Manipulating objects and telling words:
A study on concrete and abstract words acquisition

Anna M. Borghi[1,2], Andrea Flumini[1], Felice Cimatti[3], Davide Marocco[4], & Claudia Scorolli[1]

[1]Department of Psychology, University of Bologna;
[2]Institute of Cognitive Sciences and Technologies, CNR, Rome
[3] Department of Philosophy, University of Calabria;
[4]School of Computing and Mathematics, University of Plymouth;

Corresponding authors:
Anna M. Borghi, Claudia Scorolli

Department of Psychology, Viale Berti Pichat, 5, 40100, Bologna, Italy; Voice: +39-051-2091838;
e-mail: annamaria.borghi@unibo.it, claudia.scorolli2@unibo.it

Keywords: categorization, concepts, embodied cognition, grounded cognition, language grounding, language acquisition.

Abstract

Four experiments (E1-E2-E3-E4) investigated whether different acquisition modalities lead to the emergence of differences typically found between concrete and abstract words, as argued by the Words As Tools (WAT) proposal. To mimic the acquisition of concrete and abstract concepts, participants either manipulated novel objects or observed groups of objects interacting in novel ways (training1). In TEST 1 participants decided whether two elements belonged to the same category. Later they read the category labels (training2); labels could be accompanied by an explanation of their meaning. Then participants observed previously seen exemplars and other elements, and were asked which of them could be named with a given label (TEST2). Across the experiments, it was more difficult to form abstract than concrete categories (TEST 1); even when adding labels, abstract words remained more difficult than concrete words (TEST 2). TEST3 differed across the experiments. In E1 participants performed a feature production task. Crucially, the associations produced with the novel words reflected the pattern evoked by existing concrete and abstract words, as the first evoked more perceptual properties. In E2-E3-E4, TEST3 consisted of a color verification task with manual/verbal (keyboard-microphone) responses. Results showed the microphone use to have an advantage over keyboard use for abstract words, especially in the explanation condition. This supports WAT: due to their acquisition modality, concrete words evoke more manual information; abstract words elicit more verbal information. This advantage was not present when linguistic information contrasted with perceptual one. Implications for theories and computational models of language grounding are discussed.

1. INTRODUCTION

How do children acquire abstract words? This paper presents a study on novel categories focusing on what differs in the acquisition of concrete and abstract words. One standard way of differentiating between concrete and abstract words is to refer to their perceivability. Concrete words refer to entities that can be perceived through the senses. Abstract words refer to entities more detached from physical experience (Barsalou et al., 2003; Paivio et al., 1986; Crystal, 1995). However, the distinction between concrete and abstract words cannot be conceived of as a dichotomy (Wiemer-Hastings et al., 2001). For example, words referring to social roles (e.g., "physician") might be more abstract than words referring to single objects (e.g. "bottle"), but less abstract than purely definitional words (e.g. "odd number") (Keil, 1989). In addition, words referring to emotions probably require special classification (Altarriba et al., 1999). Further, basic and subordinate words, such as "cat" and "siamese cat", referring to single entities, can be seen as more concrete than superordinate words, such as "animal", that refer to sets of entities that differ in shape and other perceptual characteristics (e.g., Borghi et al., 2005). To summarize, the distinction between concrete and abstract words is not clear-cut, and should be intended as a continuum. However, we believe that this distinction captures some aspects of word meaning, and that it is important to understand how the process of abstraction occurs, from single instances to categories at different levels of abstraction. In particular, explaining the ways in which abstract words are represented constitutes a major challenge for embodied and grounded views of cognition, as well as for embodied computational models and robotics. The problem abstract words pose for embodied and grounded theories is clearly synthesized by Barsalou (2008, p. 634) as follows: "Abstract concepts pose a classic challenge for grounded cognition. How can theories that focus on modal simulations explain concepts that do not appear modal?" We will first clarify why explaining abstract concepts is a crucial challenge for embodied cognition, and later clarify its importance for research in robotics.

According to the standard propositional view (e.g. Fodor, 1998), the representation of both concrete and abstract concepts is abstract, symbolic and amodal. In contrast, according to standard embodied accounts (e.g. Barsalou, 1999) both concrete and abstract concepts are grounded in the sensorimotor system, and therefore are modal. Notice that both standard propositional and embodied accounts evoke a single kind of representation, either amodal or modal, for both concrete and abstract concepts.

In contrast, recent views propose that multiple representational systems are activated during conceptual processing (for a non embodied version of this view, see Dove, 2009). According to these views both sensorimotor and linguistic information play a role in conceptual representation. This idea is not entirely novel. The seminal dual coding theory by Paivio et al. (1986) applies two different kinds of representations, a linguistic and a sensorimotor code, to explain how concrete and abstract words are represented and recalled. Concrete words are recalled more easily because they activate both sensorimotor and linguistic information; differently abstract words are not "grounded", they only evoke linguistic information. Recent support to Paivio's theory comes from studies on brain imaging showing that abstract word processing is strongly lateralized towards the left hemisphere, while activation during processing concrete words is bilateral (for a review, see Sabsevitz et al., 2005). However, this might be due to the fact that the majority of the studies employ single words and tasks requiring a superficial level of processing. Recent studies requiring deeper processing, such as sentence sensibility evaluation tasks, do not provide evidence in favor of a pronounced laterality (e.g., Desai et al., 2010). The major difference between Paivio's view and embodied accounts is based on the concept of multiple representation; to elaborate, Paivio argues that abstract words are not "grounded" in perception and action systems, whereas according to the embodied perspective both concrete and abstract words activate both linguistic and perception-action information, even if these two kinds of information are differently distributed.

The Language And Situated Simulation (LASS) theory is probably the most well-known of the multiple representation theories (Barsalou et al., 2008). In this view both the linguistic and the sensorimotor system are activated during word processing. The understanding of word meanings always implies activation of the sensorimotor system (simulation), but for tasks which do not require deep processing the linguistic system might suffice. While presenting the LASS theory, Barsalou et al. (2008) suggest that for abstract concepts, linguistic information might be more relevant than for concrete concepts, but they do not advance clear predictions pertaining the differences in processing between concrete and abstract concepts, independently from the task. Thus, they argue that "different mixtures of the language and simulation systems support the processing of abstract concepts under different task conditions." (Barsalou et al., 2008, pp. 267).

More precise predictions concerning the difference between concrete and abstract words are advanced by the Words As Tools (WAT) proposal (Borghi & Cimatti, 2009; 2010), which assumes the existence of multiple representations. WAT is based on the idea, initially proposed by Wittgenstein (1953), that words are *tools* we use (see also Clark, 1998). Similarly to real tools, words can be considered as instruments to act in the social world, thus as social tools. The difference between concrete and abstract words is explained by WAT referring to the fact that, due to a different acquisition process, the role played by actions performed through words – by linguistic information – is more relevant for abstract than for concrete words. The present work aims to directly test the WAT proposal using novel categories and novel linguistic labels. According to WAT (Borghi & Cimatti, under review) perception and action are crucial in the acquisition of concrete words. Instead to acquire the meaning of an abstract word children also rely on verbal explanations (for example, explaining the meaning of "democracy" requires many more other words than for explaining the meaning of "bread"). In this respect, the role played by words as social tools is more important for abstract than for concrete words. Evidence relevant to this issue was obtained by Wauters et al. (2003), who studied different Modalities Of Acquisition (MOA) of words. They did not however, speak directly about concrete vs. abstract words. According to the authors, the meaning of a word like "ball" is acquired through perception, because every time the child hears the word, he/she sees a real ball, or a picture of it. The meaning of a word like "grammar", instead, has to be explained linguistically. Finally, the meaning of a word like "tundra" can be acquired in both ways, depending on the environment where it is learned. WAT predicts that this difference in the acquisition process can explain why, for concrete and abstract words both perception-action and linguistic information are activated. Linguistic and social information however, plays a more important role for abstract than for concrete words (e.g., Crutch & Warrington, 2005; Sabsevitz et al., 2005).

From a different perspective, an embodied and grounded account of the difference between concrete and abstract words is crucial in the process of developing intelligent machines capable of autonomously creating categories and using language. In computational cognitive science, robotics offers new opportunities for the design of artificial agents in which language is grounded on their ability to manipulate and experience the external world by means of physical interactions. The *symbol grounding problem* (Harnad, 1990) highlights that fact that, in traditional computational models, symbols are self-referential entities that require the interpretation of an external experimenter to identify the referential meaning of the lexical items. This issue has been widely discussed in the realm of cognitive science, and robotics offers a completely different way to solve the grounding problem. Indeed, in the last 20 years, many different models were created with the explicit aim of grounding symbols and language in perception (e.g. Steels, 2003) and, more recently, in action (Marocco et al., 2010; Sugita & Tani, 2005). Although the embodied approach to language in robotics is gaining increased interest, both in terms of cognitive modelling and applications, the current trend is strongly focused on systems capable of autonomously acquiring concrete concepts and words, that can be grounded on perception and action processes of the robot. Existing models do not focus on the acquisition of abstract words, except for highlighting that such abstract concepts and words permeate the entire domain of human language experience and cannot be neglected. Nevertheless, an extension of the actual grounding approach in robotics to abstract words is not automatic. In this regard, we believe that the WAT proposal offers an interesting way to incorporate abstract words in future cognitive robotic models without compromising the grounding and the embodied approach, which should be the milestone of the future robotics. On the other hand, a robotic model could be useful to complement traditional psychological experiments, and provide further evidence on the feasibility of a novel theory, such as the WAT proposal presented.

