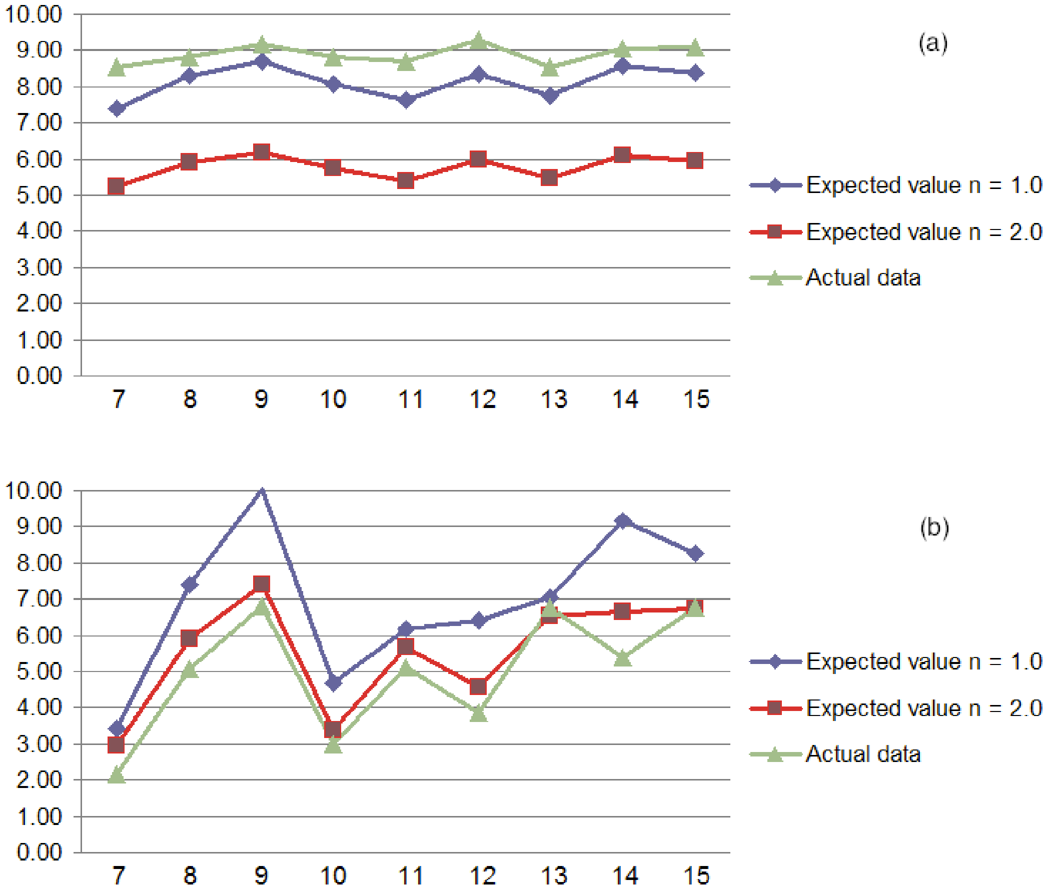

Fig. 1 Actual data compared to expected values in the two metrics, for the first type of stimuli (a) and for the second type (b)

Experiment 2

Experiment 2 was aimed at studying the relationship between types of pictures and types of linguistic symbols. Integral and Separable pictures selected as result of the first experiment were used as basis for grounding of nonsense words. Such words were associated to both types of pictures in two conditions, *compositional* (two words combined) or *holistic* (one word). According to this setup, the resulting four conditions were: (A) two-word expressions and Integral pictures, (B) single words and Integral pictures, (C) two-word expressions and Separable pictures, (D) single words and Separable pictures.

