

# How to assess abstract conceptual knowledge: construction, standardization and validation of a new battery of semantic memory tests

Pasquale Anthony Della Rosa, PhD<sup>a</sup>  
Eleonora Catricalà, PhD<sup>b,c</sup>  
Silvia De Battisti, MSc<sup>b</sup>  
David Vinson, PhD<sup>d</sup>  
Gabriella Vigliocco, PhD<sup>d</sup>  
Stefano F. Cappa, MD<sup>b,e</sup>

<sup>a</sup> Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Milan, Italy

<sup>b</sup> Vita-Salute San Raffaele University, Milan, Italy

<sup>c</sup> Laboratory of Neuropsychology, The Foundation of the Carlo Besta Neurological Institute, IRCSS, Milan, Italy

<sup>d</sup> Research Department of Cognitive, Perceptual, and Brain Sciences, University College London, London, UK

<sup>e</sup> Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy

Correspondence to: Pasquale Anthony Della Rosa  
E-mail: dellarosa.pasquale@hsr.it

## Summary

**The neuropsychological investigation of semantic memory has mainly focused on concrete concepts, while abstract concepts have been relatively neglected. We describe a new battery for assessing abstract concepts in brain-damaged patients. The battery includes three different tests: an association task, a multiple-choice naming-to-description task and a sentence completion task.**

**The three tasks are based on the same 40 stimuli belonging to different categories of abstract concepts and they are tightly controlled for variables that can account for quantitative differences between abstract concepts (i.e. concreteness, imageability, context availability, familiarity, age of acquisition, mode of acquisition, emotional valence and arousal). The three tasks showed high reliability. Normative data were collected from 108 healthy Italian adults. To assess its sensitivity, the battery was administered to 13 patients with probable Alzheimer's disease who performed worse than matched controls. Significant correlations were also found between the tests and other semantic memory tests, supporting the validity of the battery.**

*KEY WORDS: abstract concepts, battery of tests, semantic memory, standardization, validity*

## Introduction

Concrete concepts refer to persons, places or things that can be perceived and experienced, while abstract concepts are generally intangible and are communicated through words. They can refer to our knowledge of emotions, social relations, cognitive states, personality traits or ideas (Barsalou and Wiemer-Hastings, 2005). In the neuropsychological literature, the majority of case studies report superior performances for concrete concepts over abstract ones after brain damage (Howes and Geschwind, 1964; Goodglass et al., 1969; Coltheart, 1980; Coltheart et al., 1980; Bub and Kertesz, 1982; Rissenberg and Glanzer, 1987; Katz and Goodglass, 1990; Martin and Saffran, 1992; Franklin et al., 1994, 1995). However, an advantage of abstract concepts over concrete ones has also been reported, predominantly in patients with damage to the temporal lobes as a consequence of semantic dementia (Warrington, 1975; Breedin et al., 1994; Cipolotti and Warrington, 1995; Reilly et al., 2006; Reilly et al., 2007) or herpes encephalitis (Warrington and Shallice, 1984; Sirigu et al., 1991).

According to the "dual coding" theory (Paivio, 1971, 1986), the representation of concrete concepts benefits from both sensory-motor and verbal information, while abstract concepts are only verbally represented. This implies that the extent to which a concept is represented as concrete or abstract is determined by the promptness through which a word is able to evoke a sensory-motor mental image (i.e. imageability) (Paivio, 1971) or by the availability of contextual information connected with the concept (Schwanenflugel and Shoben, 1983).

A recent meta-analysis of data from patients included in 19 functional magnetic resonance imaging and positron emission tomography studies (Wang et al., 2010) suggested that distinct brain systems also underlie this distinction between concrete and abstract concepts, with a verbal system being more involved in processing abstract concepts and a perceptual system more engaged in processing concrete ones. It can be argued that abstract concepts may also have non-verbal characteristics. For example, an emotion concept such as love is defined as an abstract concept and inferred through verbal referents (e.g., the deep relationship between two human beings), but it can also be conveyed through non-verbal experiences (e.g., a kiss) (Della Rosa et al., 2010) and/or non-verbal affective experience (Wiemer-Hastings and Xu, 2005; Kousta et al., 2011).

Different categories within the abstract domain can also be identified. Some investigations (Altarriba et al., 1999; Setti and Caramelli, 2005) suggest, for

example, that emotion words can be grouped as an independent category separate from other abstract concepts (Altarriba and Bauer, 2004; Altarriba et al., 1999; Setti and Caramelli, 2005). Other studies instead tried to infer differences between abstract concepts considering the abstract domain of knowledge as a continuum. For instance, Della Rosa et al. (2010) documented that different levels of abstractness could be predicted by the mode of acquisition of a concept. Similarly, Kousta and co-workers (2009, 2011) highlighted the important role of affective information in the representation of abstract concepts, and proposed that affective associations should be considered a continuous variable encompassing words of all types rather than a variable identifying only emotion words. Currently there are only a few tests available for assessing the abstract domain of knowledge. Most semantic memory tests include only concrete concepts or do not specifically assess abstract concepts (Hodges et al., 1992; Laiacona et al., 1993; Moreno and Cañamón, 2005; Adlam et al., 2010; Catricalà et al., 2013; Savage et al., 2013). The only standardized tests that allow the assessment of abstract knowledge are the Concrete and Abstract Word Synonym Test (Warrington et al., 1998), a comprehension task in which the subject is asked to choose which of two options is semantically similar to the target item, and the Standardized Comprehension Test for Abstract Words (Uno et al., 2003), an abstract spoken word-to-picture matching task. In Italy, there exists only one standardized test (Novelli et al., 1986) assessing naming to verbal definition for both concrete and abstract words. Moreover, several studies assessing the abstract domain in patients have used tests created ad hoc, which typically did not control for all the variables that can influence performance (Yi et al., 2007; Jefferies et al., 2009). The purpose of this paper is to provide a new battery of tests for assessing abstract knowledge, which considers the possible role both of “categories” of abstract concepts and of all those variables that may account for differences in perceived abstractness and can explain a subject’s performance on neuropsychological evaluation. The battery therefore controls for those measures that have been specifically proposed to play a role in explaining variability within the abstract domain of knowledge, such as abstractness, mode of acquisition (Della Rosa et al., 2010), emotional valence and arousal (Kousta et al., 2009, 2011), as well as for those assumed to account for differences in performance between concrete and abstract concepts, such as concreteness, imageability, context availability, familiarity and age of acquisition (Paivio, 1986; Schwanenflugel et al., 1992; Altarriba et al., 1999).

