Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 57 (2007) 101–125
Journal of Memory and Language
www.elsevier.com/locate/jml

Models of relevant cue integration in name retrieval ☆

Luigi Lombardi [a,*], Giuseppe Sartori [b]

[a] Università di Trento, Dipartimento di Scienze della Cognizione e della Formazione, via Matteo del Ben 5, I-38068 Rovereto (TN), Italy
[b] Università di Padova, Dipartimento di Psicologia Generale, Via Venezia 8, I-35100 Padova, Italy

Received 1 March 2006; revision received 17 October 2006
Available online 26 December 2006

Abstract

Semantic features have different levels of importance in indexing a target concept. The article proposes that semantic relevance, an algorithmically derived measure based on concept descriptions, may efficiently capture the relative importance of different semantic features. Three models of how semantic features are integrated in terms of relevance during name retrieval are presented. The models have been contrasted with empirical results from a naming-to-description task administered to three different groups of participants: young people, healthy elderly and semantically impaired Alzheimer patients. Predictions of the empirical results made by the models are used to provide a measure of identifiability or the extent to which the models can be distinguished from one another. In three studies we show that an additive-type rule is consistently superior to multiplicative rule and winner-take-all rule in predicting naming accuracy in a naming-to-description task. Finally, we investigated the implications of the integration rules for degraded semantic knowledge.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Semantic relevance; Integration rules; Semantic memory models; Model comparison; Semantic memory disorders; DAT; Controls; Mathematical models

Introduction

Many cognitive theories assume that concepts can be considered, at least in part, as either organized structures of semantic features (e.g., Collins & Quillian, 1969; Hinton & Shallice, 1991; McRae, de Sa, & Seidenberg, 1997; Smith & Medin, 1981) or points in a high-dimensional space (e.g., Landauer & Dumais, 1997; Shepard, 1957). The purpose of the present research is to investigate how semantic features are combined in the process of concept retrieval in a naming-to-description task by contrasting differing binding mechanisms in neurological patients with diagnoses of dementia of Alzheimer's type (DAT) with probable semantic memory disorders and healthy controls. Modeling how individuals integrate multiple sources of information has been a cornerstone in many theories of cognitive processes, including perception (e.g., Anderson, 1981; Massaro, 1990), categorization (e.g., Anderson, 1991; Kruschke, 1992; Nosofsky, 1986), and generalization (e.g., Shepard, 1987). However, the modeling of integration of semantic features in the name retrieval process of natural concepts has received considerably less attention. The question we address in this paper is how normal and semantically impaired individuals integrate two or more semantic features that may have different

☆ We thank three anonymous reviewers for helpful comments on an earlier draft of the paper. We are also grateful to Jay McClelland for having noticed the commonalities between the additive semantic relevance model and the rational model of classification proposed by J.R. Anderson.
* Corresponding author. Fax: +39 0464 483554.
E-mail address: luigi.lombardi@unitn.it (L. Lombardi).

0749-596X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2006.10.003

relevance values with respect to candidate concepts (in the process of naming).

Semantic features may trigger concept retrieval on the basis of their degree of informativeness for the target concept. A concept may have uncountable semantic features, although those which are really useful in discriminating it from other closely related concepts may not be numerous. The following is a telling example[1]: ⟨has two humps⟩ is a semantic feature of high relevance for the concept CAMEL, because most subjects use that feature to define it, whereas very few use the same feature to define other concepts. Instead ⟨has 4 legs⟩ is a semantic feature with lower relevance for the same concept, because few subjects use it to define CAMEL but do use it to define many other concepts. Among all the semantic features of a concept, those with high relevance are useful in discriminating the concept from those similar to it.

Semantic relevance has been recently proposed by Sartori and Lombardi (2004) as a measure which indexes the level of importance of a semantic feature to the meaning of a concept. Semantic relevance is the result of two components: (i) a local component, which measures the importance of the semantic features for the concept, which may be interpreted as *dominance*, and (ii) a global component, which measures the importance of the same semantic feature for all the other concepts in the lexicon, which may be interpreted as *distinctiveness*. Dominance scores high when the semantic feature is frequently mentioned in defining a concept (Ashcraft, 1978). In contrast, distinctiveness scores high when a semantic feature is used in defining few concepts (Devlin, Gonnerman, Andersen, & Seidenberg, 1998; Marques, 2005; Randall, Moss, Rodd, & Tyler, 2004). Finally, semantic relevance scores high when a semantic feature is both frequently mentioned in defining a concept (high dominance), but only mentioned in defining few other concepts (high distinctiveness). Unlike distinctiveness which is a general (global) property of a feature, semantic relevance is a dimension which is concept-dependent, since the relevance of a given semantic feature usually varies across the different concepts in the lexicon (see the mathematical formulation of semantic relevance in the next section). When a set of semantic features is presented to an individual, their overall relevance results from the integration of the individual relevance values associated with each of the semantic features. The concept with the highest integrated relevance is the one that is retrieved with higher probability. For example, the features ⟨has a beak⟩, ⟨similar to a goose⟩, and ⟨lives in ponds⟩ have top relevance for the concept DUCK, followed by SWAN, and then OSTRICH (see Fig. 1). The retrieved concept, given a description containing the above three features, will be DUCK, as it has the highest integrated relevance.

A number of investigations have been conducted in order to test the role of semantic relevance in concept retrieval. At behavioral level, Sartori, Lombardi, and Mattiuzzi (2005a) have shown that semantic relevance predicts accuracy in a naming-to-description task better than other well-known psycholinguistic parameters such as age-of-acquisition, familiarity, typicality, and frequency. Also in a feature-verification task, participants are faster and more accurate in deciding whether a high-relevance feature belongs to a concept when compared to low-relevance features (Sartori, Polezzi, Mameli, & Lombardi, 2005b; Sartori, Mameli, Polezzi, & Lombardi, 2006).

The role of relevance of features has also been studied in semantically impaired neurological individuals (both herpes and DAT patients). Specifically, given that the concepts belonging to the Living category tend to have on average lower total relevance,[2] when patients with category-specific impairments for the Living category are presented with descriptions of concepts having matched integrated relevance across categories of Living and Non-living concepts, the previously reported impairment for Living concepts disappears (Sartori & Lombardi, 2004). As regards the neural correlates of semantic relevance, a number of studies using ERPs and fMRI have been conducted. The ERPs of semantic relevance were investigated by means of the N400, which is a negative deflection that can be used to measure semantic incongruity (Kutas & Federmeier, 2000). It was found that the N400 is larger for low-relevance features rather than for high-relevance features (Sartori et al., 2005b). Furthermore, when descriptions of concepts belonging to either the Living or Non-living categories are equated for their integrated relevances, the seemingly category effect at neural level disappears in a feature-verification task. Likewise, when relevance is matched across sensory and functional features types a larger N400 previously reported for sensory features disappears (Sartori et al., 2006). These results clearly parallel the findings of the studies conducted at behavioral level reported above. Finally, the neural locus in which semantic relevance exerts its effect has been studied using fMRI in the case of higher order visual features. Mechelli, Sartori, Orlandi, and Price (2006) report that relevance modulates neuronal responses in the medial fusiform *gyrus* bilaterally, demonstrating that neuronal responses during concept retrieval are affected by the semantic relevance of the higher order visual features in a picture-naming task.

[1] Concept names are printed in uppercase (e.g., DOG) and names of semantic features in angled brackets (e.g., ⟨has a tail⟩).

[2] The total relevance for a concept is computed by integrating the relevance values of all the semantic features which are listed in defining the concept in a feature-to-listing task (Sartori et al., 2005a).

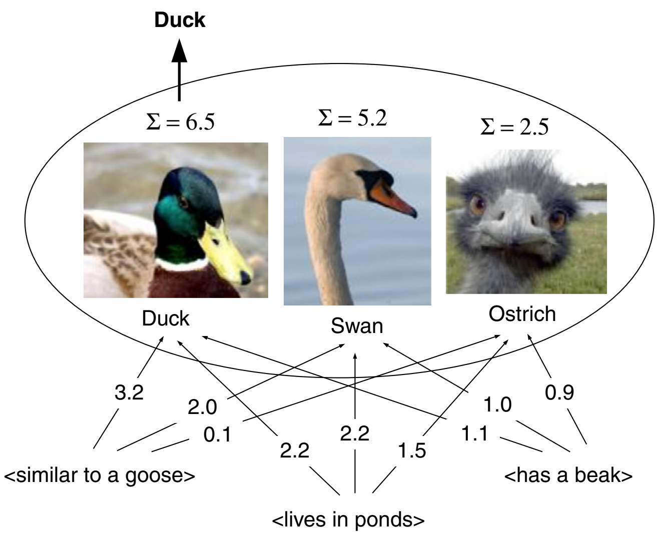

Fig. 1. Additive integration of semantic relevance (hypothetical example). A directed arrow between an attribute and a concept indicates that the attribute applies to the concept. Numerical values on the arrows denote semantic relevances. The concept with the highest summed relevance (Duck) is the one that will be retrieved.