The task was learning to name pictures. According to our premises, the hypothesis was that such learning should be easier in the holistic condition with Integral stimuli, and in the compositional condition with Separable stimuli. Condition B was then expected to be easier than A, because with Integral pictures it is difficult to get any benefit from having words referring to single attributes. This would support the hypothesis that there can be no symbolic composition when there is no immediate and simple discrimination of perceptual properties of stimuli, i.e. without a separate and independent representation. Condition C was expected to be easier than D, because with Separable pictures the attributes are easily recognized and associated with the parts of the compositional label. For example, having already learned the word that refers to "yellow", having seen a triangle of that color, one needs only to learn the word that denotes "square" when is faced with a yellow square.

It may be observed that, in absolute terms, the perceptual variation between our integral stimuli was smaller than the one between separable stimuli. This difference exists, but is not relevant to our study, because – according to the aforesaid expectations – the main focus of our analysis was the relative difference in performance between the two conditions with the same kind of stimuli (integral: A versus B; separable: C versus D).

Method

Participants

Forty students (26 female and 14 male participants) of the University of Genoa, aged 19-60 years (M = 25.2, SD = 11.26), voluntarily partecipated in return for course credit. Three other participants were excluded from the experiment because failed to understand the task. Informed consent was obtained. All participants were native Italian speakers with normal or corrected-to-normal vision. They passed a preliminary test with the Ishihara test for color blindness (Ishihara 1986).

Apparatus

The apparatus was the same as in Experiment 1. All instructions and stimuli were presented on the screen.

Stimuli

Referring to Table 1, we used pictures 1, 4, 6, 7, 10, 12, 13, 15 to show at least one example of all the logical pairs, in order to ensure different degrees of similarities among pictures. Nonsense words (Table 4) were used as names for the amphorae. They were compositional (two words, each referring to one of two features) or holistic (one word for the whole stimulus). Feature1 was *Chroma* for Integral pictures and Shape for Separable pictures. Feature2 was *Value* for Integral pictures and Color for Separable ones. In the compositional condition, the first word was referred to Feature1 and the second to Feature2.

Words were constructed according to the following criteria. All the first words had 5 letters and the same ending "-spi", in order to stress their common syntactic role. All the second words had 4 letters and a similar pattern “*o*e” (where “*” stands for a different consonant). Single words used for the holistic condition had 9 letters and simply resulted from a combination of words following the patterns for the first and second word. We choose holistic labels built on the two parts of the compositional labels, and where possible exactly the same. This method ensured that word discriminability was as much as possible comparable.

	NOLE	**BOTE**	**SOVE**	**POFE**
BASPI	*BASPINOLE*			*CUSPIPOFE*
TISPI		*TISPIBOTE*	*DOSPISORE*	
RESPI		*GISPIMOPE*		*LUSPICOBE*
CUSPI	*NASPITOGE*		*RESPISOVE*	

Table 4 Labels associated to pictures. Labels used for separate features in compositional conditions are in bold. Labels for the whole stimulus are in italics. For example, stimulus 1 was named “BASPINOLE” in holistic condition, and “BASPI NOLE” in compositional condition

Design and procedure

The design was between subjects and participants were randomly assigned to one of the four conditions.

The first stage (*verbal learning*) was aimed at making participants familiar with words. All words were presented, in alphabetical order, in a panel with 8 labeled buttons. Participants were instructed to click on each button to listen to a recorded female voice that read the corrisponding word aloud. The order of listening was chosen by participants themselves. When all words had been listened, a closing button was enabled to proceed with the next stage.

The following stage (*associative learning*) was aimed at establishing the association between pictures and words. Each picture was shown for 3 seconds and then the correspondent word (or expression) appeared under it for 4 seconds. Then the word disappeared and the picture only remained for 3 seconds. A panel with all possible words was finally shown, and participants had to click the correct answer without time limits (Fig. 2). In compositional conditions they composed a two-word expression by clicking on two words. Asking the participant to act while learning had the purpose of facilitating association and consolidating the modality of response that remained the same also in the test phase. The position of words in the panel changed randomly at every trial, in order to ensure that learning could not be based on word location. Feedback was given in case of error, as the correct response was uttered and the participants could correct their answer.