## Materials and methods

### Stimuli

#### CROSS-LINGUISTIC ITEM MATCHING

We selected an initial set of 60 words from the Della

Rosa et al. (2010) database of 417 Italian words, and their English equivalents taken from a list of 1,975 nouns extracted from the MRC database (Coltheart, 1981). The idea was to select a set of test items which could potentially be used to build different tests in either Italian or English in order to compare patient performances between languages on a completely equivalent and cross-linguistically matched set of items.

For both sets of stimuli (English and Italian) the 60 items were divided into five categories: Traits (e.g. *weakness*) (CAT1), Actions (e.g. *seduction*) (CAT2), Emotions (e.g. *fear*) (CAT3), Social concepts (e.g. *friendship*) (CAT4) and Cognitions (e.g. *ideal*) (CAT5). Each category comprised 12 items and each item’s membership of one of the five categories was decided according to the classification, based on lexico-semantic relationships between concepts, used by the cross-linguistic database MultiWordNet (multiwordnet.fbk.eu).

All the items in each category were divided into three concreteness ranges (low, medium, and high), corresponding to the 25<sup>th</sup>, 50<sup>th</sup> and 75<sup>th</sup> percentiles of the distribution of the concreteness scale values in Italian and English. The items were matched cross-linguistically in a pairwise fashion for the values of concreteness, extracted from the Della Rosa et al. (2010) database for the Italian items and from the MRC database for the English words. A matched pairs test on the mean within-pairs differences in concreteness values between the two languages for each category was carried out. No differences arose for concreteness values between Italian and English items in all categories (CAT1:  $p=.15$ ; CAT2:  $p=.12$ ; CAT3:  $p=.07$ ; CAT4:  $p=.09$ ; CAT5:  $p=.42$ ). In addition, two one-way ANOVAs were carried out on familiarity values for Italian items (extracted from the Della Rosa et al. 2010 database) and for English items (extracted from the MRC database) to assess differences between the five categories. There was no significant difference for familiarity between categories in Italian ( $F(4,55)=.90$ ;  $p=.46$ ) or English ( $F(4,55)=.91$ ;  $p=.47$ ).

#### SELECTION OF ITALIAN ITEMS

This paper focuses on the construction of an Italian battery of semantic memory tests for the assessment of abstract knowledge.

To this end, the Italian stimuli were first submitted to an additional norming procedure, in which 30 subjects were asked to indicate the category or categories to which they believed a concept belonged. To confirm a concept’s membership of a specific category as determined by the MultiWordNet database (multiwordnet.fbk.eu), we verified the correspondence with the category indicated by the majority of subjects for that specific concept.

Second, in order to control for the effect of all quantitative variables, which are assumed to affect performance on both abstract and concrete items, 40 Italian items (8 per category) were then selected from the original 60 for inclusion in the final Italian version of

the battery (see table I for the complete list). A one-way ANOVA for each variable (comparing the five categories) was carried out in order to verify whether the five categories were closely matched for all the variables of interest. The categories proved to be balanced for concreteness ( $p=.732$ ), imageability ( $p=.523$ ), context availability ( $p=.502$ ), familiarity ( $p=.848$ ), age of acquisition ( $p=.883$ ), mode of acquisition ( $p=.453$ ), abstractness ( $p=.614$ ), and number of letters ( $p=.941$ ) (Della Rosa et al., 2010).

Furthermore, in order to control for the potential impact of affective information on the processing of abstract concepts (Kousta et al., 2009), we collected norms for both emotional valence and arousal variables using the 417 words of the Della Rosa et al. database (2010).

Twenty-two subjects (mean age 24.14 years, SD 2.12; 10 males) were asked to rate concepts on nine-point scales of emotional valence and arousal. The set of 417 words was randomized into four different rating lists. The instructions were translated and adapted from those of ANEW (Affective Norms for English Words, Bradley and Lang, 1999). Emotional valence ranged from pleasant (represented by a happy face) to unpleasant (frowning face); arousal ranged from excited (face with wide open eyes) to calm (sleepy face). The 40 items were found to be matched between categories for arousal ( $p=.371$ ), but not for emotional valence ( $p=.036$ ). Post-hoc Bonferroni-corrected comparisons revealed that CAT3 was significantly different only from CAT4 ( $p=.047$ ). Emotions (CAT3) tended to have lower values for emotional valence (mean 3.95; SD 2.89) compared with the Social concepts category (CAT4) (mean 6.79; SD 1.29), given that 5/8 concepts (e.g., fear) in this category referred to negative emotions rated with lower values of emotional valence. However, CAT3 did not differ from the other categories in emotional valence (all,  $p>.19$ ).