However, the central issue of how differing semantic features are integrated in order to retrieve the corresponding concept is still awaiting specific investigations. Up to now, the proposed models of semantic relevance have been based on the standard, but empirically unverified, assumption of additivity of relevance of features (Sartori & Lombardi, 2004; Sartori et al., 2005a; Mechelli et al., 2006). In this paper, we present and compare various semantic integration models of how multiple semantic features influence concept name retrieval in a naming-to-description task. The main concerns are the integration processes assumed by the models and the resulting differences in their predictions. Each model is described and implemented in a stochastic framework and similarities and differences between the models are noted in three experiments involving three different groups of participants: normal adults, semantically impaired (DAT) patients, and healthy elderly controls.

The remainder of the paper is organized as follows. First, we review the formal framework used to define the notion of semantic relevance. Next, we describe three integration models of how semantic features are combined in terms of relevance during name retrieval. The three models are based on three different integration rules (additive, multiplicative, and winner-take-all rule) each of which subsumes a different psychological principle in the name retrieval process. We then compare the models with previous well-known models of category learning and cue integration and emphasize the commonalities and differences with respect to our new contribution. Next, the predictions of the three relevance models are derived and tested on data collected on three different groups of participants: young people, healthy elderly and semantically impaired Alzheimer patients. The predictions made by the models are used to provide a measure of identifiability from which the models can be discriminated from one another. Finally we investigate, in a Monte Carlo simulation study, the implications of the three integration models for degraded semantic knowledge. We close by considering further possible extensions of our approach.

A formal framework for semantic relevance

Many of the most influential theories of conceptual representation have been based on semantic features as their representational currency (e.g., Smith & Medin, 1981). Feature-based models have also been applied to the analysis of categorization (e.g., Estes, 1993; Medin & Schaffer, 1978), memory (e.g., Murdock, 1993), and similarity (Tversky, 1977). A main assumption made by the feature-based models is that concepts can be meaningfully reduced to sets of features. Moreover, these features are usually treated as independent in the sense that they make separable contributions to the model's output.

Semantic relevance models are feature-based models in which concepts are defined by vectors of weights codifying the intensities of features or properties used in describing a semantic domain. In a relevance context we usually refer to concept names as basic-level category names (Rosch, 1978). It is worthwhile representing a concept domain as an I (concepts) $\times$ J (features) dominance matrix $\mathbf{X} = [x_{ij}]$, where $x_{ij} \in \mathbb{R}_+$ denotes the degree of dominance (*intensity of association*) of Feature j for

Table 1
Hypothetical (7×8) dominance matrix $\mathbf{X}$

Concepts	**X** Features							
	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
c_1	1	0	0	2	0	0	3	0
c_2	3	0	0	0	0	0	0	0
c_3	1	0	0	0	2	0	0	0
c_4	0	0	2	0	0	0	0	0
c_5	0	1	0	0	0	3	0	2
c_6	0	0	0	0	0	0	1	2
c_7	0	0	0	0	0	1	0	3
I_j	3	1	1	1	1	2	2	3
$\mathbf{g}$	1.222	2.807	2.807	2.807	2.807	1.807	1.807	1.222

Rows of $\mathbf{X}$ correspond to 7 different concepts, whereas columns of $\mathbf{X}$ represent 8 different features. Entry x_{ij} of $\mathbf{X}$ denotes the frequency of occurrence of Feature j in Concept i over all subjects' descriptions; for example, $x_{43} = 2$ is the frequency of feature f_3 in concept c_4. The row denoted by I_j contains the number of concepts for which feature f_j applies at least once. So, for example, $I_6 = 2$ as the sixth column of $\mathbf{X}$ contains two non-zero frequencies. The row denoted by $\mathbf{g}$ contains the elements g_j $(j = 1, \ldots, 8)$ of the global component parameter (see Eq. (2) in the text).

Concept i. If $x_{ij} = 0$, we say that concept c_i and feature f_j are unrelated in the concept domain.

Several weighting schemes may be derived in modeling relevance (see, e.g., Mechelli et al., 2006; Sartori & Lombardi, 2004; Sartori et al., 2005a). In this paper, we refer to a simple weighting scheme called FF × ICF (*Feature Frequency × Inverse Concept Frequency*), adapted from Salton's well-known TF × IDF (*Term Frequency × Inverse Document Frequency*) information retrieval measure (Salton, 1989). The hypothetical dominance matrix (Table 1) and its derived relevance matrix (Table 2) serve as a guiding example for FI × ICI computations. Under the FI × ICI assumption the dominance matrix $\mathbf{X}$ is computed by setting entry x_{ij} as equal to the number of co-occurrences of Feature j in Concept i over all subjects' descriptions collected in a feature-to-listing task[3] (Sartori & Lombardi, 2004). Dominance can be understood as a measure of the local strength of a feature in describing a concept; for example, the higher the association x_{ij} of Feature j for Concept i, the more it is dominant for that concept (Ashcraft, 1978). A relevance process acts by transforming $\mathbf{X}$ into an $I \times J$ relevance model matrix $\mathbf{K} = [k_{ij}]$ which is derived by the following simple equation

$$k_{ij} = g_j x_{ij}, \quad i = 1, \ldots, I; \; j = 1, \ldots, J \tag{1}$$

where g_j represents a global feature weight. In particular g_j can be computed as

$$g_j = \log_2\left(\frac{I}{I_j}\right), \quad j = 1, \ldots, J \tag{2}$$

where I_j (with, $0 < I_j \leqslant I$) denotes the number of concepts in which Feature j loads a positive dominance (see Table 1, row denoted by I_j).

In substantive terms the global weigh may be read as follows: the more concepts a given feature is connected with the less distinctive is that feature (low g_j). The intuition was that a feature that occurs in many concepts is not a good discriminator, and should be given less weight than one which occurs in few concepts, and the weight g_j was an heuristic implementation of this intuition. Therefore, a feature that represents a good cue for a concept will have both high dominance and high distinctiveness. In contrast, if $k_{ij} = 0$, we say that Feature j is *irrelevant* for Concept i. Notice that, like distinctiveness, semantic relevance depends on the total number (I) of concepts in the representational model, which ideally should correspond to the size of the mental lexicon. This may be a critical point as we ignore on what set of concepts the actual computations of our mind are based. However, Sartori et al. (2005a) showed the robustness of relevance estimations. By means of a Monte Carlo simulation study these authors found that

[3] Feature-to-listing task is an empirical paradigm widely adopted in the featural based approach to construct and validate empirically derived conceptual representations. Several versions of the feature-to-listing task have been proposed in the literature (see, for example, Cree & McRae, 2003; Cree, McNorgan, & McRae, 2006; Randall et al., 2004; Rosch, 1975; Sartori & Lombardi, 2004; Storms, De Boeck, & Ruts, 2001; Tversky, 1977; Vigliocco, Vinson, Lewis, & Garrett, 2004), but, despite minor differences, they all require participants to list the features they believe to belong to a concept or stimulus. With regard to the semantic features collected in a relevance framework, they can be of any type and contain heterogeneous information about concepts, such as perceptual information (DOG: ⟨has four legs⟩), functional information (DOG: ⟨is used for hunting⟩), associative information (DOG: ⟨likes to chase cats⟩) and encyclopaedic information (DOG: ⟨may be one of many breeds⟩). This reflects the commonly believed fact that semantic memory includes heterogeneous knowledge about concepts.

Table 2
Relevance model matrix **K** derived from the dominance matrix **X** in Table 1

Concepts	K Features							
	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
c_1	1.2	0	0	5.6	0	0	5.4	0
c_2	3.7	0	0	0	0	0	0	0
c_3	1.2	0	0	0	5.6	0	0	0
c_4	0	0	5.6	0	0	0	0	0
c_5	0	2.8	0	0	0	5.4	0	2.4
c_6	0	0	0	0	0	0	1.8	2.4
c_7	0	0	0	0	0	1.8	0	3.7

Entry k_{ij} of **K** contains the relevances of a feature f_j for concept c_i. For example, $k_{43} = x_{43} \times g_3 = 2 \times \log_2(7/1) = 5.6$ is the relevance of feature f_3 for concept c_4 (see Eq. (1) in the text).

even if, in theory, the number of concepts in the representational model influences the absolute values of relevance, in practice the relevance ranking of concepts computed with varying levels of I remains substantially unchanged. Further formal and substantive interrelationships among dominance, distinctiveness and semantic relevance are discussed in Sartori et al. (2005a).

Closely related attribute dimensions

However, semantic features may differ on dimensions other than distinctiveness, dominance, and semantic relevance, and those dimensions can also influence the impact of a feature on different semantic tasks. In order to clarify our notion of semantic relevance, we distinguish it from other closely related dimensions such as salience, conceptual validity, and diagnosticity.

Salience refers in general to the intensity of a feature, the extend to which it presents a high amplitude signal in relation to background noise (Tversky, 1977). Semantic relevance is clearly not identical to salience because some features can be highly salient but have low relevance values. For example, ⟨has big teeth⟩ which is a highly salient feature for Lion has little relevance value for the same concept as many wild animals share this visual property (low distinctiveness).

Conceptual validity (or category validity) refers to the percentage of concepts displaying a given feature. Within a featural approach it can be represented as the conditional probability of a feature given a concept which has an obvious estimate in the normalized dominance: $P(f_j|c_i) \approx \frac{x_{ij}}{N}$, where N denotes a normalization factor which depends on the type of feature-listing task adopted.