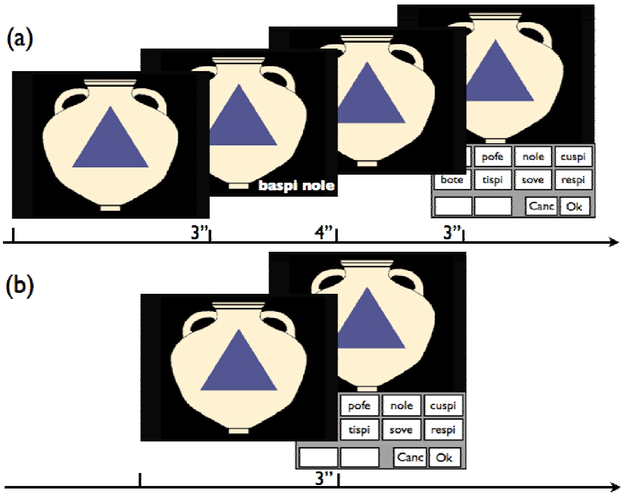

Fig. 2 Examples of screenshots and timing of stages of associative learning (a) and test (b) in Experiment 2

Picture-word associations were shown one at a time in a fixed sequence of couples, so that in the second item Feature1 was kept constant and Feature2 varied (e.g. baspi nole, baspi pofe). The same sequence of the 8 pictures, presented in order with the corresponding words, was repeated three times. No learning threshold was set to end the training session, in order to ensure that all participants had the same training.

During the final stage (*test*), each picture was presented, one at a time, without its name. Participants had to click the correct label on the panel. No feedback was given at this stage. The test included all eight pictures seen in the learning phase, each repeated twice in random order, fixed for all participants. The dependent variable was therefore the number of correct answers given during the test.

To check that participants in compositional conditions (A and C) had not learned the composed words as if they were holistic, two pictures were added at the end, which had not been seen during the associative learning. These pictures were a combination of two features seen in the learning phase, but never together. Correct answers would indicate a linguistic productivity based on compositional grounding.

During a final stage, participants were debriefed about their task understanding and personal methods used for associations. Three participants were excluded from data analysis due to misunderstandings or use of a peculiar approach (e.g. one had not considered words but only initials).

Results and discussion

Results are shown in Fig. 3. The average correct answers were A = .13 (*SD* = .33), B = .29 (*SD* = .45), C = .51 (*SD* = .50) and D = .34 (*SD* = .48). Results of additional test, performed in order to check whether pictures with never seen feature combinations were correctly named by participants, were: A = .15 (*SD* = .37) and C = .50 (*SD* = .51). Means were significantly different (*t* = 2.483, *df* = 38, *P* < .05). This suggests that participants in condition C who responded correctly to the main test also understood the semantics of the compositional parts, thus exhibiting a linguistic productivity. Participants in group A showed a performance comparable with chance both in regular and additional tests.