Table I - List of the 40 stimuli used in the battery in alphabetical order with English equivalents taken from the MRC database.

affermazione (affirmation)	libertà (liberty)
amicizia (friendship)	logica (logic)
amore (love)	mistero (mystery)
civiltà (civilization)	noia (boredom)
debolezza (weakness)	pace (peace)
definizione (definition)	paura (scare)
descrizione (description)	peccato (pity)
differenza (difference)	pericolo (danger)
discussione (discussion)	protesta (protest)
esasperazione (exasperation)	reputazione (reputation)
esitazione (hesitation)	risentimento (resentment)
giustizia (justice)	seduzione (seduction)
ideale (ideal)	sfida (challenge)
illusione (illusion)	sicurezza (safety)
importanza (importance)	soddisfazione (satisfaction)
inesperienza (inexperience)	sogno (dream)
innocenza (innocence)	sospetto (suspicion)
ira (wrath)	talento (talent)
ironia (irony)	tristezza (woe)
istinto (instinct)	vendetta (revenge)

## BATTERY DESCRIPTION

The battery based on the 40 stimuli described above includes the following three tests: an Association Task (AT), a Sentence Completion Task (SCT), and a multiple-choice, Naming-to-Description Task (NDT). For all tasks the stimuli are presented both in written and spoken form. The sequence of presentation of the stimuli is randomized for each task.

### Test 1: Association Task

In this task, which is similar to the Pyramids and Palm Trees Test (Howard and Patterson, 1992), the subjects are asked to choose the item most strongly associated with the presented stimulus.

A stimulus is presented together with three words: a target and two distractors. The subject is required to match each stimulus item (e.g. friendship) with the word (out of the choice of three) that is most closely associated with it. The three options include: a target word with high association strength (e.g. bond), a distractor with low association strength (e.g. embrace), and a distractor that belongs to another category (e.g. color).

In order to obtain association norms, 30 participants were asked to write the first three words that came to mind that were meaningfully related to or strongly associated with the item. The word reported by the highest number of participants [mean 9.25;  $p$  (between categories) = .162] was selected as the target having high association strength with the item. A target-associated word that was reported by only one participant was selected as the related distractor. The second distractor, not related to the target, was a word never cited as a word associated with the item in question. This word was, instead, one showing low association strength with another item belonging to another category. We made sure that the 'other' categories providing distractors for items belonging to a given category were all equally represented.

We also controlled for differences in familiarity values between the target and two distractors. Given the absence of familiarity norms for all the words used in the association task, we extracted familiarity values for most of the words from an Italian database (Della Rosa et al., 2010) and an English database (Coltheart, 1981), which have been shown to be strongly correlated ( $r=.62$ ) (Della Rosa et al., 2010).

Twenty-two words had familiarity values in each of the two databases, and we first performed a correlation analysis which confirmed a very strong correlation ( $r=.73$ ) between the two databases.

We then performed an ANOVA on familiarity values for the English translations, which showed no differences between targets, distractors with high association and distractors with low association ( $p=.61$ ). It must be acknowledged that the MRC database did not contain familiarity values for nine targets, eight distractors with high association and seven distractors with low association.

In order to address any potential frequency differences specifically related to Italian items, which might somehow impact on choices made on the basis of familiarity, we also applied a measure of the spoken frequency of the Italian language obtained from the 'Banca dati dell'italiano parlato (BADIP)' (<http://badip.uni-graz.at/it>). No differences between the three categories of items (targets, high- and low-association distractors) were detected ( $p=.105$ ) in this regard.

The rankings of the targets and distractors (i.e., as first, second or third choices) were balanced within and between the categories. One point is given for each correct response (range 0-40).

### Test 2: Sentence Completion Task

In this task subjects are asked to complete 40 sentences in which the final word is missing.

For example, for the item "pace" (peace), the following sentence is used "Alla fine della guerra, i due paesi hanno firmato un trattato di..." (English equivalent: At the end of the war, the two countries signed an agreement to keep the ...).

The sentences were created using definitions and phrases taken from two different Italian dictionaries (De Mauro, 2000; Garzanti, 1998) in order to exert higher levels of contextual constraint on the probability of the target word completing the sentence frame. A one-way ANOVA showed that the mean sentence length (mean number of words 12.95;  $p=.195$ ) was matched between categories.

One point is given for each correct target word produced (range 0-40). In addition, responses other than the target are classified, according to their relationship with the item, as synonyms, semantically related, semantically related but contextually inappropriate, contextually suitable but not related to the target, opposite, circumlocution, repetition of words or phrase just spoken, anomia or other.