Diagnosticity is usually referred to the informational value of a feature for one concept relative to a family of concepts. For example, if one's task is to retrieve a concept name, say Dog, the feature ⟨it barks⟩ may prove highly diagnostic as it excels at distinguishing Dog from other concepts of animals. Diagnosticity can be measured by means of *cue validity* (Rosch, 1978). Estimation of cue validity can be characterized as a process of inferring a Bayesian posterior probability for Concept i given that it has Feature j. Bayes' formula can be used to express this in terms of the prior probability of retrieving Concept i and the conceptual validity of Feature j for Concept i (see Appendix A.2 for further details). Unlike cue-validity, semantic relevance is a heuristic norm which is not based on probability estimates. Nonetheless, semantic relevance and diagnosticity share a common predictive goal as they are both concept-relative measures which integrate the informational value of a feature for one concept relative to the remaining concepts in the semantic domain.

Modeling a naming-to-description task

A widely adopted paradigm within the neuropsychology community is the so called naming-to-description task. It consists of presenting participants with a sentence describing a target concept, and including a set of semantic features. For example, the features ⟨has an handle-bar⟩, ⟨has two wheels⟩ and ⟨has two pedals⟩ is presented orally to the participant who is required to retrieve the name Bicycle (see Table 3 for other examples of concept descriptions). The paradigm can be understood as a particular kind of production task in which the retrieved concept name belongs to a possibly very large set of candidate names. According to our formal framework, this empirical task appears to involve at least three different sequential processes (Massaro, 1987; Selfridge, 1959): evaluation, semantic integration, and decision (which finally leads to name retrieval). More formally, when the concept description is presented to a participant, the overall relevance for this description results from the integration of the individual relevance values associated with each of the semantic features in

Table 3
Representative semantic features of high and medium relevance values for concepts GULL and PIPE-ORGAN

GULL	PIPE-ORGAN
Features with high relevance	Features with high relevance
⟨found near the sea⟩	⟨has pipes⟩
⟨has webbed feet⟩	⟨similar to a piano⟩
⟨has wings⟩:	⟨can be found in the cathedral⟩
Features with medium relevance	Features with medium relevance
⟨has white feathers⟩	⟨can be filled with wind⟩
⟨hunts for fish⟩	⟨has pedals⟩
⟨flies⟩	⟨used in ceremonies⟩

the description. The concept with the highest integrated relevance is the one that will be retrieved. The evaluation stage is defined as the analysis of each semantic feature in the description by the processing system. In the semantic relevance approach, it can be thought of as the retrieval, for each feature in the description, of its corresponding relevance value. In the evaluation stage, we assume that the resulting relevance value for a given feature in the description is independent of the remaining semantic features in the same description. This independence assumption rules out the possibility that, for example, in an evaluation stage the relevance values of attribute ⟨has a tail⟩ is affected by the relevance values of attribute ⟨can bark⟩. However, we do not assume that it also rules out a possible interaction of the semantic information in the final response that is due to the nature of the integration process.

Semantic integration is defined as some combination of the relevance representations made available by the evaluation process. We assume that for each concept-name c_i in the lexicon the semantic integration stage provides a single integrated relevance value K_i as a deterministic function

$$K_i = \phi(\mathbf{k}_i)$$

of the different relevance values $\mathbf{k}_i = (k_{ij_1}, k_{ij_2}, \ldots, k_{ij_q})$ provided by the evaluation stage and associated with the q different features in the concept description. Moreover, we also assume that the integrated relevance values have *no-memory* of how they were obtained. In other words, if two different sets of semantic features lead to identical outcomes of relevance integration, then the decision concerning which concept-name to retrieve would be the same in both cases. The latter means that the decision stage involved in name retrieval does not have access to the initial relevance values given by the evaluation stage and operates exclusively on the final integrated relevance value.

Finally, the retrieval accuracy for the target concept c_i given a concept description is modelled by a logistic function as follows:

$$\pi_i = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 K_i)]}. \quad (3)$$

In Eq. (3), $\beta_1 > 0$ is a steepness parameter (e.g., as $\beta_1 \to +\infty$, the logistic function reduces to a step function). Whereas $\beta_0 \leqslant 0$ is the intercept parameter (lower the value of β_0 the more the function shifts to the right). The model in Eq. (3) represents a monotonic transformation of the integrated relevance value K_i. Some examples of logistic functions compatible with a naming accuracy scenario are shown in Fig. 2. Standard relevance models (e.g., Sartori & Lombardi, 2004) are based on the assumption of *additivity* of relevance of semantic features (see Fig. 1). However, it is still to be established empirically whether participants really integrate information about features in an additive fashion. In this paper we propose three different elementary rules that could be used to integrate semantic relevance. We will describe the models that serve as representatives of the additive, multiplicative, and winner-take-all rules. The models will be contrasted against each other and compared with other existent models in the categorization or learning literature.

The additive relevance model

According to the *additive rule*, the retrieval accuracy for the target concept c_i given the set of features A_i is given by

$$\pi_i^1 = \frac{1}{1 + \exp[-(\beta_0 + \beta_1(\sum_{j \in J_i} k_{ij}))]}, \quad (4)$$

where J_i denotes the set of feature indices associated to A_i. The integrated relevance value $K_i^1 = \phi^1(\mathbf{k}_i)$ can be understood as the additive combination of its relevance values. We suppose that each concept c_i would be compared to the set of features A_i and that the matching probability would simply add to the target concept in the amount of their weighted dominance values. In psychological substantive terms the additive model implies that in the integration process the semantic features are considered independent in the sense that they make additive contributions to the probability of retrieving the target concept. In a more technical way we mean that the semantic effects of two features do not interact in the integration process, that is to say they are stochastically independent (Ashby & Townsend, 1986). Another issue that can be closely associated with semantic independence is that of semantic separability. For example, according to Garner and Morton (1969), we may consider two features as semantically separable if they act separately in the semantic system and thus can go independently of each other in the integration process. The assumption of independence is critical in both the design of behavioral experiments and the development of models and theories of category learning, concept

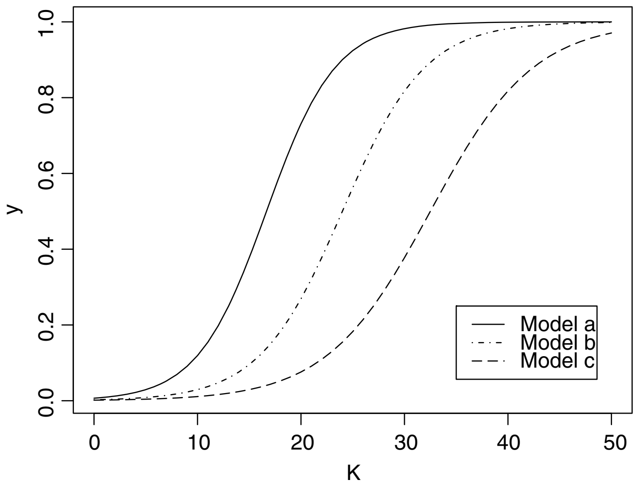

Fig. 2. Examples of logistic models of naming accuracy: Model a [$\beta_0 = -5,\ \beta_1 = 0.30$], Model b [$\beta_0 = -6,\ \beta_1 = 0.25$], Model c [$\beta_0 = -6.5,\ \beta_1 = 0.20$]. K denotes a generic integrated semantic relevance.

similarity and generalization. In general we refer to representational models that lack cross-featural interactions as being semantically independent.

The multiplicative relevance model

The assumption of featural independence may lack psychological validity. In principle, the additive relevance model can suffer from inherent fitting disadvantage because of its additive nature. For example, it can lack any capacity to allow a final integrated relevance value that increase (resp. decrease) in a not linear way. Moreover, the additive rule lacks any sensitivity for capturing eventual interactions among the informative value of semantic features. According to the *multiplicative rule*, the probability model is given by

$$\pi_i^2 = \frac{1}{1 + \exp[-(\beta_0 + \beta_1(\prod_{j \in J_i} k_{ij}))]}. \tag{5}$$

In the multiplicative relevance model, the integrated relevance value $K_i^2 = \phi^2(\mathbf{k}_i)$ is defined as the multiplicative combination of the relevance values in the concept description. This model mimics a semantic integration process that exaggerates the informativeness of high relevant descriptions relative to less informative descriptions and eventually increases the estimated percentage of correct retrievals. A further potential virtue of the multiplicative rule is that it can be sensitive to correlational structure of the featural dimensions (see Medin, 1983). However, it is sensitive to the presence of irrelevant features as well. As an example, let us consider two features f_1 and f_2 with corresponding relevance values $k_{i1} = 0.05$ and $k_{i2} = 7.5$ (for a concept c_i), respectively. The results of the integration rules are: $7.55 = 0.05 + 7.5$ for additive and $0.375 = 0.05 \times 7.5$ for multiplicative. The multiplicative rule might underperform whenever a concept description contains a majority of irrelevant features together with a small set of very relevant features (like, for example, ⟨has to humps⟩ for CAMEL). In this latter case we speculate that individuals may still be able to correctly retrieve the target name independently of the presence of irrelevant features.