In this research we used novel categories to mimic the different ways in which concrete and abstract word meanings are acquired and then represented. Reported experiments are designed in a way that allows for replication with a computational model. Similar stimuli and training processes can be used to create a cognitive based controller for a humanoid robot (Tikhanoff et al., 2008) that will be able to perform an identical categorization task. We defined concrete concepts as having a concrete, manipulable object as a referent. Abstract concepts, on the other hand, do not have a single, manipulable object as referent; instead they refer to rather complex relations between entities. We acknowledge that the distinction we made for operational simplicity is not exhaustive and that it covers only a subset of items. For example, it leaves out word meanings referring to perceivable but not manipulable objects or entities, such as "cloud", "mountain", and "moon". Even if the referents of these words cannot be manipulated, we would consider them as concrete, as their referents are clearly perceivable, can be scanned (acted upon) with the eyes, and are easy to imagine. We decided to address the distinction between concrete and abstract words starting from the extremes of the continuum: for this reason we decided to focus on concrete, manipulable objects. As for abstract word meanings, here we did not refer to purely definitional abstract word meanings, simply based on verbal explanations (as it might be the case for a word like "philosophy") but to word meanings that evoke complex relationships between entities; due to their complexity, we suspect applying a linguistic label and explaining their meaning is crucial in order to form categories. Consider that the referents of our abstract categories were interacting moving objects – thus they were perceivable, similarly to the referents of concrete categories. As a matter of fact, in our view the formation of abstract categories always starts with some form of perception, be it visual, acoustic, tactile or otherwise.

Due to the difficulties involved in reproducing the acquisition of different kinds – concrete vs. abstract – of novel concepts/words in an artificial setting (i.e. laboratory), we operationalized the acquisition process considering two phases – the experience and the word acquisition – as follows:

a - Novel concepts acquisition: Training 1 (Experience) was designed to mimic the acquisition of concrete and abstract concepts. The idea underlying these two different acquisition processes is that, where typically concrete concepts refer to category members which are perceptually similar or elicit similar actions, abstract concepts refer to entities that show complex interactions, or do not share an evident perceptual similarity (i.e. common features are not perceptually salient). We showed

participants 3D figures of novel objects vs. 3D figures of objects interacting in novel ways. Then participants were tested (TEST 1: Categorical Recognition).

b - Novel labels acquisition: During Training 2 (Words Acquisition) participants were taught the category name; in some conditions a verbal explanation of the category meaning was added. Then participants were tested (TEST 2: Words-Objects Match). We predicted that in both tests participants would produce less errors with concrete than with abstract categories, as the first can be formed more easily on perceptual and motor basis. This difference should be reduced when a category label and a linguistic explanation of what the category members had in common were given.

The manipulation of TEST3 in the different experiments allowed us to check for the effectiveness of our operationalization of acquisition process (Experiment 1), as well as to test if the verbal labeling, possibly strengthened by a verbal explanation, reinforces learning of both concrete and abstract categories in different ways (Experiment 2, 3 and 4).

c - Real words evidence match: TEST 3 of Experiment 1 consisted of a feature production task. We predicted that the pattern of produced properties would match that typically obtained in feature generation tasks with concrete and abstract words.

d - Linguistic vs. Manual Information: In Experiments 2, 3, and 4, TEST 3 consisted of a property verification task. We chose to ask participants to respond to the objects' color because color was not relevant to the motor response. In one condition participants were required to provide a manual response (i.e. to press a key on the keyboard), and in another a verbal response (i.e. to respond "yes" with the microphone; *see* Scorolli & Borghi, 2007). We predicted facilitation for manual responses with concrete words and for mouth responses with abstract words. This would demonstrate that language is part of the representation of abstract words meanings. The rationale is the following: If linguistic information is more relevant for the representation of the meaning of abstract compared to concrete words, with abstract words phono-articulatory aspects should be accessed more easily compared to sensorimotor manual ones. Therefore, a linguistic response (even a simple "yes" response) should be facilitated compared to a manual one (see for example Scorolli & Borghi, 2007).

2. EXPERIMENT 1

The experiment was designed to mimic the acquisition of concrete and abstract categories and to verify whether the novel categories we used reproduced the acquisition process that occurs with real world categories. As anticipated, in Experiment 1, TEST 3 consisted of a production task. Before starting the experiment, participants were randomly assigned to two groups. One group was first shown the category and then tested on concrete items; later participants were shown and then tested on abstract items; the other group first learned and then was tested on the two kinds of items in reverse order. Across the experiment the order of presentation of the two blocks (concrete block; abstract block) was counterbalanced. The same methodological choice was applied to all the other three experiments.

2.1. METHOD

2.1.1. Participants
16 students of the University of Bologna took part in the study (3 men; mean age = 20.31 years; s.d. = 1.62). All were native Italian speakers, both right- and left-handed (2 left-handed) and all had normal or corrected-to-normal vision. The study was approved by the local ethics committee.

2.1.2. Materials

2.1.2.1. *3D figures of novel objects and related new labels*
We invented four novel words (*calona*, *fusapo*, *norolo*, *tocesa*) all having the same number of syllables and letters. We avoided using new words with ambiguous accents. Two of the four words ended with the vocal "a", which in Italian characterizes the female gender; the remaining two words ended with the vocal "o", which in Italian characterizes the male gender. The new words corresponded

to four new categories of objects, composed of twelve exemplars each (4X12). The criteria we followed to construct the "original" three new objects were the following:

1. CALONA was a 3D concave figure ("C" shaped). The colors we used were sky-blue and light-grey;
2. FUSAPO was a 3D figures with five protuberances ("*"shaped). The colours we used were blue and yellow (Fig. 1);
3. NOROLO was a 3D figure with small convex nooks ("N" shaped). The colors we used were red and grey;
4. TOCESA was a 3D figure shaped as wavy slash, without internal convexities or concavities ("I" shaped). The colors we used were violet and beige.

   The other nine exemplars for each category were both built by inverting the surface and depth colors (3X2), and by rotating the original figures by 180 degrees (6X2). Finally, we built 40 3D figures that were used as fillers: they did not belong to a category and were not assigned a name.

2.1.2.2. *3D figures of novel relations and related new labels*

   We invented four new words (*cofiro*, *latofo*, *panifa*, *rodela*) by following the same criteria as described for the linguistic labels used for the 3D figures of novel objects. These new words referred to new categories of *relations* between two 3D figures; each of these categories was composed by twelve exemplars (4X12). We followed the following criteria to construct the "original" three new relations (that is, novel groups of 3D interacting objects):

a. COFIRO: two 3D moving figures. After the contact just one 3D figure remained, and it moved in a straight line or in a curved line;
b. LATOFO: one 3D static figure and two 3D moving figures. After the contact two 3D figures appeared at the opposite diagonal sides of the computer screen (e.g., one at the top right of the screen and the other at the bottom left of the screen), and they moved converging towards the central point of the screen;
c. PANIFA: two 3D moving figures. After the contact one of them moved in a straight line; the other one executed a turning movement with a different velocity (Fig. 2);
d. RODELA: one 3D static figure and two 3D moving figures. After the contact the two 3D figures moved in a same (straight) line and with the same velocity, but in an opposite direction, as if the figures were pushed away from each other.

   All the 3D figures were sky-blue cylinders; they were arranged horizontally, one came from one part of the screen and the other from the other side. For LATOFO and RODELA we added a 3D static figure to the two interacting ones. This aimed to reproduce real life abstract word acquisition: some abstract words can evoke both relations between entities and static visual images (e.g., "freedom" can evoke a bird flying in the sky as well as an image of the Statue of Liberty). In other words, it can happen that objects which would be first categorized as exemplars of a concrete category (e.g. a statue) can be re-categorized and evoked by abstract words.

   The other nine exemplars for each category were built by using parallelepipeds (3X2) instead of cylinders; the movement of the 3D figures followed a vertical instead of a horizontal direction (6X2). Finally, we built 40 3D figures to use as fillers, and we constructed 40 relations between 3D figures to use as fillers. They did not belong to a category and were not assigned a name. The duration of each relation was the same for both the categories' exemplars and the fillers (4 seconds).
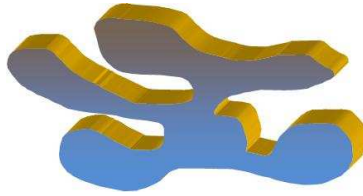
Figure 1. An exemplar of the concrete category FUSAPO; all other category members were perceptually similar to the shown exemplar.