### Test 3: Multiple-choice, Naming-to-Description Task

In this task subjects are asked to select a word, from a choice of four, that best matches a verbal definition (see Yi et al., 2007 for a similar task). Forty definitions were created, and for each definition a target word (TW) and three foils were presented: one foil semantically related to the target word (SRTW), one with an opposite meaning to that of the target word (OMTW), and one semantically related to the opposite meaning (SROTW). For example, for the TW "mistero" (mystery) we used the definition "un evento che non si riesce a spiegare chiaramente e razionalmente" (English translation: an event that cannot be clearly and rationally explained). For this definition the SRTW was "complotto" (plot), the OMTW "soluzione" (solution) and the SROTW "identificazione" (identification). Given the absence of tools for assessing semantic relatedness and neighborhood between concepts in Italian, definitions were first created in English using

four British English dictionaries (Cambridge Dictionaries Online, Wordsmyth Advanced Dictionary, Encarta World English Dictionary and The Merriam-Webster on-line English dictionary) and two North American ones (OneLook® Dictionary and The Free Dictionary). Each definition was given to a group of ( $n=28$ ) native English subjects who were asked to produce a target word. Only definitions with a cloze probability value above .4 were included in the final list. Antonyms and synonyms of the target were selected using the same dictionaries. The strength of the semantic relationship between the meaning of the TW and the SRTW was computed through latent semantic analysis (Landauer and Dumais, 1997) in the "General Reading up to 9<sup>th</sup> Grade" semantic space with 300 dimensions and term-to-term comparisons. Semantic neighborhood values were matched between the categories ( $F(4,55) = .86$ ,  $p=.49$ ) (mean .257; SE .0227). The same analysis was performed on the OMTW and the SROTW in order to match the strength of the relationship between the opposite and the foil between all categories ( $F(4,55) = .44$ ,  $p=.77$ ) (mean .264; SE .0229). The Italian definitions were then adapted from the English ones, according to two Italian dictionaries (De Mauro, 2000; Garzanti, 1998) and were modified into everyday Italian. Definitions were kept as short as possible and, to reduce the likelihood that synonyms or related words would be produced, these were included in the definitions where possible. The number of words for the definition of each item was balanced between categories ( $p=.214$ ). Finally we used the same two Italian dictionaries to select the synonyms and the opposites of the target. One point is given for each correct response (range 0-40).

### INTERNAL CONSISTENCY AND INTER-RATER RELIABILITY

To determine the internal consistency and reliability of our measures (40 items) for each of the three tasks (AT, SCT and NDT) included in our battery, named DeCABS (Della Rosa **C**atricalà **A**BSTRACT battery), Spearman-Brown split-half coefficients based on a dichotomy of odd-even items for each of the five sub-categories (CAT1, CAT2, CAT3, CAT4, CAT5) included in each test were computed. The coefficients were computed on an initial sample of 102 healthy native Italian participants. The Spearman-Brown corrected odd-even split-half reliability was .806 for AT, .886 for SCT and .783 for NDT, demonstrating the high reliability of our measures across all three tasks.

Relative inter-rater reliability was assessed for the three tasks using the intraclass correlation coefficient (ICC) two-way mixed model and an absolute agreement definition (McGraw and Wong, 1996). We chose the absolute agreement definition of reliability as a more stringent criterion accounting for the tendency of participants ( $n=102$ ) to give the same response when they respond to the same item. ICCs were calculated for reliability statistics and their strength was interpreted using Munro's classification (Munro, 2005), in which 0-0.25 = little if any, 0.26-0.49 = low, 0.50-0.69

= moderate, 0.70-0.89 = high, and 0.9-1.0 = very high correlation. The ICC values across all participants were .830 (95% confidence interval [CI]: .746, .897) for the AT; .956 (95% CI: .934, .973) for the SCT; and .875 (95% CI: .812, .926) for the NDT. The ICC values for all three tests ranged from high to very high.

**Standardization of the DeCabs battery**

**SUBJECTS**

A total of 108 healthy native Italian subjects (54 females) took part in the study. Their mean age was 54.31 years (SD 17.31, range 25-84 years) and their mean duration of education was 11.52 years (SD 4.23; range 5-20). Table II presents the distribution of the demographic data. Subjects with past or present neurological or psychiatric illnesses or a corrected score of less than 24 on the Mini-Mental State Examination (MMSE, Folstein et al., 1975) were excluded. All volunteers completed a written consent form to participate in this study, which was reviewed and approved by the Ethics Committee of the San Raffaele Hospital and performed in accordance with the Declaration of Helsinki II.

**PROCEDURES**

The battery was administered to all participants in a single session. The Sentence Completion Task was administered as the first task, while the order of administration of the other two tests was randomized across the subjects. For all tests the stimuli were presented both visually and orally. All tests were scored as described above.

**DATA ANALYSES**

For each test, different simple linear regression analyses were performed in order to assess which demographic variables, i.e. age, years of education (or their transformations) and gender, were to be included in the final models as more effective in reducing the residual variance. Multiple regression analyses, which included only those variables found to be significant on the previous simple regression analyses, were then carried out in order to generate prediction equations. For each test score we obtained correction coefficients from the prediction equations. Correction grids were also derived by reversing the signs of the coefficients in order to adjust the original score by adding or subtracting the contributions of the significant variables (Table III). This stan-

Table II - Distribution of demographic data for the 108 subjects.

Education	Age						Tot F/M	Tot
	25-34	35-44	45-54	55-64	65-74	75-85		
5-8 yrs	1/1	3/3	2/1	2/2	7/5	5/5	20/17	37
9-13 yrs	1/2	4/4	6/5	5/5	2/3	2/3	20/22	42
14-20 yrs	6/6	2/2	3/3	1/2	1/2	1/0	14/15	29
Tot F/M	8/8	9/9	11/9	8/9	10/10	8/8	54/54	108
Tot	17	18	20	17	20	16	108	

\*The number of participants in each cell is reported as females / males.