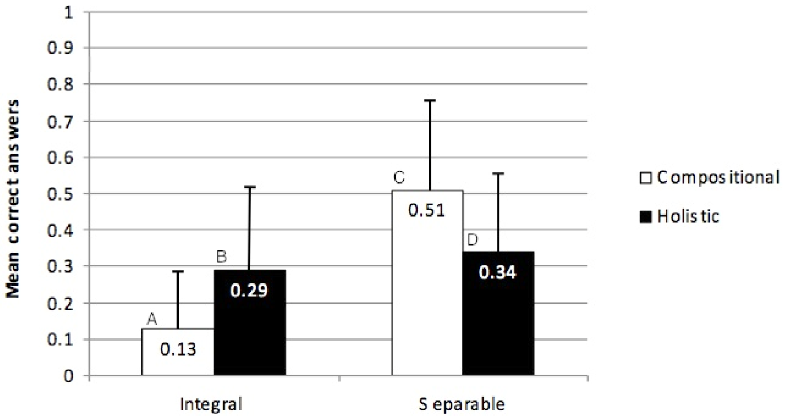

Fig. 3 Mean proportion of correct answers for each group in Experiment 2

These results approximately correspond to the expected trend, if comparison between compositional and holistic conditions regardless of the type of picture (Integral or Separable) is made. In fact, as expected, condition B was easier than A and C easier than D. Considering the global performance, however, it is evident that performance in the holistic condition was not related to feature separability. In order to assess the overall performance, a 2 (picture type: Integral vs. Separable) by 2 (verbal condition: Compositional vs. Holistic) between-subjects analysis of variance (ANOVA) was conducted on mean proportion of correct responses in the test stage. There was a significant main effect for picture type, *F* (1, 636) = 38.535, *MSE* = 1.199, *P* < .0001, but no effect at all for verbal condition, where marginal means were equal (.32). This reveals that verbal manipulation had effect only relatively to the type of picture, as if the two picture conditions were somewhat uncomparable.

We considered that this effect of the type of picture might have been due to a bias in our learning condition. Participants learned the picture-word association by observing pictures first, then words, then pictures again. Then a verbal response was required, to be done by clicking on words in a panel. We had assumed that this procedure would have encouraged a better learning of the compositional mechanism since participants in compositional condition were actively engaged in constructing composite two-word expressions. In this setting, however, the initial attention could have been too focused on pictures being associated with verbal elements.

A new experiment was therefore designed, displacing the initial attentional focus from pictures to the verbal part and the response selection from words to pictures. The experimental setup and the pictures were the same, but the whole procedure was reversed: the procedure started from word-picture association and required participants to respond by selecting the appropriate grounding picture. This procedure appears also more conforming to a most natural conception of symbol grounding: symbols in this case are the starting point that require the identification of grounding features.

Experiment 3

Method

Participants

Thirty-two students (17 female and 15 male participants) of the University of Genoa took part in this study for course credit. Informed consent was obtained. No participant in this experiment had also participated in Experiment 2. Their ages ranged between 19 and 47 (*M* = 22.75, *SD* =

6.94). All participants were native Italian speakers with normal or corrected-to-normal vision. They passed also a preliminary Ishihara test for color blindness (Ishihara 1986).

Design and procedure

The design included the same four conditions as in Experiment 2: (A) two words and Integral pictures, (B) single words and Integral pictures, (C) two words and Separable pictures, (D) single words and Separable pictures. Also the apparatus and all the stimuli were exactly the same as in Experiment 2. The only difference was the order of presentation. After the *verbal learning* phase, in the *associative learning* stage each word was shown for 3 seconds and then the correspondent picture appeared under it for 4 seconds. Then the picture disappeared and the label only remained for 3 seconds. A panel with all eight possible pictures was finally shown, and the participant had to click the correct answer without time limits (fig. 4). Feedback was given in case of error as the correct picture was shown in a separate screen area and the participants could correct their response. Word-picture associations were shown one at a time in a fixed sequence of couples, so that in the second item Feature1 was kept constant and Feature2 varied (e.g. baspi nole, baspi pofe).

The final stage (*test*) started after three repetitions of the sequence of the 8 words presented in order with corresponding pictures. As in Experiment 2, learning was not assessed at this stage.

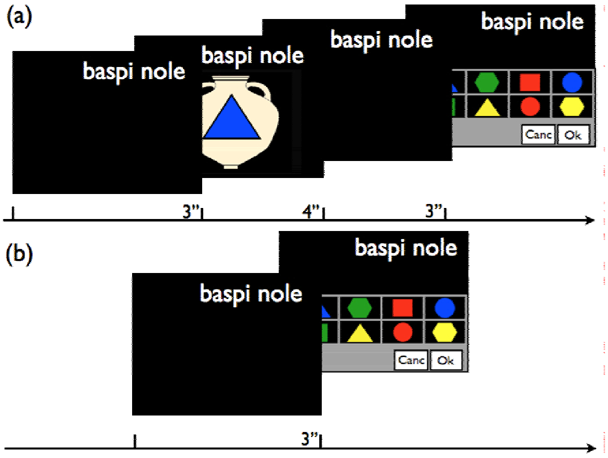

Fig. 4 Examples of screenshots and timing of stages of associative learning (a) and test (b) in Experiment 3

In the additional test, participants were asked to guess the picture connected to a label that had not been seen during associative learning. This label was constructed using two words seen in the learning phase, but never together. In the panel with the pictures the correct amphora, never seen before, was therefore also present. A possible right answer had to be based on the recombination of features associated separately to different words.