The winner-take-all relevance model

According to the winner-take-all (WTA) relevance model, a concept retrieval can be based on only one cue, whatever the total number of cues presented in the concept description. WTA relevance model searches for the best cue in the order of its semantic relevance for the concept. Formally it can be defined as follows:

$$\pi_i^3 = \frac{1}{1 + \exp[-(\beta_0 + \beta_1(\max\{k_{ij} : k_{ij} \in \mathbf{k}_i\}))]}. \tag{6}$$

This model allows the retrieval decision maker to follow a simple exhaustive search of the best relevant cue: the feature with the maximum relevance for the concept is the selected one. Such one-cue selection does not need to combine different features, and so no common currency between cues need to be determined. The rationale for basing a concept retrieval on only a single piece of semantic information rather than on a combination of several cues is that combining the information from different cues requires converting them into a common final currency, an integration that can be, depending on the nature of the semantic task and the state of the cognitive system, expensive in terms of cognitive capacity. In contrast, this heuristic employs a minimum knowledge and computation to make choices in a semantic retrieval environment. However, such simplicity need not necessarily lead to a disadvantage in retrieval accuracy, as a simple max-type rule can help the retrieval model to be more robust than those that are based on multiple-cues integration. In general, the WTA relevance model may be considered more robust as it can better ignore the noise inherent in many semantic cues by looking instead to the most relevant cue. Thus, simply using the best cue can automatically yield robustness (Gigerenzer, Hoffrage, & Kleinblting, 1991). Furthermore, best relevant cues are likely to remain important even when the semantic representation changes to some degree, for instance due to semantic perturbation or semantic deficits. In contrast, low relevant cues can be more affected by random perturbation.

Relation to other models

The relevance models in Eqs. (4)–(6) are closely related to other integration models such as, for example, the rational Bayesian model of categorization (e.g., Anderson, 1991), the linear integration model (Anderson, 1981), the additive-prototype model (Smith & Minda, 1998), the multiplicative-prototype model (Estes, 1986; Nosofsky, 1987), the generalized context model (Nosofsky, 1984, 1986), and the probabilistic mental models (Gigerenzer et al., 1991). Here we will mainly focus on the general aspects of the comparison.

Bayesian models. The additive relevance model in Eq. (4) shares many commonalities with the rational Bayesian model of categorization. One obtains a rational Bayesian model (Anderson, 1991) if the notions of dominance and semantic relevance are replaced with those of conceptual validity and cue validity, respectively. The Bayesian model allows for a normative optimal solution of the retrieval problem. Like the additive relevance model also the Bayesian model assumes independence between features. More precisely the critical assumption is that the probability of observing a conjunctive combination of features given a target concept is independent of the probabilities of each single feature in the concept description. In general, however the specific representations of the two models are substantially different. In particular, given a concept-feature pair (c,f), the Bayesian model focuses on the derivation of the prior probability $P(c)$ and the conceptual validity $P(f|c)$. The prior probability $P(c)$ reflects the *a priori* probability of retrieving a concept name before the concept description is presented to the participant.[4] Unlike the Bayesian model, the additive relevance model does not use any concept parameter. It seems plausible that this difference depends upon the nature of the empirical task modelled. The Bayesian model of categorization has been developed for modeling data collected from researches studying the acquisition of artificial categories in the laboratory (e.g., Anderson, 1991; Friedman, Massaro, Kitzis, & Cohen, 1995) in which stimuli are characterized by few selected features. From its part semantic relevance models are feature-based models of non-artificial concept descriptions which are represented by high-dimensional co-occurrence data matrices. These co-occurrence data matrices immediately reveal the problem of *data sparseness*, also known as *zero-frequency problem* (Witten & Bell, 1991). A typical concept-feature matrix derived from a feature-to-listing task may only have a small fraction of non-zero entries (typically well below 1%), which reflects the fact that only very few of the features in the feature set are actually used in any single concept. Therefore, most of the counts in the matrix will thus typically be zero or at least significantly corrupted by sampling noise. If normalized frequencies (probability estimates) are used in predicting naming accuracy, a large number of co-occurrences is observed which are judged to be impossible based on the data set. This means that the Bayesian model may be sensitive to the zero-frequency problem. In contrast, the relevance models are based on a information retrieval measure ($FF \times ICF$) which, in turn, has been proven to efficiently overcome the sparseness problem (Salton, 1991; Salton & Buckley, 1988).

Linear integration models. According to Anderson (1981) a semantic integration can be computed by the addition of the quantities representing the evaluation of each cue of information. The retrieval is assumed to be linear; that is the integrated value can be mapped linearly into a rating scale. Like the additive relevance model and unlike the Bayesian model, the linear integration model is non-optimal (in a normative sense). The retrieval given two features supporting the same concept is a sort of average between the retrievals given to the separate features presented in isolation. In contrast, optimal integration (i.e., Bayes's theorem) dictates that the retrieval given two independent features be more extreme than either of the retrievals given the separate features supporting the same concept. However, the linear integration model and the additive relevance model differ with respect to the statistical tool adopted. Anderson (1981) proposed a comprehensive framework for the analysis of integration based on the analysis of variance (ANOVA) and interval-response scales. Unlike the additive relevance model, the linear integration model does not provide any explicit procedure, such as distinctiveness, to weight informative cues. From its part the additive relevance model is based on a generalized linear model which is able to work with both non-normal data (e.g., accuracies, reaction times) and non-linear functions mapping integrated relevance into responses

[4] In our opinion there is no straightforward way to estimate such prior probability from data collected using a feature-to-listing task.

(see Appendix A.1 for further details). Finally, like the Bayesian model, also the linear integration model has been applied mainly to experimental data collected in the laboratory.

Prototype and exemplar models. Prototype models are parametric models assuming that people abstract and store the prototype of categories. The prototype is defined as the most typical, or representative, category member (e.g., Rosch, 1977) and is usually computed as the centroid for all points in the multidimensional space that are associated with the category. From their part, relevance models are based on a semantic representation in which basic categories are described by sets of relevant features. According to the featural representation proposed in this paper, the relevance profiles of the I distinct concepts do not correspond to the centroids of the associated I basic categories. Therefore, the relevance profile of a concept cannot be interpreted as a sort of prototype pattern. The difference between relevance models and exemplar models is even more extreme as exemplar models assume that humans represent categories by storing every exemplar (together with its category label) in memory. Moreover, like the Bayesian models, also the prototype and exemplar models have been prevalently applied in researches studying the acquisition of simplified artificial categories (but see Smits, Storms, Rosseel, & De Boeck (2002) and Storms et al. (2001) for an application of the generalized context model to the domain of natural concepts). Nonetheless, we can still observe some commonalities between these families of models. Like the multiplicative relevance model, also most currently popular categorization models such as, for example, the multiplicative-prototype model (Estes, 1986; Nosofsky, 1987) and the generalized context model (Nosofsky, 1984, 1986) satisfy the product rule. In particular, the global match between the probe stimulus and the exemplars (or prototypes) is usually defined as an exponentially decreasing similarity function satisfying the product rule property. However, unlike these models, the multiplicative relevance model does not use any kind of similarity matching function in the modeling of the retrieval process.

Probabilistic mental models (PMM). Gigerenzer et al. (1991) proposed a formal framework for modeling inferences from memory. The main assumption of this approach is that cognitive processes are based on simple and plausible mechanisms of inference. Those mechanisms are so simple that a cognitive system can carry them out under limited time and knowledge. In particular, in PMM the search for information is reduced to a minimum, and in general there is no integration of elements of information. The *take-the-best* algorithm is probably the most known implementation of this family of models. Like, the WTA relevance rule, the take-the-best algorithm assumes a subjective rank order of the cues according to their informativeness. Hence, the best cue is selected and the remaining cues are ignored. However, unlike the WTA relevance rule, the take-the-best algorithm usually works for binary cues and is a simple heuristic that cannot be considered a standard statistical tool for inductive inference. Nonetheless, like the WTA relevance rule, it is a noncompensatory algorithm as only the best discriminant cue determines the final inference and no combination of other cues can contribute to this inference.

Rationale of the empirical studies

In order to evaluate which of the previously proposed relevance integration rules best predicts naming accuracy, we report three experiments on three different groups of participants: (a) normal adults (b) semantically impaired (DAT) patients, and (c) elderly healthy controls. All the experiments in this paper employed a particular paradigm, the so called naming-to-description task, which permits full control over the presented semantic features and is an empirical paradigm usually adopted to investigate semantic memory disorders (e.g., Lambon Ralph, Graham, Ellis, & Hodges, 1998; Silveri & Gainotti, 1988). Given the set of plausible integration models described above, the real basis of comparison is the predictive power of the models. Reliable assessments will be possible only after the models have been contrasted in a broad range of experimental tasks. Here we begin this project by providing a simple experimental setting in which the performance of the models is evaluated in different groups of individuals.

Study 1

In this first study we evaluate the performance of the three relevance integration models in predicting naming accuracy in a naming-to-description task administered to a group of young adults.

Method

Participants. Twenty-six undergraduate and postgraduate students (22 women, 4 men; mean age = 23.4) from the Faculty of Psychology of the University of Padua (Italy) participated voluntarily in the first experiment.