Table III - Age, gender and education adjustment grids for the three tests.

a) Association task													
EDUCATION	5	8	10	13	17								
	2.18	1	0.33	-0.55	-1.58								
Corrected score = Raw score - [1.993 × (√EDUCATION - 3.33)]													
b) Sentence completion task													
EDUCATION	5	8	10	13	17								
	5.4	2.47	0.83	-1.36	-3.91								
Corrected score = Raw score - [4.932 × (√EDUCATION - 3.33)]													
c) Multiple-choice, naming-to-description task													
EDUCATION	AGE												
	25	30	35	40	45	50	55	60	65	70	75	80	85
5	3.32	3.59	3.86	4.13	4.40	4.67	4.94	5.21	5.48	5.75	6.02	6.29	6.56
8	0.66	0.93	1.20	1.47	1.74	2.01	2.28	2.55	2.82	3.09	3.36	3.63	3.90
10	-0.83	-0.56	-0.29	-0.02	0.25	0.52	0.79	1.06	1.33	1.60	1.87	2.14	2.41
13	-2.82	-2.55	-2.28	-2.01	-1.74	-1.47	-1.20	-0.93	-0.66	-0.39	-0.12	0.15	0.42
17	-5.14	-4.87	-4.60	-4.33	-4.06	-3.79	-3.52	-3.25	-2.98	-2.71	-2.44	-2.17	-1.9
Corrected score = Raw score - [-0.054 × (AGE - 54.31)] - [4.48 × (√EDUCATION - 3.33)]													

standardization procedure has been widely applied (e.g. Frasson et al., 2011) and the detailed methodological implementation is described in detail elsewhere (Capitani, 1987; Capitani and Laiacona, 1997). Finally, we classified the adjusted scores into five categories (from 0 to 4) according to the Equivalent Scores method (Capitani, 1987; Capitani and Laiacona, 1997) in order to allow comparison of performances on different tests along the same scale of values, once the influence of the demographic variables has been accounted for (for a detailed description of this method see Capitani and Laiacona, 1997). The first region (score 0) falls outside the outer unidirectional non-parametric tolerance limit (normality threshold) with a confidence of 95% (the second observation for the population of 108 subjects). At the other extreme, there is a region in which the adjusted scores lying between the median and best score (score 4) fall. Between these two regions, three other regions, equispaced between the normality threshold and the median, can be differentiated. In this way, performance on each test can be coded as an adjusted score falling on a new scale of five ordered performance levels, namely Equivalent Scores, where 0 corresponds to not normal or pathological performances, 4 to performances equal to mean or superior values, and 1, 2 and 3 to intermediate performances.

## RESULTS

The multiple linear regressions analyses revealed that while the AT and the SCT were influenced only by education [respectively,  $F(1,107)=10.634$ ,  $p<.005$ ,  $t=3.261$ ,  $p<.005$ ;  $F(2,107)=35.037$ ,  $p<.001$ ,  $t=6.515$ ,  $p<.001$ ], with better scores recorded for subjects with higher levels of education, the NDT was influenced by both education and age [ $F(2,107)=37.941$ ,  $p<.001$ ,  $t=6.246$ ,  $p<.001$ ;  $t=-2.370$ ,  $p<.05$ ], with better scores recorded for younger and more highly-educated subjects. The adjustment grids and equivalent scores for each task are reported, respectively, in tables III (a,b,c) and IV.

### **Sensitivity and validity of the DeCAbs battery**

In order to assess the sensitivity of the tests, we administered the complete Italian battery to a group of probable Alzheimers' disease (pAD) patients and to a group of matched controls.

## SUBJECTS

Thirteen patients with a diagnosis of pAD (4 males; mean age 74.85 years, SD 7.27; mean education 7.92 years, SD 2.4) and 13 healthy individuals from the previous normative group (4 males; mean age 74.15 years, SD 8.73; mean education 7.92 years, SD 4.07) were enrolled. The pAD patients were matched with the normal elderly controls for age ( $p=.829$ ) and education ( $p=1$ ).

The diagnosis of pAD was made according to the criteria developed by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's disease and Related Disorders Association (ADRDA). The pAD group included subjects with mild to moderate levels of dementia (uncorrected MMSE range: 16-24). All the pAD patients were recruited from the San Raffaele Hospital.

The Ethics Committee of the San Raffaele Hospital approved the study, which was conducted in accordance with the Declaration of Helsinki II. Written informed consent was obtained from each patient.

Both the patients and the controls were administered the complete battery of tests (with the exception of one pAD patient, who did not perform the AT) and three tests included in the Italian battery for the assessment of semantic memory disorders for concrete concepts (Catricalà et al., 2013) (i.e. picture naming, naming in response to an oral definition and sentence verification task).

## DATA ANALYSIS

A t-test was used to assess the differences between the pAD patients and the controls on all three tests. Furthermore, the validity of the three tests was evaluated by considering, for all subjects ( $n=26$ ), the correlation of scores on each DeCabs task with performance on the three tests included in the Italian battery for concrete concepts (Catricalà et al., 2013).

## RESULTS

The pAD patients showed significantly impaired performances with respect to the control subjects on all three tests for abstract concepts (Table V).

Overall, the correlations between scores on the three DeCAbs tests and the three tests included in the

Table IV - Equivalent scores for the three tasks included in the DeCAbs battery.