Finally, participants were debriefed about their task understanding and personal methods used for associations. No particular difficulty emerged.

Results and discussion

As for Experiment 2, with Integral pictures condition B (holistic) was espected to be easier than A (compositional) and, with Separable pictures, condition C easier than D. Average correct answers (Fig. 5) for the four groups were A = .14 (*SD* = .35), B = .27 (*SD* = .45), C = .78 (*SD* = .42) and D = .40 (*SD* = .49). Results of additional tests (performed, as previously explained, with never seen expressions) were: A = .21 (*SD* = .42) and C = .46 (*SD* = .51), significantly different (*T*

= 2.011, *df* = 54, *P* < .05). Also in Experiment 3, then, participants in condition C showed a compositional learning.

A 2 (picture type: Integral vs. Separable) by 2 (verbal condition: Compositional vs. Holistic) between-subjects two-way analysis of variance (ANOVA) was conducted on mean proportion of correct responses in the test stage. There were significant main effects for picture type, *F* (1, 508) = 101.965, *MSE* = .184, *P* < .0001, for verbal condition, *F* (1, 508)=10.872, *MSE* =.184, *P* < .001, and also for interaction, *F* (1, 508) = 46.247, *MSE* = .184, *P* < .0001.

The effect for verbal condition suggests that the change in learning conditions (i.e. asking to ground words into pictures and not the converse) in this experiment reduced the possible bias due to the initial attentional focus placed on pictures. But the amplitude of the picture type effect, still high, suggests again that feature discriminability is a crucial factor in grounding. In fact, these results, together with ones from Experiment 2, suggest that expectations about the comparison between conditions A and B and between C and D were correct. As expected, both A and B had a poor performance because of the difficulty of the task with low discriminable pictures. In particular, condition A, where component words could not be connected to discriminable features, resulted more difficult and performance was comparable with chance. In this case, capturing separately brightness and saturation resulted difficult. Most likely, pictures were perceived as different shades of blue. Learning with Integral pictures was better in the Holistic condition (B). In this case pictures represented as a whole were associated with single labels and grounding was easier.

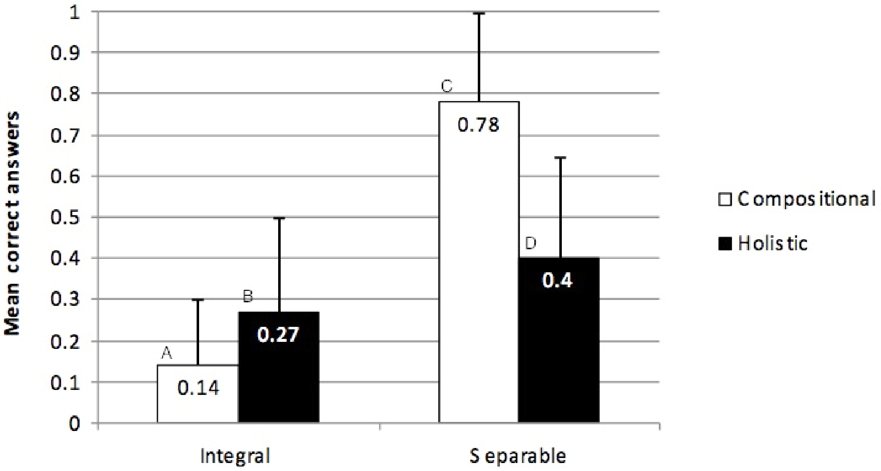

Fig. 5 Mean proportion of correct answers for each group in Experiment 3

With separable pictures, in line with expectations, performance was better in condition C than D. In condition C, the association between separable features and component words made the task straightforward. In fact, in condition C features could be easily identified and associated with words in the composite expression. This result can also be explained as a saving in working memory load, as each part of the labels could be used a second time. Hence, informational advantage could have prevailed over the effort in discovering the association rule, and over the resulting load on working memory throughout the task.