Stimuli and procedure. Fifty concepts were used, randomly selected from a larger pool of 254 concepts belonging to the dominance database of Sartori and Lombardi (2004), which furnished the starting dominance matrix. The dominance database contained concepts included in 13 different categories (i.e., birds, buildings, clothes, flowers, furniture, fruits, houses, wares, mammals, musical instruments, vegetables, vehicles and weapons; Dell'Acqua, Lotto, & Job, 2000) and 2619 distinct features. The 50 concepts selected for use in this investigation guaranteed: (i) a sufficient number of stimuli in order to run the accuracy models, (ii) the various semantic relevancies spanned within range $[1.39, 33.22]$ with mean $= 9.39$ and $SD = 6.56$. Each concept c_i $(i = 1, \ldots, 50)$ was described by a sentence containing a set A_i of n_i semantic features (n_i ranging from 1 to 6) randomly selected from the set of all features that applied to it. For each concept c_i, the n_i semantic features were presented orally and in random order to the participants, who were required to retrieve the corresponding concept with an emphasis on accuracy rather than speed. The required responses were oral. The dependent variable of this study was naming accuracy, y_i (with mean $= 0.471$ and $SD = 0.383$), computed as the proportion of subjects who correctly retrieved the name of the target concept given the feature set A_i. The independent variables were the three integrated semantic relevance values K_i^h $(h = 1,2,3)$ computed according to Eqs. (4)–(6).

Model fitting. The data collected from the experiment were fit using the three logistic models described by Eq. (10) (see Appendix) on the basis of maximum-likelihood estimation (MLE). MLE parameters values and their standard errors for the three models are presented in Table 4. The standard errors were also estimated by taking 1000 nonparametric bootstrap samples from the data. More precisely, for each observed accuracy y_i, the resampling was done by drawing an integer number between 0 and 26 from a binomial distribution, $n_i^* \sim \text{Bin}(y_i, 26)$, therefore the observed proportion y_i was replaced with $y_i^* = n_i^*/26$. The fit of the model based on the additive integration rule ($\text{AIC}_1 = 395.15$, $R_1^2 = 0.705$, $r_1 = 0.879$)[5] was better than the fit yielded by the other two competing models (multiplicative rule: $\text{AIC}_2 = 779.81$, $R_2^2 = 0.297$, $r_2 = 0.573$; max rule: $\text{AIC}_3 = 674.75$, $R_3^2 = 0.408$, $r_3 = 0.678$). Fig. 3 (left panels) shows the data and the model predicted values. Fig. 3 (right panels) shows the residual scatterplots for the three logistic models.

Evaluating model mimicry. The problem of identifying the true integration relevance model cannot be based exclusively on a generic evaluation of the goodness-of-fit for the observed data, as the obvious mathematical dependency between the three integration rules yields strong inter-correlations between the empirical predictors (see Fig. 4). Therefore, it can be difficult to discriminate between the real adequacy of the three rules. In particular, one of the models may happen to show a general larger flexibility in fitting the data. This larger flexibility may be due to a sort of *chameleon* model, which is able to give a relatively good account for data it did not actually generate. Hence, it is important to evaluate to which extent a given integration rule is able to mimic the behavior of other rules. In this section we will asses the relative flexibility of the models using the *parametric bootstrap cross-fitting method* (PBCM) proposed by Wagenmakers, Ratcliff, Gomez, and Iverson (2004). This sampling procedure consists in generating distributions of differences in goodness-of-fit expected under each of the competing models. Therefore, the original observed differences in goodness-of-fit can be compared to the generated distributions of differences which allows for a quantitative evaluation of model adequacy. In particular, for each pair of different models (h, h'), the PBCM procedure yields two distributions of AIC differences, one derived under the assumption that model h is true, and one derived under the assumption that model h' is true. The observed difference computed on the original data set can then be assessed with reference to these two distributions. When the observed difference has higher probability under the distribution with model h-true than under the alternative distribution (model h'-true) this is an evidence supporting that model h is more adequate than model h' (see Appendix for a detailed description of the PBCM procedure implemented in this paper).

With respect to our data, the difference distributions are shown in Fig. 5. The inspection of Fig. 5 allows several observations. First, on the basis of the nominal criterion 0 of no difference in AIC value, all simulated data sets are correctly classified across all difference distributions. That is to say, there is no overlap between

Table 4
Parameter point estimates, nonparametric bootstrap mean parameter estimates and nonparametric standard errors for the three integration models (additive, multiplicative, and WTA) fitted to the observed naming accuracy y_i

	β_0	z	p	β_1	z	p
Additive: K_i^1						
Point	−3.137	−18.59	<.01	0.151	19.59	<.01
Mean	−3.147			0.151		
SE	0.147			0.007		
Multiplicative: K_i^2						
Point	−0.807	−11.07	<.01	0.002	11.78	<.01
Mean	−0.809			0.0018		
SE	0.065			0.0001		
WTA: K_i^3						
Point	−1.844	−15.39	<.01	0.165	15.18	<.01
Mean	−1.848			0.165		
SE	0.096			0.009		

[5] The Akaike criterion (AIC) is an index of fit that takes into account the parsimony of the model. Smaller AIC values are indicative of a better fit to the data. R^2 is a generalization of the residual sum of squares for linear models which is defined as $1 - (\log L_1 / \log L_0)$, where $\log L_1$ and $\log L_0$ are the deviance under the target model and the deviance under the null model (that is the model based only on the intercept parameter β_0), respectively. Finally, r denotes the Pearson-correlation between the predicted probability π and the observed accuracy y.

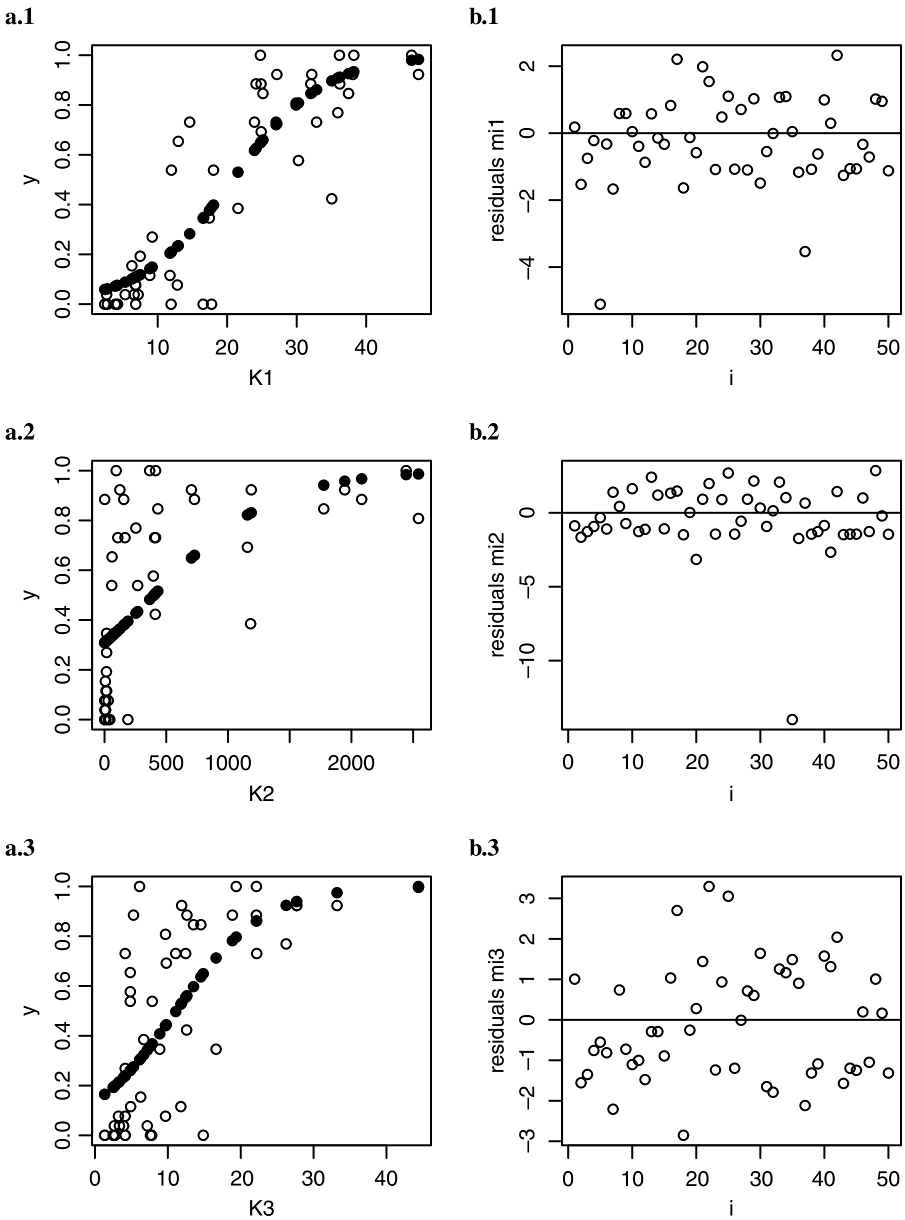

Fig. 3. Logit models for naming accuracy y_i. Panel (a.h): scatterplot of y_i on K^h where filled dots represent the model predictions π^h. Panel (b.h): residual plots of the linear predictor μ^h. Index value: $h = 1$ (additive), *2* (multiplicative), *3* (WTA).

the difference distributions across all comparisons. Second, the observed difference in AIC values is located near the middle of the additive rule model (top panel and bottom panel of Fig. 5). This result indicates that the data are much more likely under the additive integration rule than under either the multiplicative rule or WTA rule.