Equivalent scores	Association Task	Sentence Completion Task	Multiple-choice, Naming-to-Description Task
0	≤ 26.18	≤ 13.40	≤ 24.02
1	26.19 – 29.78	13.41 – 16.83	24.03 – 27.78
2	29.79 – 33.39	16.84 – 20.28	27.79 – 31.54
3	33.40 – 37.00	20.29 – 23.72	30.55 – 35.51
4	≥ 37.01	≥ 23.73	≥ 35.32

Italian battery for concrete concepts (Catricalà et al., 2013) were high and significant for all three tests. The list of correlation coefficients is reported in Table VI.

### Discussion

This paper describes the development and standardization of an Italian battery of three semantic memory tests for abstract words. Normative data showed that education influenced performance on all three tasks, while age contributed to the prediction of performance only on one task (NDT). Furthermore, all three tests were found to be highly reliable and the battery was shown to be sensitive to impairment in the abstract domain of knowledge in a group of pAD patients, given the lower performances of these patients compared with a group of normal controls matched for age, education and gender. The validity of the three tests was also assessed through correlations with other tests included in the Italian battery for the assessment of semantic memory deficits (Catricalà et al., 2013). High correlations were found between each of the three tasks and the tests of the Italian battery for concrete concepts (Catricalà et al., 2013).

The main aim of the battery is to provide a new, comprehensive tool for investigating impairments in the domain of abstract knowledge in different neurological conditions taking into account both categories of stimuli and other variables that could explain subjects' performances.

As detailed below, there are several novel aspects to this work.

First, well-supported empirical evidence for categories of abstract concepts is still lacking, given that, to date, dissociations have been reported only between abstract and concrete concepts (Warrington, 1975) and, in the concrete domain, between distinct categories of living (animals, fruits and vegetables) and non-living entities (tools, vehicles, furniture) (for a review see Capitani et al, 2003).

The idea of a categorical organization underlying the representation of abstract knowledge has been suggest-

ed, although it is mainly supported by studies on healthy subjects, with, as yet, no clear evidence reported in patients (Martin and Fedio, 1983; Hsieh et al., 2012). Altarriba and Bauer (2004) indicated emotions as a peculiar category, as they were found to be less concrete and to show greater imageability and context availability than other abstract words. Setti and Caramelli (2005) tried to provide evidence for the existence, also, of other categories in the abstract domain. They considered four categories of abstract knowledge (emotions, cognitive processes, states of self and nominal kinds) and showed that they differ with regard to the concreteness, imageability, context availability and abstractness dimensions. Similarly, Ghio et al. (2013) traced boundaries between three different categories: mental state, emotion and mathematics related meanings. All these studies seem to converge in showing the existence of at least one distinct category in the abstract domain, that of emotion words. In the same direction, Kousta and co-workers (2009, 2011; see also Vigliocco et al., 2013) highlighted the important role of affective information in the representation of abstract concepts, proposing, however, that affective associations (emotional valence and arousal) should be considered a continuous variable encompassing words of all types rather than a variable referring only to emotion words.

Evidence for abstract categories deriving from studies of patients showing semantic memory impairments is instead poor or controversial. For instance, Martin and Fedio (1983) reported preserved comprehension of abstract words with respect to other categories (objects, actions, modifiers) in a group of AD patients. However, more recently, Hsieh et al. (2012) reported no differences between emotion words, concrete and abstract concepts in AD patients, who performed similarly to controls on all categories.

In this battery of tests we outlined five different categories of abstract concepts. However, it needs to be stated here that we do not claim that categories of abstract concepts can be clearly delineated and distinguished as concrete ones can. We acknowledge that

Table V - Means (standard deviations) and statistical analyses, for each group, of the three tests comprising the battery

Test	Controls Mean (SD)	pAD Mean (SD)	t-value	p-value
Association Task	35.69 (4.42)	29.75 (6.81)	2.608	.016
Sentence Completion Task	21.38 (3.66)	12.31 (4.59)	5.573	<.001
Multiple-choice, Naming-to-Description Task	32.46 (6.39)	20.85 (8.58)	3.913	.001

Table VI - Spearman-rho correlation matrix for the performances of the 26 subjects (i.e. patients and controls) on the three abstract tests and on the three concrete tests.

	Naming of colored photographs	Naming in response to an oral description	Sentence verification
Association Task	.575**	.763**	.592**
Sentence Completion Task	.755**	.836**	.783**
Multiple-choice, Naming-to-Description Task	.578**	.743**	.640**

\*\* Correlation significant at  $p=.01$

there may be more overlap between abstract categories, however this does not necessarily exclude the possibility that at least some categories of abstract concepts may be structured in a hierarchical way.

Therefore, this battery could probably be used to establish whether or not there are some categories of abstract concepts that are more impaired than others in different pathological conditions, controlling for most of the variables that can account for differences in subjects' performances, and also for the fact that controls do not perform at ceiling. In addition the battery could be used together with a battery assessing concrete concepts (Catricalà et al., 2013) in order to assess the aforementioned abstract-concrete dissociation (Warrington, 1975).

Second, in this battery of tests, the stimuli were controlled for several variables as such as concreteness, familiarity, imageability, context availability, age of acquisition, mode of acquisition, emotional valence and arousal in order to investigate the possible role of these variables in predicting overall performance and differences between subsets of items with specific characteristics on each scale considered.