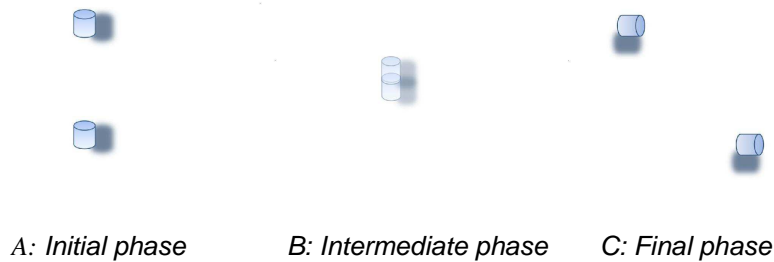*A: Initial phase* *B: Intermediate phase* *C: Final phase*

Figure 2. An exemplar of the abstract category PANIFA; the figure shows three phases of the interacting movement. All the other category members were not perceptually similar, but showed similar complex interactions.

2.1.3. Procedure

Across all experiments, participants were trained and tested individually in a quiet laboratory room. They sat on a comfortable chair in front of a computer screen. All participants were submitted to 2 training phases (Experience; Word Acquisition) and to 3 different tests (Categorical Recognition; Word-Object Match; Production).

2.1.3.1. Training 1: Experience

Training 1 aimed to reproduce the different processes underlying the acquisition of concrete and abstract concepts. Whereas typically concrete words refer to category members which are perceptually similar or elicit similar actions, abstract words refer to entities that show complex interactions or do not share an evident perceptual similarity (i.e. common features are not perceptually salient). For example, the word "truth" binds experiences and situations that might be rather complex and different. During this training session participants were sitting in front of the computer screen. They were exposed to 20 trials. In each trial either three 3D figures (in the *concrete concept* acquisition condition) or three relations between 3D figures (in the *abstract concept* acquisition condition), were shown. Both the 3D figures and the relations were novel, i.e. participants had never experienced them before. In order to mimic the acquisition of concrete concepts (e.g., BOTTLE), participants were presented with 3D figures of novel objects as previously described. They were instructed to verify whether the objects could be inserted inside a doughnut shaped 3D figure. The experimenter invited them to manipulate the objects with the mouse for 12 seconds each. In order to simulate the acquisition of abstract concepts (e.g., TRUTH), participants were instructed to observe the groups of dynamic objects until the end of their interaction (12 seconds). The 3D figures interacted in ways that revealed the existence of a common structure. For example, two objects moved toward each other, then only one of them remained on the screen, moving in a straight line (COFIRO).

2.1.3.2. TEST 1: Categorical Recognition

Training 1 was followed by a categorical recognition task (TEST 1). Participants were instructed to look at a fixation cross that remained on the screen for 500 ms. Then they were shown two exemplars of the same or different categories, and were asked to judge whether the stimuli belonged to the same category or not by pressing two different keys (left, right). The key-response mapping was

counterbalanced. They were shown 24 randomly ordered trials, with different combinations of the exemplars or of the exemplars and fillers, that is:

1) two exemplars of the same category;
2) two exemplars belonging to two different categories;
3) one exemplar of a category and one filler, that did not belong to any learned category.

Concrete concepts' exemplars remained on the screen for 2 seconds, while abstract concepts' exemplars were displayed for 10 sec. The 24 experimental trials were preceded by 2 training trials.

The Categorical Recognition task aimed to verify whether the training phase allowed participants to form a category on a purely sensorimotor basis, and to contrast it with a different category. We collected and analyzed errors, as this is the more reliable and informative measure for this particular task. Across all studies, percentages of errors are reported. We predicted that participants would produce less errors with concrete than with abstract categories, as the first can be formed more easily without the aid of language.

2.1.3.3. Training 2: Words Acquisition

After TEST 1, participants were trained to associate a linguistic label to each learned exemplar. Five exemplars from each category were randomly selected and they were presented once to participants together with the appropriate linguistic label. In order to mimic the acquisition of concrete words participants were shown 20 3D figures together with the related linguistic labels ("calona", "fusapo", "norolo", "tocesa"), presented in random order. Each trial lasted 2 seconds. Symmetrically, in order to simulate the acquisition of abstract words, participants observed the 20 relations together with the related linguistic labels ("cofiro", "latofo", "panifa", "rodela"), presented in random order. Each trial lasted 4 seconds. Participants were instructed to learn the linguistic labels associated with the 3D figures and with the relations.

2.1.3.4. TEST 2: Words-Objects Match

After the Training 2 participants had to perform a Words-Objects Match task. They were presented with 24 trials. One of the learned names and 2 figures/relations were displayed on the computer screen: the target object, corresponding to the label, and another nearby, which in half of the trials was novel and in the remaining 12 trials was an exemplar already associated with a different label. One of the two figures/relations was located on the left of the screen, the other on the right; the figure location was counterbalanced. Participants were required to decide by pressing a different key (left, right) on the keyboard which of the two was named with the shown label. This second test aimed to verify whether participants had associated a label with a category, and whether they were able to generalize it to a different category. We predicted that participants would produce fewer errors with concrete than with abstract categories, as the first rely more than the second on perception and action. However, the difference between concrete and abstract categories should be reduced compared to TEST 1, given that participants could now rely on linguistic labels as well.

2.1.3.5. TEST 3: Production task

After TEST 2, TEST 3 consisted of a feature production task with novel category names. The experimenter told participants each category name (in 4 random orders) asking them to produce the first properties that came to their mind. They were prompted to produce properties until they stopped for about 15 seconds. Properties produced were transcribed; both their frequency and production order was recorded. We predicted that the pattern of produced properties would match that typically obtained in production tasks with concrete and abstract words. Behavioral studies with production tasks, such as word association and property generation tasks, have shown that, whereas concrete words activate mainly perceptual and thematic relations, abstract words typically elicit more taxonomic relations (Borghi & Caramelli, 2001); in addition, they elicit more situations and introspective relations compared with concrete words (Barsalou & Wiemer-Hastings, 2005).

2.2. RESULTS

Across all experiments, significant results will be reported.

2.2.1. TEST 1: Categorical Recognition

We performed a one-way ANOVA on errors produced in the categorical recognition task, in which the factor Concept (Concrete vs. Abstract) was manipulated within participants. As predicted, Abstract Concepts (M = 5.21%) elicited more errors than Concrete Ones (M = 2.34%), F (1, 15) = 12.70, MSe = 5.17, p <.005 (*see* Tab. 1).

2.2.2. TEST 2: Words-Objects Match

An ANOVA was performed on the errors produced. Consider that on the screen two objects were presented, the target one and another object. Therefore in the ANOVA two factors were entered, both manipulated within participants: the factor Word (Concrete vs. Abstract) and the factor Other Exemplar (Novel vs. Learned). Both factors reached significance; Abstract Words (M = 5.01%) elicited more errors than Concrete Ones (M = 1.37%), F (1, 15) = 11.96, MSe = 17.79%, p <.005, and more errors were produced when the target exemplar was presented with a Learned (M = 4.17%) than with a Novel Other Exemplar (M = 2.21%), F (1, 15) = 15.70, MSe = 3.89, p <.005 (*see* Tab. 1).

2.2.3. TEST 3: Production task

Different analyses were performed on the production task. The number of produced properties did not differ significantly between Concrete (M = 4.18) and Abstract Words (M = 3.73); p = .29. The properties produced with each word were put together, organized in 2 different random orders, and an independent group of 12 participants were asked to rate the produced properties on a 7 point scale. They were asked to select 1 if they believed that the property was typical of words having "concrete" referents, such as bottles, screwdriver, building, cellular, and cat, and 7 if they thought the property was typical of words having "abstract" referents, such as happiness, philosophy, risk, fantasy, democracy. The raters did not know which situation the properties had been produced in. We performed an ANOVA on the ratings of the properties produced with concrete and abstract words. As predicted, we found that abstract words elicited significantly higher scores than concrete words (M =3.93; M = 3.13), F (1, 11) = 27.51, MSe = 0.14, $p <.$ 001. In addition, the scaled ratings were applied to the individual protocols in order to verify whether the properties produced and the production order of the properties for each word reflected the properties typically produced for concrete or for abstract words (the same method was used by Borghi, 2004; Borghi and Barsalou, in preparation; Wu and Barsalou, 2009). The average rating of each property was multiplied by the frequency of the produced property for each of the participants. A one-way ANOVA was performed on the obtained means, with participants as the random factor. The only factor manipulated was significant, F (1, 11) = 27.51, MSe = 0.14, $p <.$ 001, as the means obtained with Abstract Words (M = 4.14) were higher than those produced with Concrete Words (M = 3.04), indicating that the novel Abstract Words we created elicited properties typical of real-life abstract words (e.g., "singularity"; "variation"; "linear motion"); this was symmetrically true for the novel Concrete Words which elicited a higher number of properties such as "hole in the middle", "stick-shaped", "crab-shaped". In addition, the average rating on each property was multiplied by the position of the property produced for each participant according to the formula *(n+1-p) / (n-1)*r*, where *n* is the total number of properties produced by each participant for each word, *p* the position in which each property was produced and *r* the average rating on that particular property (for a similar procedure, *see* Wu & Barsalou, 2009). This normalized *p* is the position in which each property was produced, in relation to *n*, the total number of properties produced by each participant. One ANOVA was performed on the obtained means, with participants as random factor; the factor manipulated was the kind of Word (Abstract vs. Concrete Words). The ANOVA again revealed lower means for Concrete (M = 3.11) than for Abstract Words (M = 4.48), F (1, 15) = 55.38, MSe = 0.27, $p <$ .00001. This indicates that with our novel Concrete Words properties typically elicited by real concrete

words were elicited earlier, and the same was symmetrically true for our novel Abstract Words (*see* Tab. 1).