In order to control the possibility that single words suggested associations with existing words in natural language, it was checked whether performance was affected by this factor. A univariate ANOVA was performed on mean correct answers, across both Experiments 2 and 3, with First-words in composite expressions, Second-words, and Group as fixed factors. In the Compositional condition there was a considerable uniformity: no significant main effects resulted for First-words (*F* (3, 537)=1.108, *MSE* =.132, *P*=.345), for Second-words (*F* (3, 537)=.931, *MSE* =.132, *P*=.426), for interaction First-words * Group (*F* (3, 537)=.413, *MSE* =.132, *P*=.744) nor Second-words * Group (*F* (3, 537)=.679, *MSE* =.132, *P*=.565). In the Holistic condition, some words resulted significantly easier (*F* (7, 560)=3.555, *MSE* =.213, *P*=.001), but there was no interaction between Words and Group (*F* (7, 560)=.308, *MSE* =.213, *P*=.950). This confirmed that the peculiarity of some words was not related to a particular group.

An overall comparison of data from the two experiments (see fig. 3 and 5) shows clearly that a very similar result pattern was obtained, with the exception of the already discussed high performance in group C (compositional and separable condition). A final analysis was carried out in order to reveal if training change, which differentiated experiments 2 and 3, had a significant effect on performance. A further univariate ANOVA was run treating data from experiments 2 and 3 together, with Experiment (2 vs. 3) as a fixed factor, and the other variables being the same as in the previous analysis: mean correct answers as dependent measure; picture type (Integral vs. Separable) and verbal condition (Compositional vs. Holistic) as other fixed factors. As expected, significant effects were found for picture type, *F* (1, 1144)=133.926, *MSE* =.192, *P* < .0001 and for verbal condition, *F* (1, 1144) = 5.783, *MSE* = .192, *P* < .05. For the Experiment condition, the effect was significant, *F* (1,1144)=10.152, *MSE* =.192, *P* =.001, showing that the result pattern was robust across different learning conditions.

General discussion

The present study is placed into a general framework going beyond the classical cognitivist approach that considered concepts as a symbolic system regardless of their grounding in sensorimotor and embodied aspects. The research was aimed in particular at exploring whether feature discrimination is a central requisite for grounding compositional symbols. For this purpose, pictorial stimuli whose component features are easily or barely discriminable (Separable or Integral) were selected in a first experiment, on the basis of the psychological distance metric (City-block or Euclidean) that best fit empirical similarity judgment data. These stimuli were then used, in subsequent experiments, as different types of grounding representations (GREPs) for associated symbols. Grounding was empirically tested for single or composite nonsense labels, associated with such two kinds of stimuli. Additionally, this experiment showed that Integral stimuli are less stable than Separable, therefore – in principle – less appropriate for being used as primitive GREPs in compositional manner. In Experiment 2 pictures were associated with words and learning was tested. Integral pictures resulted best matched with single words, whereas Separable pictures were best grounded by composite two-word expressions. The same result pattern was obtained in Experiment 3, where the learning procedure was reversed, so that words were associated with pictures and tested responses concerned correct picture selection. Having obtained similar results reversing the learning procedure also seems to support the idea of a bidirectionality of the grounding process, where words are based on discriminated features, but feature discrimination can benefit as well from the association with verbal elements.

Globally, our results support the hypothesis that there can be no composition when there is no immediate and simple discrimination of perceptual properties of grounding stimuli, where a separate and independent representation cannot be established. These results seem also consistent with the perceptual symbol systems conception (Barsalou 1999, Barsalou et al. 2008) of a modal symbolic foundation, but extend that approach showing that primitive grounding representations should correspond to discriminable features.