Third, the three tasks included in the battery allow investigation of impairments in the abstract domain of knowledge through both comprehension and production tasks, while using the same set of stimuli (i.e. AT and NDT – comprehension; SCT – production) to establish the presence of a true semantic deficit. However, it must be acknowledged that at least two factors should be carefully considered when evaluating patients' performances on the single tasks. First, the material included in each task is required by the nature of abstract conceptual knowledge to be exclusively verbal (i.e. words or sentences); second, the presence of cognitive deficits in other domains may influence, or even cause, poor performances on some of the tasks in the battery, for example, those requiring more linguistic and executive resources, as in the case of sentences that include more than a single meaning needing to be computed and kept temporarily available in order to give a response.

In conclusion, we think that this new battery of tests may allow inferences to be drawn about abstract representations through both the evaluation of a selective impairment/advantage for items belonging to a specific category or for subsets of items with similar characteristics in relation to specific variables. In this way, the assessment of abstract knowledge in patients with semantic memory impairments could make an important contribution to understanding how abstract knowledge is organized.

The complete battery of three tests and materials will be made available upon request on condition that it is not used for commercial purposes.

### Acknowledgments

This study was conducted under the supervision of Stefano F. Cappa and Gabriella Vigliocco and fully sup-

ported by a European Union (FP6-2004-NEST-PATH) grant (028714) (<http://www.abstract-project.eu>) to Stefano F. Cappa. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### References

- Adlam AL, Patterson K, Bozeat S, Hodges JR (2010). The Cambridge Semantic Memory Test Battery: detection of semantic deficits in semantic dementia and Alzheimer's disease. *Neurocase* 16: 193-207.
- Altarriba J, Bauer LM (2004). The distinctiveness of emotion concepts: a comparison between emotion, abstract and concrete words. *Am J Psychol* 117: 389-410.
- Altarriba J, Bauer LM, Benvenuto C (1999). Concreteness, context availability and imageability ratings and words associations for abstract, concrete and emotion words. *Behav Res Methods Instrum Comput* 31: 578-602.
- Banca dati dell'italiano parlato (BADIP). <http://badip.uni-graz.at/it>.
- Barsalou LW, Wiemer-Hastings K (2005). Situating abstract concepts. In: Pecher D, Zwaan R (Eds) *Grounding Cognition: The Role of Perception and Action in Memory, Language and Thought*. New York, Cambridge University Press, pp. 129-163.
- Bradley MM, Lang PJ (1999). *Affective Norms for English Words (ANEW)*. Instruction manual and affective ratings (Tech. Rep. No. C-1). Gainesville, FL, University of Florida, The Center for Research in Psychophysiology
- Breedin SD, Saffran EM, Coslett HB (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology* 11: 617-660.
- Bub D, Kertesz A (1982). Deep agraphia. *Brain Lang* 17: 146-165.
- Cambridge Dictionaries Online, <http://dictionary.cambridge.org>
- Capitani E (1987). Statistical methods. In: Spinnler H, Tognoni G (Eds) *Italian normative values and standardization of neuropsychological tests*. *Ital J Neurol Sci* 6 (Suppl. 8) 14-20.
- Capitani E, Laiacona M (1997). Composite neuropsychological batteries and demographic correction: standardization based on equivalent scores, with a review of published data. The Italian Group for the Neuropsychological Study of Ageing. *J Clin Exp Neuropsychol* 19: 795-809.
- Capitani E, Laiacona M, Mahon B, et al (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cogn Neuropsychol* 20: 213-261.
- Catricalà E, Della Rosa PA, Ginex V et al (2013). An Italian battery for the assessment of semantic memory disorders. *Neurol Sci* 34: 985-993.
- Cipolotti L, Warrington EK (1995). Semantic memory and reading abilities: a case report. *J Int Neuropsychol Soc* 1: 104-110.
- Coltheart M (1980). A right hemisphere hypothesis. In Coltheart M, Patterson KE, Marshall JC (Eds), *Deep Dyslexia*. London, Routledge and Kegan Paul, pp. 326-380.
- Coltheart M (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology* 33A: 497-505.
- Coltheart M, Patterson KE, Marshall JC (1980). (Eds) *Deep Dyslexia*. London, Routledge and Kegan Paul.
- De Mauro T. (2000) *Il dizionario della lingua italiana*, Turin, Paravia.