Assessing the relevance integration assumptions. Semantic relevance integration rules are based on the main assumption requiring the existence of a monotonic function between integrated relevance and naming accuracy. Results of our analysis have confirmed the plausibility of this hypothesis. However, it is not clear whether naming accuracy is really independent from the particular size and type of set of features used in the concept description, that is to say, if naming accuracy depends exclusively on the final value of the integrated semantic relevancies. In order to check whether the number n_i of semantic features used in the concept description might have affected the goodness-of-fit of the additive

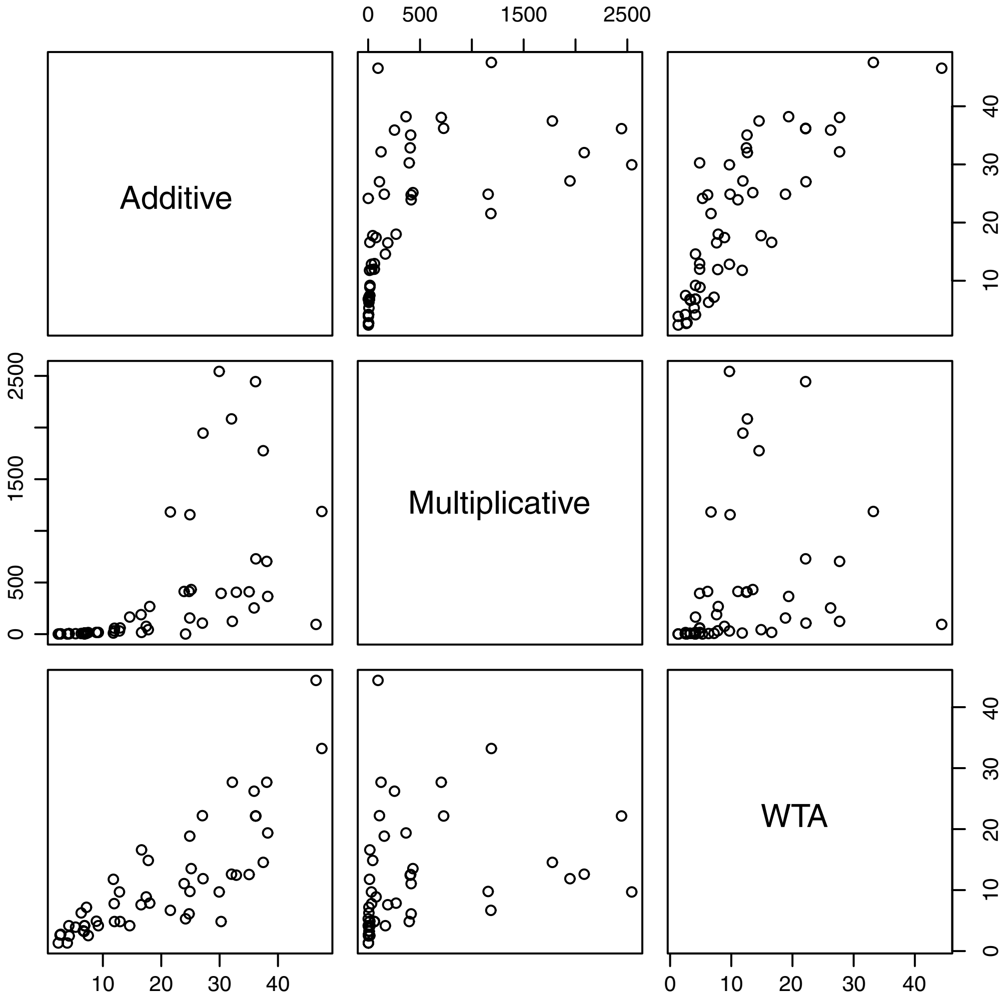

Fig. 4. Scatterplot matrix for the three empirical predictors of naming accuracy.

rule model, we ran a new analysis by introducing n_i as an additional factor in the logit model. The result of the likelihood-ratio test between the original logistic model and the new model showed a significant reduction of the deviance due to the n_i term ($\Delta d = 11.64$, with $p < 0.001$). However, the difference between the amount of deviance explained by the new model and that explained by the original model is substantially negligible, $R^2_{\text{new}} - R^2_{\text{old}} = 0.717 - 0.705 = 0.012$, as is the difference in the AIC values, $\text{AIC}_{\text{new}} - \text{AIC}_{\text{old}} = 385.52 - 395.15 = -9.63$. Therefore, adding the new term n_i does not seem to really challenge the pattern of results obtained by the original model, although it suggests that future investigations concerning the possible effect of description length on naming accuracy might clarify the real validity of the no-memory assumption.

Performance of the Bayesian model. Given that the Bayesian model shares many commonalities with the additive relevance model, one may wonder whether the two models differ with respect to data fitting. We can construct a Bayesian model (Anderson, 1991) by integrating into the Bayes' rule the notions of conceptual validity and cue validity (see Appendix A.2 for further details). In order to make the comparison valid the two models should have equivalent degrees of freedom. Therefore, we set to $1/(I = 254)$ the prior probability of each of the 50 target concepts in the experimental task. This latter constraint yielded a simplified version of the Bayesian model. To overcome possible mis-estimation of the Bayesian model, we fit the dependent variable by means of a simple linear model $y_i = \beta_0 + \beta_1 p(c_i|A_i)$, where $p(c_i|A_i)$ denotes the Bayesian conditional probability of retrieving c_i given the description A_i (see Eq. (14), Appendix). MLE parameters values and their standard errors for the Bayesian model were -0.021 (0.057) for β_0 ($p = 0.718$) and 0.664 (0.075) for β_1 ($p < 0.001$), respectively. Although the fit of the Bayesian model ($R^2 = 0.573$) was better than the fit

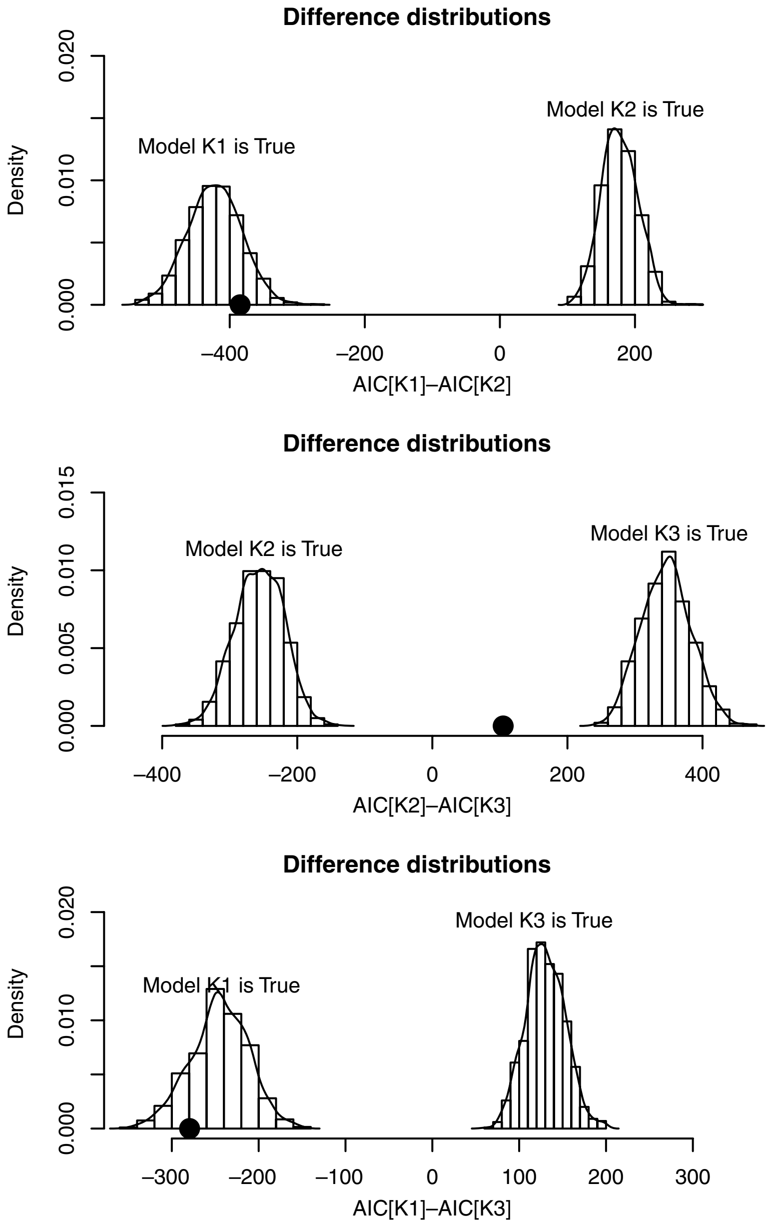

Fig. 5. Difference distributions obtained by the PBCM procedure applied to the data of the first study. The solid dot indicates the observed AIC's difference computed on the original data set.

yielded by the multiplicative relevance model and WTA relevance model, the additive relevance model outperformed the Bayesian one.