EXPERIMENT 1

Categorical Recognition

Concept	Concrete	Abstract
	2.34 (2.62)	5.21 (2.95)

Words-Objects Match

Word	Concrete	Abstract
	1.37 (1.64)	5.01 (4.90)
Other Exemplar	Novel Exemplar	Learnt Exemplar
	2.21 (3.66)	4.17 (4.27)

Production

Scaled ratings (1 concrete referent → 7 abstract r.) applied to the individual protocols		
Word	Concrete	Abstract
	3.04 (0.29)	4.14 (0.29)
Normalized position in which each property was produced (n+1-p) / (n-1)*r		
Word	Concrete	Abstract
	3.11 (0.25)	4.48 (0.46)

Table 1. Errors percentages and standard deviations (in parenthesis) for TEST 1 and 2 of Experiment 1. For TEST 3 we reported results on ratings' scores.

2.3. DISCUSSION

Results of Experiment 1 indicate that with our training with novel categories and words we were able to recreate the real-life situation in which concrete and abstract words are learned.

Results for TEST 1 (categorical recognition) indicated that it is more difficult to form abstract categories than concrete ones. In addition, results of TEST 3 (property generation task) showed that the properties produced for the concrete and abstract words we created corresponded to those typically obtained with existing concrete and abstract words. Results of TEST 1 and TEST 3 revealed that abstract categories are more difficult to form, and that abstract words are represented differently from concrete ones, as they elicit less perceptual properties, such as properties related to shape, and more abstract and relational properties.

The higher difficulty of abstract words compared to concrete ones was also maintained in TEST 2 (Words-Objects Match), when participants learned to associate a novel word to a category. Results on TEST 2 showed that the use of linguistic labels did not further facilitate the acquisition of abstract in comparison to concrete words. This reveals that the higher complexity of abstract concepts is not reduced thanks to the use of linguistic labels. A possibility is that, in order to reduce the complexity of abstract words, a verbal explanation of the category meaning is needed.

3. EXPERIMENT 2

Given our results on Words-Objects Match (TEST 2) in Experiment 1, in Experiment 2 we decided to add a verbal explanation to the linguistic label used for abstract categories. This should mirror the way the acquisition process works. Abstract words differ from concrete words insofar that the first refer to a variety of situations, states, events. Due to this complexity, linguistic labels should be more relevant for abstract than for concrete words acquisition, and the first might also require a verbal explanation of their meaning. This is often not the case for concrete words, for which the linguistic label is usually associated with the presence of the object. Experiment 2 aimed to test whether there is a facilitation effect when the meaning of abstract words is explained linguistically, compared to when only the linguistic label is provided.

In addition, the aim of Experiment 2 is to verify whether the different acquisition modality has an impact on the response modality. We designed a property verification task (TEST 3), to be performed in substitution of the production task of Experiment 1 in order to address this aim. We chose to use color as the target property as color was not relevant to the motor response and to the response device that we used.

Specifically, we predicted that, given that for concrete words manual information is more relevant than for abstract ones, participants should be faster to perform a property verification task with concrete words when they had to respond using a keyboard instead of a microphone. Symmetrically, if it is true that linguistic information is more important for the acquisition of abstract word meanings than for concrete ones, faster responses should be noted with regard to abstract words while responding with the microphone than with the keyboard. We expect a stronger effect when abstract words are presented not only with novel verbal labels but with the explanations as well.

3.1. METHOD

3.1.1. Participants

32 students of the University of Bologna took part in the study (8 men; mean age = 20.44 years; standard deviation = 1.41). All were native Italian speakers and right handed.

3.1.2. Procedure

All participants were submitted to 2 training phases (Experience; Word Acquisition) and to 3 different tests (Categorical Recognition; Word-Object Match; Property verification task). Training 1 and TEST 1 were identical to Experiment 1. However, Training 2 varied, as participants were randomly assigned to two different conditions, the Explanation or No Explanation condition. In the Explanation condition with abstract words half of the participants were told the name of the abstract category and were given an explanation clarifying the similarities of the members of a given category; in the No Explanation condition only the name was associated to the category. Training 2 for concrete categories was the same of Experiment 1.

In TEST 3 participants took part in a color verification task. Questions appeared on the screen, for example, "Is LATOFO yellow?". To respond "yes" or "no" they had to press two keys on the keyboard in one block (24 trials), or to pronounce the word "yes" or "no" in the microphone in another block (24 trials). The block order was counterbalanced. Both response times and errors were recorded. Forty-eight responses were recorded; "yes" responses corresponded to questions on five different colors (blue, red, violet, yellow for concrete words and sky-blue for abstract), and "no" responses corresponded to questions about 5 wrong colors (black, brown, green, orange, white).

3.2. RESULTS

3.2.1. TEST 1: Categorical Recognition

In the one-way ANOVA conducted on error rates the factor Concept (Concrete vs. Abstract), which was manipulated within participants, was highly significant. As predicted and as in Experiment 1, Abstract Concepts (M = 6.18%) elicited more errors than Concrete Ones (M = 1.82%), F (1, 31) = 51.32, MSe = 5.92, p <. 0000001 (*see* Tab. 2).

3.2.2. TEST 2: Words-Objects Match

We performed two different ANOVAs on the errors produced, one for the No Explanation group (A) and another for the Explanation group (B). In the first ANOVA two factors were manipulated within participants, Word (Abstract vs. Concrete, both without explanation) and Other Exemplar (Novel vs. Learned). In the second ANOVA the same factors were manipulated but, as far as the Word factor is concerned, we contrasted Abstract Words *with* Explanation vs. Concrete Words *without* Explanation. In the first ANOVA, Abstract Words (M = 4.04%) elicited more errors than Concrete Ones (M = 1.17%), *F* (1, 15) = 12.01, *MSe* = 10.93, *p* < .005, and more errors were produced when the target exemplar was associated with a Learned (M = 3.52%) than with a Novel Other Exemplar (M = 1.69%), *F* (1, 15) = 13.35, *MSe* = 3.98, *p* < .005 (*see* Tab. 2). In addition, the interaction between Word and Other Exemplar was significant, *F* (1, 15) = 5.46, *MSe* = 3.19, *p* < .04. Post-hoc LSD showed that all differences were significant (*p* <.05), with the exception of the difference between Concrete Words accompanied with a Learned vs. Novel Exemplar. With Abstract Words, instead, a Target Exemplar presented together with a Learned Exemplar elicited more errors than a Target Exemplar associated with a Novel Exemplar (*p* < .0005). In the second ANOVA both main effects were significant: Abstract Words with Explanation (M = 3.19%) elicited more errors than Concrete Words without explanation (M = 0.98%), *F* (1, 15) = 6.09, *MSe* = 12.87, *p* < .05, and more errors were produced when the target exemplar was associated with a Learned (M = 2.67%) than with a Novel Other Exemplar (M = 1.50%), *F* (1, 15) = 6.09, *MSe* = 12.87, *p* < .05 (*see* Tab. 2).

3.2.3. TEST 3: Property verification task with keyboard vs. microphone

In TEST 3 we collected both RTs and errors, for a number of reasons. First, previous work on the influence of action sentences on keyboard and microphone response devices was performed recording response times (e.g., Scorolli & Borghi, 2007). Second, differently from TEST 1 and TEST 2, no figures were presented, and participants had to read and respond to verbal questions. Thus there were no differences in the presentation timing of concrete categories (static figures) and abstract ones (videos). We will report results based on LSD test (*p* <.05) and discuss the results crucial for our hypotheses. Even though we collected RTs as well, we believe that, given that we study word acquisition, accuracy probably represents the most important measure of participants' performance.

24.77% of the trials were removed as errors. RTs above or below two standard deviations from each participant's means for correct trials were excluded from this analysis. This trimming method leads to the removal of further 3.39% of the data. The mean RTs for correct responses for true trials for each participant were submitted to two ANOVAs, one for the No Explanation group (A) and another for the Explanation group (B). In the first ANOVA two factors were manipulated within participants: Word (Abstract vs. Concrete, both without explanation) and Response Device (Keyboard vs. Microphone). In the second ANOVA we manipulated the same factors but, with the factor Word, we contrasted Abstract Words *with* Explanation vs. Concrete Words *without* Explanation. In both ANOVAs the factor Word was significant. Abstract Words (M = 958 ms; M = 950, respectively) were responded to significantly faster than Concrete ones (M = 1192 ms; M = 1200 ms), F (1, 15) = 12.52, *MSe* = 69871.63, *p* <. 005; *F* (1, 15) = 57.04, *MSe* = 17525.17, *p* < .0001 (*see* Tab. 2). Crucially in the second ANOVA we found an interaction between the kind of Words and the kind of Device, *F* (1, 15) = 11.18, *MSe* = 91173.10, *p* <. 005: Concrete Words were responded to significantly faster with the keyboard (M = 1057 ms) than with the microphone (M = 1343 ms) (LSD post hoc, *p* < .05); symmetrically Abstract Words were responded to faster with the microphone (M = 841 ms) than with the keyboard (M = 1059 ms) (LSD post hoc, *p* <.06; *see* Fig. 3).