- Della Rosa PA, Catricalà E, Vigliocco G et al. (2010). Beyond the abstract-concrete dichotomy: mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. *Behav Res Methods* 42: 1042-1048.
- Encarta, Encarta World English Dictionary. London, Bloomsbury 1999.
- Franklin S, Howard D, Patterson K (1994). Abstract word meaning deafness. *Cognitive Neuropsychology* 11: 1-34.
- Franklin S, Howard D, Patterson K (1995). Abstract word anomia. *Cognitive Neuropsychology* 12: 549-566.
- Folstein MF, Folstein SE, McHugh PR (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 12: 189-198.
- Frasson P, Ghirelli R, Catricalà E, et al (2011). Free and Cued Selective Reminding Test: an Italian normative study. *Neurol Sci* 32: 1057-1062.
- Garzanti, Il Nuovo Dizionario Italiano. 1996. Garzanti editore, 1998.
- Ghio M, Vaghi MM, Tettamanti M (2013). Fine-grained semantic categorization across the abstract and concrete domains. *PLoS One* 8:e67090
- Goodglass H, Hyde MR, Blumstein S (1969). Frequency, pictureability and availability of nouns in aphasia. *Cortex* 5: 104-119.
- Hodges JR, Salmon DP, Butters N (1992). Semantic memory impairment in Alzheimer's disease: failure of access or degraded knowledge? *Neuropsychologia* 30: 301-314.
- Howard D, Patterson KE (1992). The Pyramids and Palm Trees Test: A test of semantic access from words and pictures. Thames Valley Test Company.
- Howes D, Geschwind N (1964). Quantitative studies of aphasic language. In: Riosch DM, Wenstein EA (Eds) *Disorders of Communication*. Baltimore, Williams and Wilkins, pp. 229-244.
- Hsieh S, Foxe D, Leslie F, et al (2012). Grief and joy: emotion word comprehension in the dementias. *Neuropsychology* 26: 624-630.
- Jefferies E, Patterson K, Jones RW, et al (2009). Comprehension of concrete and abstract words in semantic dementia. *Neuropsychology* 23: 492-499.
- Katz RB, Goodglass H (1990). Deep dysphasia: analysis of a rare form of repetition disorder. *Brain Lang* 39: 153-185.
- Kousta ST, Vigliocco G, Vinson DP, et al (2011). The representation of abstract words: why emotion matters. *J Exp Psychol Gen* 140: 14-34.
- Kousta ST, Vinson DP, Vigliocco G (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition* 112: 473-481.
- Laiacona M, Barbarotto R, Capitani E (1993). Perceptual and associative knowledge in category specific impairment of semantic memory: a study of two cases. *Cortex* 29: 727-740.
- Landauer TK, Dumais ST (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104: 211-240.
- Martin A, Fedio P (1983). Word production and comprehension in Alzheimer's disease: The breakdown of semantic knowledge. *Brain Lang* 19: 124-141.
- Martin N, Saffran EM (1992). A computational account of deep dysphasia: evidence from a single case study. *Brain Lang* 43: 240-274.
- McGraw KO, Wong SP (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1: 30-46.
- Moreno FJ, Cañamón S (2005). Presentation and preliminary results of Semantic Battery Nombela: A new instrument to evaluate semantic categorial impairment. *Spanish J Clin Psych* 10: 205-219.
- Munro BH (2005). *Statistical Methods for Health Care Research* (5th Ed.). Philadelphia, Lippincott Williams & Wilkins.
- MultiWordNet. URL: <http://multiwordnet.fbk.eu/english/home.php>
- Novelli G, Papagno C, Capitani E et al. (1986). Tre test clinici di ricerca e produzione lessicale: taratura su soggetti normali. *Archivio italiano di Neurologia, Psicologia e Psichiatria* 47: 477-503.
- OneLook® Dictionary. URL: <http://www.onelook.com>
- Paivio A (1971). *Imagery and Verbal Processes*. New York, Holt, Rinehart & Winston.
- Paivio A (1986). *Mental Representation: A Dual coding Approach*. New York, Oxford University Press.
- Reilly J, Cross K, Troiani V, et al. (2007). Single-word semantic judgements in semantic dementia: do phonology and grammatical class count? *Aphasiology* 21: 558-569.
- Reilly J, Grossman M, McCawley MC (2006). Concreteness effects in lexical processing of semantic dementia. *Brain Lang* 99: 147-148.
- Rissenberg M, Glanzer M (1987). Free recall and word finding ability in normal aging and senile dementia of Alzheimer's type: the effect of item concreteness. *J Gerontol* 42: 318-322.
- Savage S, Hsieh S, Leslie F, et al (2013). Distinguishing subtypes in primary progressive aphasia: application of the Sydney language battery. *Dement Geriatr Cogn Disord* 35: 208-218.
- Schwanenflugel PJ, Shoben EJ (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *J Exp Psychol Learn Mem Cogn* 9: 82-102.
- Schwanenflugel PJ, Akin C, Luh WM (1992). Context availability and the recall of abstract and concrete words. *Mem Cognit* 20: 96-104.
- Setti A, Caramelli N (2005). Different domains in abstract concepts. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, pp. 1997-2002.
- Sirigu A, Duhamel JR, Poncet M (1991). The role of sensorimotor experience in object recognition. A case of multimodal agnosia. *Brain*, 114: 2555-2573.
- The Free Dictionary. URL: <http://www.thefreedictionary.com>
- The Merriam-Webster on-line English dictionary. URL: <http://www.merriam-webster.com>
- Uno A, Haruhara N, Kaneko M (2003). *The Standardized Comprehension Test for Abstract Words*. Tokyo, Interuna Publishing.
- Vigliocco G, Kousta ST, Della Rosa PA, et al (2013). The neural representation of abstract words: the role of emotion. *Cereb Cortex* doi: 10.1093/cercor/bht025.
- Wang J, Conder JA, Blitzer DN, et al. (2010). Neural representation of abstract and concrete concepts: a meta-analysis of neuroimaging studies. *Hum Brain Mapp* 31: 1459-1468.
- Warrington EK (1975). The selective impairment of semantic memory. *Q J Exp Psychol* 27: 635-657.
- Warrington EK, Shallice T (1984). Category specific semantic impairment. *Brain* 107: 829-854.
- Warrington EK, McKenna P, Orpwood L (1998). Single word comprehension: a concrete and abstract word synonym test. *Neuropsychol Rehabil* 8: 143-154.
- Wiemer-Hastings K, Xu X (2005). Content differences for abstract and concrete concepts. *Cogn Sci* 29: 719-736.
- Wordsmyth Advanced Dictionary. URL: <http://www.wordsmyth.net>
- Yi HA, Moore P, Grossman M (2007). Reversal of the concreteness effect for verbs in patients with semantic dementia. *Neuropsychology* 21: 9-19.