Discussion

The results are more consistent with the model based on the additive integration rule. Moreover, the evaluation of the model mimicry also confirms the hypothesis that adding distinct relevance information yields a better explanation of the empirical data in a naming-to-description task. The probability of retrieving the name of the target concept is higher when the integrated relevance is also high and, in general, the retrieval function is nearly linear over much of its range, say between about $K^1 = 10$ and $K^1 = 30$. Moreover, an almost perfect concept retrieval is observed when descriptions with integrated relevance values greater than $K^1 = 40$ are presented to participants. In conclusion, the prediction based on the additive rule is striking compared with the other two competing models and the overall result is in line with what was already observed in previous studies on the predictive power of semantic relevance (Mechelli et al., 2006; Sartori et al., 2005a, 2005b).

Study 2

Neuropsychological studies conducted on patients with specific knowledge impairments have been a useful source of data for addressing issues on the organization of conceptual-semantic knowledge in the human brain (Forde & Humphreys, 2000). Semantic relevance was originally proposed as a model for explaining category specificity in neuropsychological patients with semantic memory disorders (Sartori & Lombardi, 2004). Semantic memory patients show a degradation of concept representation which causes poor performance in semantic memory tasks. They are impaired in concept description, in word-picture matching, in picture naming, and also in naming-to-description (Hodges & Patterson, 1995). An interesting issue is whether the impairment exerts its effect through a modification of the original semantic relevance values or through a modification of the featural integration mechanism. In order to verify these contrasting hypotheses, in this second experiment a naming-to-description task was administered to a DAT group. The experimental design is identical to that of Study 1, with the following two exceptions: first, a new set of 80 concept descriptions was presented; second, given that patients with DAT might have severe memory problems, each of our concept descriptions was limited to only three features.

Method

Participants. Twenty-three patients with diagnoses of dementia of Alzheimer's type (DAT) (mean age = 76.8 years, $SD = 8.31$; mean education = 6.8 years, $SD = 5.05$) were selected for this second study. All patients were native Italian speakers. Some degree of semantic impairment is commonly seen in the early stages of dementia of Alzheimer's type (Chertkow & Bub, 1990; Hodges & Patterson, 1995), and this investigation was conducted on DAT patients with this characteristic. The background neuropsychological data collected on the DAT patients are reported in Table 5. The 23 DAT patients (15 women, 8 men) met the National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria for probable Alzheimer's disease (McKhann et al., 1984). All 23 patients had Hachinski scores (Hachinski et al., 1975) below 4 and an MMSE (Folstein, Folstein, & Mc Hugh, 1975) below 24/30. All patients with DAT were at least 2 *SD* below average scores of the normative sample on two anterograde and two semantic memory tests. All underwent CT or MRI scanning, together with a screening battery to exclude treatable causes of dementia. Patients with major depression, past history of known stroke or TIA, alcoholism, head injury or major medical illnesses were excluded. Patients were recruited in three hospitals and four nursing homes located in the Veneto district (North-East Italy).

Stimuli and procedure. Eighty concepts were used in this second study. The concepts were randomly selected from the same database used in Study 1. Each concept was described by a sentence consisting of only three semantic features randomly selected from the set of all the features that applied to the target concept (relevance range $[1.12, 51.03]$ with mean = 22.01 and $SD = 11.33$). The three semantic features were presented orally in random order to the participants, who were required to retrieve the corresponding concept. Each sentence was presented to 15 different patients.[6] As for Study 1, the responses were oral.

Model fitting. Data collected from the DAT group were fit using the three logistic models described by Eq. (10). The average accuracy of the DAT group on the 80 items was 0.284 ($SD = 0.272$). Table 6 reports the best MLE parameters and their standard errors. As for Study 1, we also computed bootstrap standard errors for the estimated parameters. The fit of the model based on the additive integration rule ($AIC_1 = 353.81$, $R_1^2 = 0.625$, $r_1 = 0.815$) was better than the fit yielded by the multiplicative rule ($AIC_2 = 522.75$, $R_2^2 = 0.289$, $r_2 = 0.542$) and the max rule ($AIC_3 = 464.24$, $R_3^2 = 0.405$, $r_3 = 0.651$). A large proportion of participants showed medium-low accuracy and only one subject had a perfect retrieval performance. In general, the retrieval function was approximately linear over all its empirical range (Fig. 6).

Evaluating model mimicry. With respect to DAT data, the difference distributions are shown in Fig. 7. In the DAT group the differences in the AIC values are qualitatively similar to those observed in Study 1. Like Study 1, the additive integration model is better able to account for the observed AIC difference computed on the original data set, as it is closer to the middle of its distribution as compared to the distribution of the alternative competing model.

[6] Data were collected in two different periods of time. In the first period, a test with 50 concept descriptions was administered to 15 DAT patients. Next, in a subsequent period, 30 supplementary concept descriptions were presented to a new group of 15 patients. This new group was composed of seven of the original 15 patients plus eight new DAT patients. The two DAT groups (first group: 15 patients, new group: 8 patients) did not differ with respect to the basic neuropsychological tests presented in Table 5.

Table 5
Background neuropsychological and semantic memory screening tests for group of DAT patients and normal controls

Description	DAT		Controls		Difference
	Mean	*SD*	Mean	*SD*	
Neuropsy. tests					
MMSE (correct max = 30)	19.06	2.51	25.81	1.45	$p < .05$
Prose memory test	1.98	0.63	9.01	2.88	$p < .05$
Incid. phon. memory (max = 20)	1.33	0.71	3.67	1.38	$p < .05$
Semantic memory tests					
Picture naming					
Non-living % ($N = 32$)	50.5	12.33	82.41	9.88	$p < .05$
Living % ($N = 32$)	48.6	18.12	84.55	11.07	$p < .05$
Naming to description					
Verbal description % ($N = 14$)	57.33	15.81	91.22	9.39	$p < .05$
Visual description % ($N = 11$)	31.56	16.31	81.02	15.08	$p < .05$

Table 6
DAT group

	β_0	z	p	β_1	z	p
Additive: K_i^1						
Point	−4.445	−16.98	<.01	0.060	14.96	<.01
Mean	−4.465			0.062		
SE	0.261			0.004		
Multiplicative: K_i^2						
Point	−1.717	−16.93	<.01	0.0002	10.86	<.01
Mean	−1.711			0.0001		
SE	0.1015			0.00001		
WTA: K_i^3						
Point	−2.670	−17.23	<.01	0.054	13.21	<.01
Mean	−2.632			0.058		
SE	0.155			0.004		

Parameter point estimates, nonparametric bootstrap mean parameter estimates and nonparametric standard errors for the three integration models (additive, multiplicative, and WTA) fitted to the observed naming accuracy y_i.

Discussion

Overall, the results are more consistent with the model based on the additive integration rule. This conclusion is supported by both the results of the model fitting and PBCM analysis. We believe that one important factor is, indeed, the different task adopted in Study 2. In contrast to the modality of feature presentation of Study 1, where an experimental sentence could contain a number of different features ranging from 1 to 6, the task administered to the DAT group involved only three different semantic features. In theory, this description length constraint might have an effect in diminishing the difference between the integration rules. In particular, the task in Study 1 would naturally yield larger semantic relevance variability, as compared to that of Study 2. In order to avoid this problem we augmented the number of concepts in the experimental task (from 50 to 80 concepts), as well as the range of the relevance values as compared to Study 1. This provided enough information to discriminate the predictive power of the three integration rules.

Study 3

In the third study, the performance of the DAT group from Study 2 on the naming-to-description task was compared to that of a group of older healthy adults, matched for age, education, and vocabulary.

Method

Participants. A group of 31 normal controls (mean age = 75.01 years, $SD = 7.78$; mean education = 5.41 years, $SD = 2.25$), matched for age and education to the DAT group. Controls performed better than patients with DAT on picture naming and naming-to-description tests, used here as semantic memory screening tests (Table 5).

Stimuli and procedure. A new set of 80 concepts were used in this second study (relevance range [1.17, 44.21] with mean = 21.07 and $SD = 10.44$). Selection procedure, modality of stimuli presentation and required response were the same as in Study 2. In particular, each sentence was presented to 25 different controls.[7]

[7] Like for Study 2, also in this third study data were collected in two different periods of time. In the first period, a test with 50 concept descriptions was administered to 25 controls. Next, in a subsequent period, 30 supplementary concept descriptions were presented to a new group of 25 controls. This new group was composed of 14 of the original 25 patients plus eleven new controls. Like for patients with DAT, they did not differ with respect to the basic neuropsychological tests presented in Table 5.