The main effect of Word on both the analyses is of marginal interest, as it is probably due to the fact that the task was easier to perform when using Abstract Words, as the figures / entities referred to through abstract words were always light blue colored, whereas objects referred to by concrete words differed in colors. Much more crucial for our hypotheses is the interaction between Word and Response Device found in the second ANOVA (group B): as predicted, with Abstract Words provided by a verbal Explanation RTs were faster with the microphone than with the keyboard; symmetrically with Concrete Words RTs were slower with the microphone than with the keyboard (*see* Fig. 3). Finally it is interesting to notice the difference between Abstract and Concrete Words, still present *without* the Explanation (group A, 234 ms), was increased by the introduction of the verbal Explanation (group B, 250 ms), particularly in case of mouth responses.

Two further ANOVAs on errors were performed, in which the same factors were manipulated. In both analyses the factor Word reached significance: Concrete Words (group A: M = 15.69%; group B: M = 15.04%) elicited more errors than Abstract Words (group A: M = 10.55%, $F(1,15) = 4.49$, $MSe = 94.38$, $p < .05$; group B: M = 7.75%, $F(1,15) = 26.04$, $MSe = 32.69$, $p < .0005$), probably due to the different difficulty level involved in processing the color property. Crucially, the introduction of the explanation strongly reduced errors with Abstract Words (10.55% vs. 7.75%) (*see* Tab. 2).

EXPERIMENT 2

Categorical Recognition

Concept	Concrete	Abstract
	1.82 (2.17)	**6.18 (3.45)**

Words-Objects Match

Group A	Word	Concrete without explan.	Abstract without explan.
		1.17 (1.75)	**4.04 (3.99)**
	Other Exemplar	Novel Exemplar	Learned Exemplar
		1.69 (2.62)	**3.52 (3.83)**
Group B	Word	Concrete with explan.	Abstract with explan.
		0.98 (1.75)	**3.19 (3.55)**
	Other Exemplar	Novel Exemplar	Learned Exemplar
		1.05 (2.44)	**2.67 (3.40)**

Property verification task: keyboard vs. microphone

Group A	Word	Concrete without explan.	Abstract without explan.
		15.69 (8.47)	**10.55 (8.52)**
Group B	Word	Concrete without explan.	Abstract with explan.
		15.04 (5.77)	**7.75 (5.63)**

Table 2. Errors percentages and standard deviations (in parenthesis) for each TEST of Experiment 2.

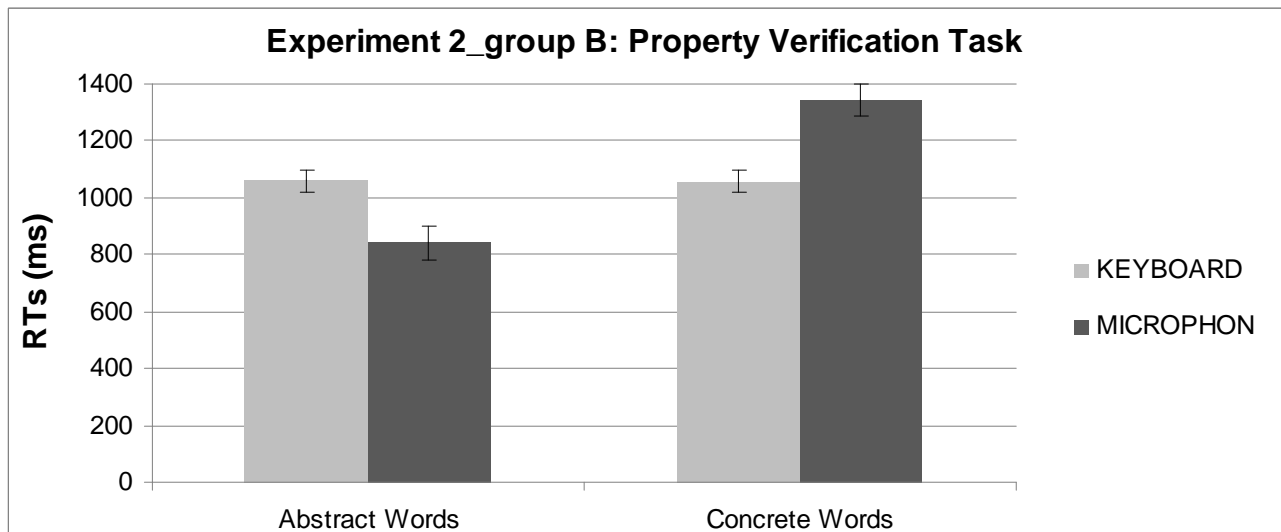**Experiment 2_group B: Property Verification Task**

Figure 3. Experiment 2, group B: Interaction between Words (Abstract with Explanation, Concrete) and Response Device (Keyboard, Microphone).

3.3. DISCUSSION

Results of Experiment 2 confirmed and extended those obtained in Experiment 1. Results on the recognition test confirm the results of Experiment 1, indicating that it is more difficult to form abstract categories than concrete ones. As in Experiment 1, TEST 2 showed that when participants learned to associate a novel word with a category, abstract words caused more difficulty in comparison to concrete words. Interestingly, abstract words without Explanation (group A) produced a significantly higher frequency of errors when the exemplar nearby has already been learned: this suggests that the categorical boundaries are less marked with exemplars referred to by abstract rather than by concrete nouns. By adding an Explanation to the label (group B), the categorical boundaries with exemplars referred to by abstract nouns become marked as the ones referred to by concrete nouns.

More crucial to our hypotheses are the results of TEST 3. As predicted, we found that Abstract Words produced faster responses with the microphone than with the keyboard; by introducing the Explanation (group B) this difference becomes significant. Symmetrically, Concrete Words (group B) were responded to more quickly with the keyboard than with the microphone. This clearly supports the WAT proposal, as it suggests that concrete words evoke more manual information, whereas abstract words elicit more verbal information.

4. EXPERIMENT 3

A potential problem of Experiment 2 was that TEST 3 (the property verification task) was submitted separately for concrete and abstract words. It is possible that, because abstract words always referred to blue objects, participants did not have to retrieve the perceptual properties of the single categories to respond, whereas this was necessary for concrete words. This could explain why RTs were faster with abstract than with concrete words. Experiment 3 is very similar to Experiment 2, with some modifications. First, given the interesting results obtained with explanations, we decided to use only the explanation condition with abstract words. Second, we balanced color information of objects referred to by both concrete and abstract categories, coloring the abstract figures. We used both concrete and abstract figures of different colors. We introduced this variation in order to solve the potential limitations of Experiment 2, thus to avoid any facilitation with abstract words in responding to the property verification task due to the fact that all abstract words' referents were blue in color. Third, in order to precisely control for the influence of learning the new labels of categorization we decided to

perform the category recognition task both before and after learning the category labels. Fourth, and most importantly, we decided to perform the property verification task at the end of the experiment, so that both concrete and abstract words were presented. This modification was introduced in order to be sure that participants referred to the learned category names to respond.

4.1. METHOD

4.1.1. Participants
18 students of the University of Bologna took part in the study (9 men; mean age = 23.00 years; standard deviation = 2.30). All were native Italian speakers, both right and left-handed (1 left-handed).

4.1.2. Procedure
All participants were submitted to 2 training phases (Experience; Word Acquisition) and to 4 different tests (Categorical Recognition without labels; Categorical Recognition with labels; Word-Object Match; Property verification task). The procedure was identical to that of Experiment 2. We only introduced three variations: 1. all abstract words were presented using both the noun and the explanation, thus the No Explanation condition for abstract words was eliminated; 2. we added a further categorical recognition task after Training 2, in order to verify whether using category labels (for both concrete and abstract words) and explanations (for abstract words) would facilitate recognition; 3. the entities to which the abstract words referred to were presented in different colors. Similarly to what we did with concrete ones, we assigned to each abstract category a specific color (light blue, light green, orange, and pink).

4.2. RESULTS

4.2.1. TEST 1: Categorical Recognition
In an ANOVA conducted on errors two factors were manipulated within participants, the factor Concept (Concrete vs. Abstract), and the factor Label Studied (Before vs. After learning the label designating the category). Only the factor Language was significant, showing that more errors were produced before (M = 1.01%) than after learning the label (M = 0.29%), $F(1, 17) = 36.26$, $MSe = 0.26$, $p < .00005$. Thus, both concrete and abstract category formation appears to benefit from language (*see* Tab. 3).

4.2.2. TEST 2: Words-Objects Match
An ANOVA was performed on errors produced in the word-object match. Both the factors Word and Other Exemplar were significant. Abstract words (M = 4.46%) elicited more errors than concrete ones (M = 2.20%), $F(1, 17) = 8.42$, $MSe = 10.89$, $p < .01$, and more errors were produced when the exemplar nearby had already been learned (M = 4.57%) than when it had not (M = 2.08%), $F(1, 17) = 61.85$, $MSe = 1.80$, $p < .000001$ (*see* Tab. 3).