Symbol grounding is a process where nonsense arbitrary "new" labels must be associated with "given" sensorimotor information in order to acquire their meaning. It is possible to speculate that, in this process, some "expectation of grounding" is generated for each new item, i.e. the search for a corresponding grounding feature or stimulus is activated. This might be seen as opening a sort of "semantic space" (Fauconnier 1985) or, more appropriately, a space for featural dimensions generated by discrimination and judgments of similarity, much like to "conceptual spaces" devised by Gärdenfors (2004). In compositional conditions, then, a two-word expression may have generated the expectation of a separate grounding for each word. The difficulty in this case could have been produced by the impossibility of filling both the spaces with corresponding features. Conversely, a straightforward grounding could have taken place in the situation where a single label was connected with a whole stimulus (not further analyzable into component features) or two labels could be associated with different discriminated features.

A limit of the present study might be related to the use of nonsense words. It is obviously almost impossible to create absolutely nonsense words, because it is impossible to insure that participants did not produce free associations with meaningful words in their natural language. In order to minimize this bias, the same pattern in constructing words was used and the same or very similar words were used in compared groups. Results showed that, nonetheless, there was no effect deriving from a particular word being exceptionally linked to some extra association.

Another critical point of our framework is that only 8 pictures and 8 words were used. From the informational point of view, linguistic compositionality is mostly effective with a high number of stimuli, that can be described more economically by the combination of a smaller number of component words. For practical reasons, however, it is difficult to set-up experimental sessions including a great number of stimuli with human participants. A neural network simulation is planned in order to extend this experimental setup; the advantage of compositional conditions is expected to increase with the number of stimuli.

The interest of the present study goes beyond perceptual discriminability. In many situations people are asked to understand symbols having referents with interacting, but not immediately separable, features. This might be the case, for example, of some causal interactions (Rehder 2003). Moreover, children are often supposed to be able to ground composite symbols based on separated features that they might not (yet) be able to identify. For example it has been shown (Shepp 1976) that younger children perceive as integral combinations that are separable for older children. If this happens in a relatively low-level perceptual domain, there is all the more reason to believe that this may be a problem with higher-level dimensions, such as for example heat and temperature (Smith et al 1985).

Acknowledgments This study was supported by European Union grant FP7-ICT-2007-1, as part of the research program I-TALK. We are grateful to Claudio Caneva for his indispensable help and to Simona Garofalo for valuable suggestions.

References

Barsalou LW (1999) Perceptual symbol systems. Behav Brain Sci 22: 577-660

Barsalou LW, Santos A, Simmons WK, Wilson CD (2008) Language and simulation in conceptual processing. In: De Vega M, Glenberg AM, Graesser AC (eds) Symbols, embodiment, and meaning. Oxford University Press, Oxford, pp 245-283

Burns B, Shepp BE (1988) Dimensional interactions and the structure of psychological space: The representation of hue, saturation, and brightness. Percept Psychophys 43: 494-507

Cangelosi A, Greco A, Harnad S (2002) Symbol grounding and the symbolic theft hypothesis. In: Cangelosi A, Parisi D, Simulating the evolution of language. Springer Verlag, London, pp 191-210

Cangelosi A (2011) Solutions and open challenges for the Symbol Grounding Problem. Int J Signs Semiot Syst (IJSSS) 1, 1: 49-54

Cantoni L, Di Blas N (2008) Pensare e comunicare. Apogeo, Milano

Chambers CG, San V (2008) Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. Cognition 108: 26-50

Fauconnier G (1985) Mental spaces: Aspects of meaning construction in natural language. MIT Press, Cambridge

Fodor JA (1975) The language of thought. Harvard University Press, Cambridge

Fodor JA (1980) Methodological solipsism considered as a research strategy in cognitive psychology. Behav Brain Sci 3: 63-109

Fodor JA, Lepore E (2002) The Compositionality Papers. Oxford University Press, Oxford

Frege G (1884) Die Grundlagen der Arithmetik. Breslau, Köbner

Gärdenfors P (2004) Conceptual spaces as a framework for knowledge representation. Mind and Matter 2, 2: 9-27

Goldstone RL (1994) Influences of categorization on perceptual discrimination. J Exp Psychol Gen 123: 178-200