4.2.3. TEST 3: Property verification task with keyboard vs. microphone
In TEST 3 we collected both RTs and errors for the reasons previously explained (*see* 3.2.3). 12.93% of the trials was removed as errors. The same trimming method of Experiment 2 was used; this lead to the removal of 3.22 % of the data. An ANOVA was performed with two factors Words (Abstract vs. Concrete) and Response Device (Keyboard vs. Microphone) manipulated within participants. As expected, the difference between Abstract and Concrete Words found in Experiment 2 disappeared (means were respectively M = 1150 and 1151 ms): this demonstrates that this difference was due to the fact that processing color was easier in Experiment 2 for abstract words, as the entities they referred to were all of the same color. Crucial to our aims, the interaction between Word and Response Device was significant, $F(1, 17) = 5.69$, $MSe = 6173.39$, $p < .03$ (*see* Fig. 4). LSD post-hoc showed that responses with the keyboard were slower than responses with the microphone for both Abstract and Concrete Words; however, with the first the difference was more marked (p = .000005)

than with the second (p = .005). In addition, responses with the Microphone in trend were faster with Abstract than with Concrete Words (p =.09).

The interaction was also significant in a further ANOVA we performed on errors with the same factors, F (1, 17) = 35.62, MSe = 0.80, $p <$. 00005. Post-hoc LSD showed that, as predicted, Abstract Words (M = 4.17%) elicited more errors than Concrete Words (M = 2.35%) with the Keyboard (M = 2.87%, 3.56% respectively, p < .0005), while they elicited less errors than Concrete Words with the Microphone (p < .04). Abstract Words using the Keyboard produced more errors than all other conditions except Concrete Words using the Microphone. Whereas Concrete Words using the Keyboard produced fewer errors than all other conditions except for Abstract Words using the Microphone (*see* Tab. 3).

EXPERIMENT 3

Categorical Recognition: without/with label

Label Studied	Before learning the label	After learning the label
	1.01 (0.68)	**0.29 (0.53)**

Words-Objects Match

Word	Concrete	Abstract
	2.20 (1.99)	**4.46 (3.63)**
Other Exemplar	Novel Exemplar	Learned Exemplar
	2.08 (2.49)	**4.57(3.22)**

Property verification task: keyboard vs. microphone

		Device	
		Keyboard	Microphone
Word	Abstract	**4.17 (1.87)**	**2.87 (1.34)**
	Concrete	**2.35 (1.72)**	**3.56 (1.92)**

Table 3. Errors percentages and standard deviations (in parenthesis) for each TEST of Experiment 3.
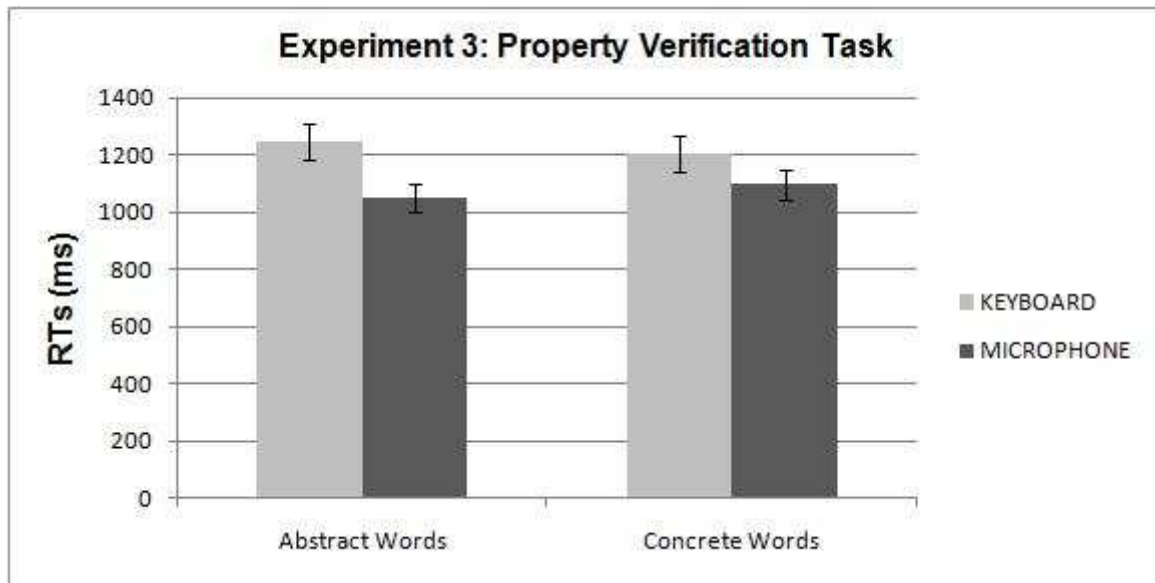
Figure 4. Experiment 3: interaction between Word (Abstract, Concrete) and Response Device (Keyboard, Microphone).

4.3. DISCUSSION

Results of Experiments 3 confirmed and extend those of Experiment 2, eliminating some potential problems. Differently from Experiments 1 and 2, in TEST 1 (Categorical Recognition Task) we found no difference between abstract and concrete words, probably due to the fact that adding a property (color) to referents of abstract words increased their difference from contrast categories. Interestingly for us, in this experiment results of TEST 1 allowed us to conclude that the introduction of category labels facilitated categorization. The comparison between the same tasks performed before and after the linguistic training reveals this.

In TEST 2, the same pattern of results as Experiment 1 and 2 emerged: abstract words elicited more errors than concrete ones, thus confirming their higher complexity as well as the fact that their borders are not so clearly marked as those observed between referents of concrete words.

In TEST 3, as expected, the advantage of abstract words over concrete ones disappeared. This confirms that it was due to the modifications we made: we introduced color differences between the entities to which abstract categories referred to, in order to be certain that the task did not differ in difficulty for concrete and abstract words. The interaction between Response device and Words revealed that responses with the keyboard were always slower than responses with the microphone but that the discrepancy between microphone and keyboard was more marked with abstract than with concrete words. The pattern was complemented by the results on errors, which were fully in line with our predictions: more errors were elicited by abstract words using the keyboard, and by concrete words when using the microphone.

5. EXPERIMENT 4

The two last experiments left two issues unsolved. In Experiment 2 we manipulated the presence of explanations, but only for abstract words. In Experiment 3 participants were given explanations to clarify abstract word meanings because this would mirror the typical acquisition process of abstract categories. However, manipulating explanations only for abstract words did not allow us to precisely determine if there is an effect of explanation also for concrete words. Therefore, in Experiment 4 we presented only the category label or the label and the explanation for both concrete and abstract words. In addition, in this experiment for abstract words the information provided by perceptual input and that provided by the verbal label plus explanation were disentangled. To dissociate these two sources of information we used different colors for the members of abstract categories, in order to induce

participants to categorize them on the basis of color, but the labels and explanations for these items still rested on items' reciprocal interaction, rather than on their perceptual features. Therefore, with concrete items the label and the explanation converged with the category formed on the basis of perceptual Experience (Training 1), whereas with abstract items the verbal and perceptual experience did not match. This manipulation was introduced in order to verify whether the advantage of the microphone responses was simply due to phono-articulatory aspects of the words or to their conceptual content as well. Our major predictions concerned TEST 3: 1) If the mouth activation found in Experiment 3 (TEST3, vocal responses) is due to a motor phono-articulatory activation pertaining to the superficial linguistic information, in Experiment 4 (TEST 3) we should find an advantage of vocal responses both with concrete and abstract words, as well as a main effect of the verbal explanation. 2) If, consistent with the WAT proposal, the previously found advantage for vocal responses pertains also the category content, then it should play a major role if it complements information given by perception and action, not if it contrasts with it. Therefore we should find a difference with results of Experiment 3: there should be an advantage of the microphone over the keyboard only when the label and the explanation do not contrast with perceptually-based categories. In this experiment, this contrast characterizes abstract categories.

5.1. METHOD

5.1.1. Participants
18 students of the University of Bologna took part in the study (7 men; mean age = 24.55 years; standard deviation = 3.66). All were native Italian speakers and right handed.

5.1.2. Procedure
The procedure was similar to that of Experiment 3, except for two variations. First, during Training 2 (Words Acquisition) half of the participants were taught the linguistic labels (Label group) vs. the linguistic labels plus the verbal explanation (Label+Explanation group), both for abstract and concrete items. The verbal explanations for abstract items were the same used in Experiment 2 and 3, so they basically described the kind of interaction. For concrete items the verbal explanations focused on the figure shape, avoiding any reference to its color (e.g. CALONA: "a figure having a concavity"). The number of words for each explanation across both the abstract and the concrete blocks was equal.

Second, in Experiment 4 we used different colors for each category member: for both concrete and abstract items, two members of each category had the same color as two members of another category. For example, FUSAPO surface could be yellow, blue, red or sky blue; its thickness was always the same, i.e. dark blue. NOROLO surface shared with FUSAPO surface yellow and blue colors, but it could be also green or violet; the color of the thickness was always dark blue.

5.2. RESULTS

5.2.1. TEST 1: Categorical Recognition
We performed two different ANOVAs on errors: one for the Label group and another for the Label+Explanation group. In the first ANOVA two factors were manipulated within participants, the factor Concept (Concrete vs. Abstract), and the factor Label Studied (Before vs. After learning the category label). In the second ANOVA we manipulated the same factors, but the levels of Label Studied factor differed (Before vs. After learning the category label with explanation). In the first ANOVA, both main effects were significant: more errors were produced with Abstract (M = 7.41%) than with Concrete Concepts (M = 3.36%), $F (1, 8) = 7.73$, $MSe = 19.12$, $p < .03$, and more errors were produced before (M = 6.54%) than after learning the label (M = 4.22%), $F (1, 8) = 17.31$, $MSe = 2.79$, $p < .005$. The factor Concept was also significant in the second ANOVA: more errors were produced for Abstract (M = 11.17%) than for Concrete Concepts (M = 3.60%), $F (1, 8) = 32.61$, $MSe = 15.38$, $p < .0005$. The factor Label Studied did not reach significance, but we found a significant interaction between Concept and Label Studied, $F (1, 8) = 7.26$, $MSe = 1.83$, $p < 0.05$ (see Table 4),

due to the fact that after learning label+explanation errors decreased with concrete words (LSD post-doc, $p > .005$), but not with abstract ones.

5.2.2. TEST 2: Words-Objects Match

We performed two different ANOVAs on errors: one for the Label group and another for the Label+Explanation group. In the first ANOVA two factors were manipulated within participants: Word (Concrete vs. Abstract) and Other Exemplar (Exemplar already learned, with only linguistic label vs. Exemplar not learned). In the second ANOVA the same factors were manipulated, but the levels of the Other Exemplar factor differed (Exemplar already learned, with label+explanation vs. Exemplar not learned).

In both ANOVAs we found a significant main effect of the factor Word: fewer errors were produced with Concrete than With Abstract Words (group A: M = 2.55%, M = 6.48% respectively, $F (1, 8) = 8.31$, $MSe = 16.77$, $p < 0.05$; group B: M = 2.66%, M = 7.29% respectively, $F (1, 8) = 22.13$, $MSe = 8.71$, $p < 0.005$; *see* Tab. 4).

5.2.3. TEST 3: Property verification task with keyboard and microphone

In TEST 3 for RTs 35.63% of the trials was removed as errors. We used the same trimming method as previous experiments; this lead to the removal of 2.38% of the data. An ANOVA was performed with three factors: Word (Abstract vs. Concrete), Response Device (Keyboard vs. Microphone) and Verbal Explanation (Without vs. With), the last one manipulated between participants. We found that vocal responses (M = 1128.73 ms) were 147.57 ms faster than manual responses (M = 1276.30 ms), even if the factor Response Device did not reach significance, $F (1, 16) = 3.48$, $MSe = 112633$, $p < .08$. The interaction between the factors Word and Response Device was significant, $F (1, 16 = 4.58)$, $MSe = 47804.8$, $p < .05$. The advantage of the microphone over the keyboard did not reach significance with abstract words (M = 1221.67 vs. M = 1184.37, respectively), while with concrete words responses with the microphone (M = 1073.09) were faster than responses with the keyboard (M = 1330.93 ms) (LSD, $p > .01$) (*see* Fig. 5)

Finally in the ANOVAs on errors with the same factors, we found that abstract words (M = 20.01%) elicited more errors than concrete ones (M = 15.63%), $F (1, 16) = 7.84$, $MSe = 44.13.08$, $p < .05$. The significant interaction between Word and Response Device, $F (1, 16) = 5.87$, $MSe = 37.90$, $p < .05$, was due to the fact that abstract words with the microphone (M = 21.79%) elicited more errors than concrete words with both the keyboard (M = 17.36%) and the microphone (M = 13.89%) (LSD post-hoc, $p > .05$) (*see* Tab. 4).

EXPERIMENT 4

Categorical Recognition: without/with label

Group A: label	Concept	
	Concrete	Abstract
	3.36 (2.59)	**7.41 (5.04)**

Group B: lab.+explanation	Concept	
	Concrete	Abstract
	3.70 (2.35)	**11.17(4.47)**

Words-Objects Match

Group A: label	Word	
	Concrete	Abstract
	2.55 (4.08)	**6.48 (4.57)**

Group B: lab.+explanation	Word	
	Concrete	Abstract
	2.66 (4.45)	**7.20 (4.36)**

Property verification task

		Device	
		Keyboard	Microphone
Word	Abstract	**18.23 (8.37)**	**21.79 (6.17)**
	Concrete	**17.36(6.38)**	**13.89 (5.19)**

Table 4. Errors percentages and standard deviations (in parenthesis) for each TEST of Experiment 4.
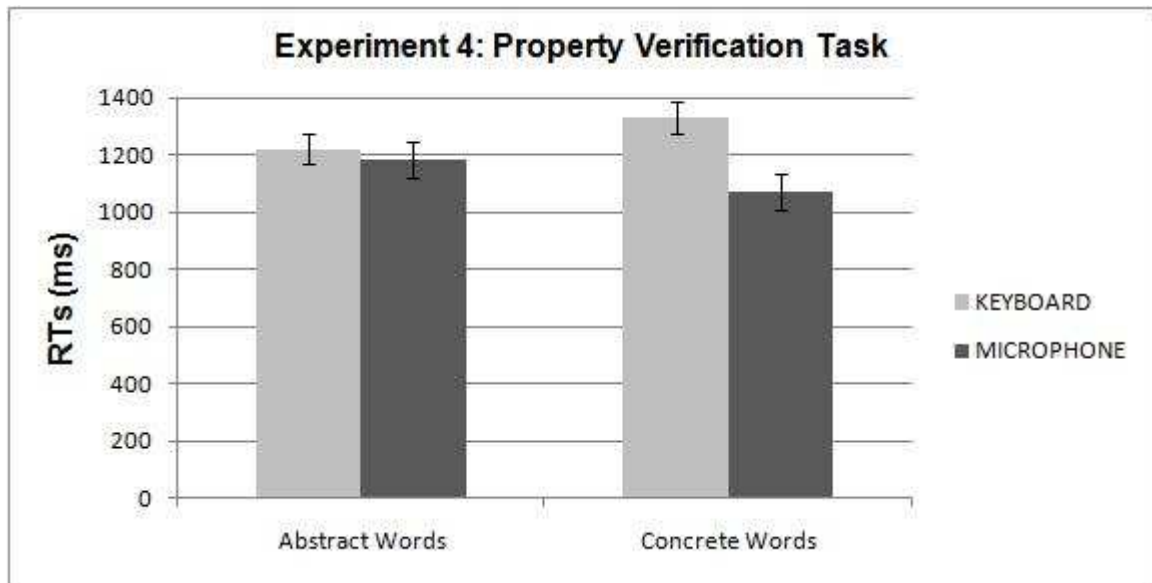
Figure 5. Experiment 4: interaction between Word (Abstract, Concrete) and Response Device (Keyboard, Microphone).

5.3. DISCUSSION

Results of TEST 1 indicate that the difference between the condition Label and No-Label increases when an explanation is added to the category name. Thus explanations facilitate categorization, as they render clearer category boundaries. However, the contribution of explanations is relevant only for concrete categories. For abstract categories, explanations do not help, as the information they provide is in contrast with perceptually based categorization.

Results of TEST 3 are the most intriguing. As predicted, participants were faster to respond with the microphone than with the keyboard with all words: this suggests that the phono-articolatory aspect of the words pronounced during acquisition affects performance. It is unclear, however, why no effect of explanation was present. The most important result is the interaction showing that the advantage of the microphone over the keyboard is more marked with concrete than with abstract words, both in RTs and accuracy. This suggests that not only phono-articolatory but also conceptual information is at play in explaining the advantage of responses with the microphone. In fact this advantage shows up only when there is a convergence between the linguistic information (label and explanation) and the category formed on sensorimotor basis, that is only with concrete words. One could object that the effect is due to the fact that explanations used with concrete words might have reduced the importance of manual interaction with objects. However, this doesn't account for the advantage of the microphone with concrete over abstract words.

6. GENERAL DISCUSSION

Four experiments were designed to study the acquisition of concrete and abstract categories and words. We chose to use novel categories, in order to avoid confounds often associated with research on concrete and abstract words. We identified some characteristics which are typical of abstract but not of concrete categories, and we created novel categories according to these criteria. First, abstract categories do not refer to a single object but rather to a complex relationship between different objects. In addition, the entities to which abstract categories refer are not manipulable, even though they are perceivable, as they are interacting moving objects. Notice that our distinction does not cover the whole continuum ranging from abstract to concrete categories. Further work is needed for a thorough investigation of different typologies of concrete and abstract words (for attempts in this direction, see Setti & Caramelli, 2005). Here we used two different examples of concrete and abstract words and have shown that different processes are involved in their acquisition.

In Experiment 1 we ascertained, using a production task, that the pattern of produced properties with our novel concrete and abstract categories was similar to that typically elicited by concrete and abstract words.