Greco A (2011) Some shifts for discussing symbol grounding. Int J Signs Semiot Syst (IJSSS) 1, 1: 65-67

Greco A, Caneva C (2010) Compositional symbol grounding for motor patterns. Front Neurorobot. doi: 10.3389/fnbot.2010.00111

Grush R (2004) The emulation theory of representation: motor control, imagery, and perception. Behav Brain Sci 27(3): 377-96

Handel S, Imai S (1972) The free classification of analyzable and unanalyzable stimuli. Percept Psychophys 12: 1B

Handel S, Imai S (1980) Dimensional, similarity, and configural classification of integral and separable stimuli. Percept Psychophys 28(3): 205-212

Harnad S (1990) The symbol grounding problem. Physica D 42: 335-346

Harnad S (1996) The origin of words: A psychophysical hypothesis. In: Velichkovsky BM, Rumbaugh DM (eds) Communicating meaning: The evolution and development of language. Lawrence Erlbaum Associates, Mahwah NJ

Horwich P (2005) Deflating Compositionality. In: Reflections on Meaning. Oxford University Press, Oxford, pp 198–221

Howell S, Jankowicz D, Becker S (2005) A model of grounded language acquisition: Sensorimotor features improve lexical and grammatical learning. J Mem Lang 53(2): 258-276

Ishihara S (1986) The series of plates designe das a test for colour-blindness. Kanehara, Tokyo

Landauer TK, Dumais ST (1997) A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. Psychol Rev 104: 211–242

Lund K, Burgess C (1996) Producing high-dimensional semantic spaces from lexical co-occurrence.

Behav Res Meth Ins C, 28: 203–208

Mitchell J, Lapata M (2010) Composition in Distributional Models of Semantics. Cognitive Sci 34: 1388–1429

Münsell AH (1905) A Color Notation. G H Ellis Co., Boston

Nosofsky RM (1985) Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. Percept Psychophys 38(5): 415-432

Pagin P (2003) Communication and strong compositionality. J Philos Logic 32: 287–322

Poirier P, Hardy-Vallée B (2005) Structured thoughts: the spatial-motor view. In: Machery E, Werning M, Schurz G (ed) The compositionality of meaning and content: Applications to linguistics, psychology and neuroscience. Ontos Verlag, pp 37-58

Pylyshyn ZW (1980) Computation and cognition: Issues in the foundations of cognitive science. Behav Brain Sci 3:111-169

Rehder B (2003) Categorization as causal reasoning. Cognitive Sci *27*(5): 709-748

Riegler A (2002) When is a cognitive system embodied? Cogn Syst Res 3(3): 339-348

Schyns PG, Goldstone RL, Thibaut JP (1998) The development of features in object concepts. Behav Brain Sci 21(1): 1-17

Searle J R (1980) Minds, brains and programs. Behav Brain Sci 3: 417-457

Shepard RN (1964) Attention and the metric structure of the stimulus space. J Math Psychol 1: 54-87

Shepp B (1976) Selective attention and the processing of integral and nonintegral dimensions: a developmental study. J Exp Child Psychol 22(1): 73-85

Smith C, Carey S, Wiser M (1985) On differentiation: a case study of the development of the concepts of size, weight and density. Cognition 21: 177-237

Smolensky P (1988) On the proper treatment of connectionism. Behav Brain Sci 11: 1-74

Turney P, Pantel P (2010) From frequency to meaning: Vector space models of semantics. J Artif Intell Res 37: 141–188

Van Gelder T (1990) Compositionality: A Connectionist Variation on a Classical Theme. Cognitive Sci *14*(3): 355-384

Werning M (2005) Right and Wrong Reasons for Compositionality. In: Machery E, Werning M, Schurz G (ed) The Compositionality of Meaning and Content: Foundational Issues. Vol. I, Ontos, Frankfurt/Lancaster, pp 285–309

Wu L, Barsalou LW (2009) Perceptual simulation in conceptual combination: evidence from property generation. Acta Psychol 132: 173–189