In Experiment 2, 3, and 4 we introduced a modification: abstract words were not only learned by associating a label with the entities / relations they referred to, but also when an explanation of their meaning was provided. This learning situation should resemble the learning process of children, as studies on MOA show. We found that this learning process influenced a later property verification task: participants responded earlier to concrete words while using the keyboard, while responses with abstract words were faster while using the microphone. Similar results with action words and effectors showed that, while comprehending sentences referring to mouth-related actions, response times were faster with the microphone than with the keyboard (Scorolli & Borghi, 2007). In addition, in line with WAT, participants' performance with abstract words was improved when provided with a verbal explanation (Experiment 2, group B; Experiment 3). This effect was not observed in concrete words. The fact that the advantage of the explanation was confined to abstract words revealed that the difference is not simply due to phono-articulatory aspects, but that for accessing the meaning of abstract words linguistic information plays a major role. This was confirmed in Experiment 4, in which we found that, due to the fact that with abstract words the verbal label and explanation were in contrast with the already formed perceptually-based category, the advantage of the microphone over the keyboard was reduced compared to the other experiments.

Our results are in line with embodied and grounded theories of categorization and language comprehension. Namely, both the concrete and the abstract categories we used are grounded, as they have objects or relations as referents. We were able to demonstrate that they are not only grounded in perception and action systems, but that for forming them language is important. This leads to the prediction that abstract words would not only activate linguistic areas in the brain, but also classic motor and sensorimotor areas (for initial fMRI results with existing concrete and abstract word combinations, see Menz, Scorolli, Borghi & Binkofski, submitted).

Results are in line with the predictions advanced by the WAT proposal. They reveal that a different learning process might lead to differences in performance on different tasks, such as a production task versus a property verification task. In addition, it provides evidence that for representing abstract concepts motor linguistic information is more important than manual information, whereas for representing concrete concepts the pattern is opposite.

One result was not predicted by the WAT proposal. Our results show that the formation of both concrete and abstract categories benefits from learning a linguistic label. As it emerges from the categorical recognition task in Experiment 3, language is relevant because it helps to better differentiate between categories (Mirolli & Parisi, in press). The recognition test in Experiment 4 (TEST 1a-1b) shows that labeling is mostly helpful when an explanation of the category meaning is added. As shown in test 2 and test 3 of Experiment 2, the benefit provided by language is higher in the case of abstract categories when a verbal explanation (group B) supports the linguistic label. Nevertheless, when no explanation is provided, labeling is useful for both concrete and abstract categories. In sum: labeling helps categorization, independently of category complexity. However, even when no explanation is provided, given that abstract words do not refer to manipulable objects and are linked by complex relational properties, language plays a major role in their representation.

This opens an interesting scenario. Language is relevant for both concrete and abstract words because it helps better differentiate between categories. However, in tasks for which categorization is not relevant, such as the color verification task, it is more accessible in the representation of abstract than of concrete word meanings. This might occur because: a. the members of abstract categories are not manipulable; and b. more linguistic information is typically associated with the acquisition of abstract word meanings.

Further work should address unsolved issues in this research. One important expansion could be to introduce the social development component implied in word acquisition. We used language in a very simple way, through adding labels or explanations to read and to associate with the relevant categories. Thus, language was not associated with a real social experience, as in real life. Further work should

take aspects of social development which characterize language acquisition into account. A further limitation is that the variety of real-life concrete and abstract words is very broad, and we were able to model only a subset of these.

Finally, we believe this work has important implications for modeling. The design of the task is particularly suitable for further modeling applications and replication. We succeeded in isolating some properties we believe to be relevant in real life categories and built novel categories based on our assumptions. We could verify that the behavioral responses produced within these novel categories were similar to the ones produced within real life categories and settings. This procedure demands an additional modeling development. We believe that computational models can integrate and generate a more general description the experimental results. For example, a robotic model, as discussed in the introduction, can benefit from a psychological theory, that provides a possible way to tackle a new and complex problem for robotics itself, such the theory described focusing on the grounding and acquisition of abstract words. On the other hand, the same robotic model can be tested in many different experimental situations, some of them not even applicable to human subjects. Experiments of this nature can complement and integrate experiments with human participants and can offer new insights and hypotheses to test. Moreover, the process of developing artificial cognitive models always requires a profound articulation of the theory implemented. This fact forces the researcher to well define and to operationally describe every aspect of the theory and, at the same time, it emphasizes the importance of the central aspects of the theory, that can be fully exploited and validated by the model, at least as a preliminary proof of concept.

References

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers*, *31*, 578-602.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.

Barsalou, L.W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*, 617-645.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, & A. C. Graesser (Eds.). *Symbols, embodiment, and meaning* (pp. 245–284). Oxford, U.K.: Oxford University Press.

Barsalou, L.W., Simmons, W. K., Barbey, A.K., & Wilson, C.D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Science*, *7*, 84-91.

Barsalou, L.W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher and R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought*, pp. 129-163 . New York: Cambridge University Press.

Boot, I. & Pecher, D. (2009). Similarity is closeness: Metaphorical Mapping in a Conceptual Task. *Quarterly Journal of Experimental Psychology*, *29*, 1-13.

Borghi, A.M. (2004). Objects concepts and action: Extracting affordances from objects' parts. *Acta Psychologica*, *115*, 1, 69-96.

Borghi A., & Barsalou L. (in preparation). Perspective in the conceptualization of categories.

Borghi, A. & Caramelli, N. (2001), Taxonomic relations and cognitive economy in conceptual organization, in Moore, J.D., & Stenning, K. (eds.), *Proceedings of 23rd Meeting of the Cognitive Science Society*, London: Erlbaum, 98-103.

Borghi, A.M., Caramelli, N., & Setti, A. (2005). Conceptual Information on Objects' Locations. *Brain and Language, 93,* 140-151.

Borghi, A.M., & Cimatti, F.(2009). Words as tools and the problem of abstract words meanings. In N. Taatgen & H. van Rijn (eds.). *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2304-2309). Amsterdam: Cognitive Science Society.

Borghi A. M., & Cimatti, F. (2010). Embodied cognition and beyond: Acting and sensing the body. *Neuropsychologia*, 2010, *48*, 763-773.

Borghi, A.M., & Cimatti, F. (under review). Words are not just words – The social acquisition of abstract words.

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

Crystal, D. (1995). *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, *128*, 615–627.

Desai, R.H., Binder, J.R., Conant, L.L., & Seidenberg, M.S. Activation of sensory-motor areas in sentence comprehension. *Cerebral Cortex*, *20*, 468-78.

Dove, G. (2009). Beyond Conceptual Symbols. A Call for Representational Pluralism. *Cognition, 110*, 412-431.

Fodor, J.A. (1998). *Concepts. Where cognitive science went wrong*. Oxford University Press.

Harnad, S. (1990). The Symbol Grounding Problem. *Physica D: Nonlinear Phenomena*, *42*, 335-346.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: Mit Press.

Marocco, D., Cangelosi, A., Fischer, K., & Belpaeme, T. (2010) Grounding action words in the sensorimotor interaction with the world: experiments with a simulated iCub humanoid robot. *Frontiers in Neurorobotics*, *4*.

Menz, M., Scorolli, C., Borghi, A.M., & Binkofski, F. (submitted). An fMRI study on abstract and concrete sentences processing.

Mirolli, M., & Parisi, D. (in press) Towards a Vygotskyan Cognitive Robotics: The Role of Language as a Cognitive Tool. *New Ideas in Psychology*.

Paivio, A., Yuille, J.C., & Madigan, S.A. (1986). Concreteness, imagery and meaningfulness values for 925 words. *Journal of Experimental Psychology*, *Monograph Supplement*, *76*, 3, part 2.

Sabsevitz, D.S., Medler, D.A., Seidenberg, M.,& Binder, J.R. (2005). Modulation of the semantic system by word imageability. *Neuroimage*, *27*, 188-200.

Scorolli, C., Binkofski, F., Buccino, G., Nicoletti, R., Riggio, L., & Borghi, A.M. (submitted). Abstract and concrete sentences, Embodiment and languages.

Scorolli, C., & Borghi, A. (2007). Sentence comprehension and action: Effector specific modulation of the motor system. *Brain research*, *1130*, 119-124.

Steels, L., (2003) Evolving grounded communication for robots. *Trends in Cognitive Sciences*, *7*, 308-312.

Sugita, Y., & Tani, J. (2005) Learning Semantic Combinatoriality from the Interaction between Linguistic and Behavioral Processes. *Adaptive Behavior, 13*, 211 - 225.

Tikhanoff ,V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L., & Nori, F. (2008) An open-source simulator for cognitive robotics research: The prototype of the iCub humanoid robot simulator. *Proceedings of IEEE Workshop on Performance Metrics for Intelligent Systems Workshop (PerMIS08)*.

Wauters, L.N., Tellings, A.E.J.M., Van Bon, W.H.J., & Van Haaften, A.W. (2003). Mode of acquisition of word meanings: The viability of a theoretical construct. *Applied Psycholinguistics*, *24*, 385-406.

Wiemer-Hastings, K., Krug, J., & Xu, X. (2001). Imagery, context availability, contextual constraint and abstractness. In J. D. Moore and K. Stenning (Eds.), *Proceedings of the XXIII Annual Conference of the Cognitive Science Society*. Edinburgh, Scotland, August 1-4, 2001 (p. 1106-1111). Mahwah, NJ: Lawrence Erlbaum Associates.

Wittgenstein, L. (1953). *Philosophische Untersuchungen*, Blackwell, Oxford.

Wu, L.L, & Barsalou, L.W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*, 173-189.