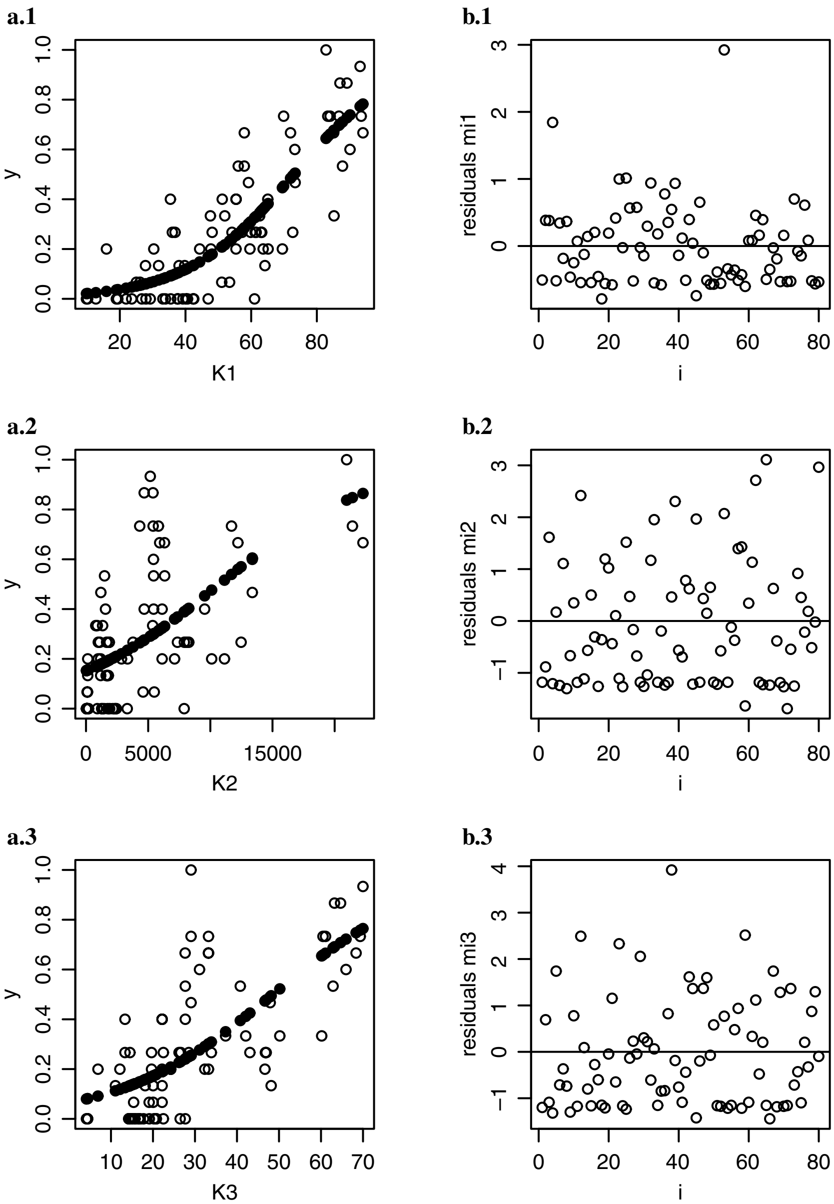

Fig. 6. Logit models for naming accuracy y_i [DAT group]. Panel (a.h): scatterplot of y_i on K^h where filled dots represent the model predictions π^h. Panel (b.h): residual plots of the linear predictor μ^h. Index value: $h = 1$ (additive), 2 (multiplicative), 3 (WTA).

Model fitting. The average accuracy of the controls group on the 80 items was 0.475 ($SD = 0.335$). Table 7 reports the best MLE parameters and their standard errors. The fit of the model based on the additive integration rule ($\text{AIC}_1 = 575.38$, $R_1^2 = 0.672$, $r_1 = 0.822$) was better than the fit yielded by the multiplicative rule ($\text{AIC}_2 = 771.66$, $R_2^2 = 0.495$, $r_2 = 0.715$) and the max rule ($\text{AIC}_3 = 842.5$, $R_3^2 = 0.431$, $r_3 = 0.671$). Fig. 8 (left panels) shows the data and the model predicted values. Fig. 8 (right panels) shows the residual scatterplots for the three logistic models. In contrast to the DAT group, the control group had a large proportion of participants who showed medium-high accuracy (Fig. 8, panel a.1).

Evaluating model mimicry. As for the DAT group, also in the control group, the additive integration model seems to better account for the observed AIC difference computed on the original data set, as it is closer to the middle of its distribution as compared to the distributions of the alternative competing models (Fig. 9).

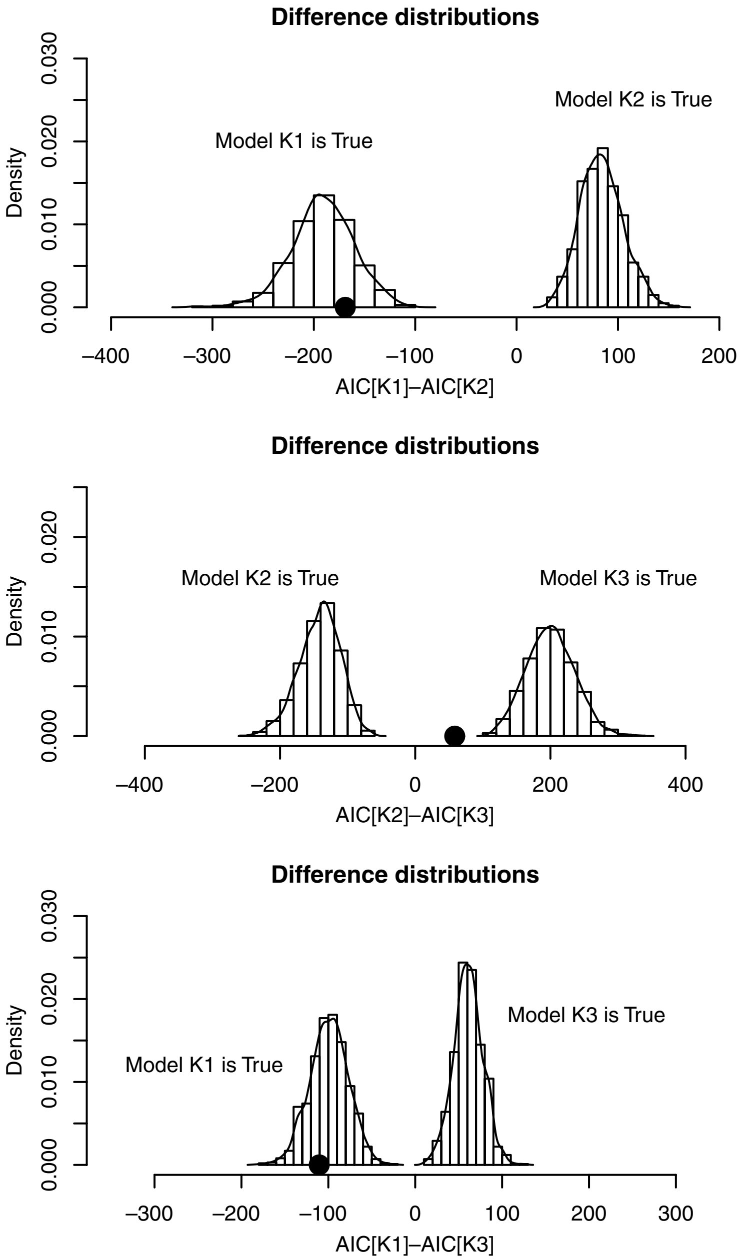

Fig. 7. Difference distributions obtained by the PBCM procedure applied to the data of the second study [DAT group]. The solid dot indicates the observed AIC's difference computed on the original data set.

Discussion

The results in the control group were consistent with what observed in the previous studies. In particular, the model based on the additive integration rule yields, overall, a better support for the observed data. This seems to be an almost robust result across all the three studies. In sum the results suggest that both the multiplicative and WTA models have something amiss in accounting for the empirical data.

Implications for degraded semantic knowledge

The sensitivity of the three integration rules to degraded semantic knowledge can differ in substantive

Table 7
Control group

	β_0	z	p	β_1	z	p
Additive: K_i^1						
Point	−3.127	−20.61	<.01	0.072	21.57	<.01
Mean	−3.189			0.069		
SE	0.151			0.003		
Multiplicative: K_i^2						
Point	−1.119	−15.93	<.01	0.0002	17.34	<.01
Mean	−1.121			0.0002		
SE	0.070			0.00001		
WTA: K_i^3						
Point	−2.627	−18.02	<.01	0.128	18.56	<.01
Mean	−2.591			0.131		
SE	0.145			0.006		

Parameter point estimates, nonparametric bootstrap mean parameter estimates and nonparametric standard errors for the three integration models (additive, multiplicative, and WTA) fitted to the observed naming accuracy y_i.

ways. Semantic memory deficits can be the result of several types of neuropathology. For example, the clinical course of Alzheimer's disease is generally characterized by progressive gradual memory deterioration. In particular, Alzheimer's disease is a pathology of widespread, patchy damage affecting many regions in the brain such as temporolimbic and frontoparietal regions. Neuroanatomical investigations in semantically impaired patients demonstrated a considerable decrease in the synapse to neuron ratio, due to synaptic deletion (e.g., Davies, Mann, Sumpter, & Yates, 1987). Semantic memory impairments can be simulated by damaging computational models of normal semantic system. For example, a common assumption in connectionist models of semantic memory deficits (e.g., Devlin et al., 1998; Farah & McClelland, 1991; Rogers et al., 2004) implies a diffuse, progressive deletion of the connections between units in the semantic system. Within our featural approach, a good analogy to the Alzheimer's pathology is provided by subjecting the featural representation to diffuse, progressive deletion of the connections between features and concepts in the semantic system. However, like other authors (e.g., Devlin et al., 1998) we do not assume that our featural representation maps directly onto a network of neurons and that connections between features and concepts correspond to synapses. Rather, we may assume that a single semantic feature is represented in the brain by a possibly large subnetwork of neurons with complex interactions. Consequently, under this assumption, a concept corresponds to a structure of these subnetworks.

The aim of this section is twofold. On one hand, we were interested in understanding how a diffuse, progressive deletion of relevance values in the model can affect the integrated relevance of a concept description. In particular, we evaluated the impact of deletion on the three integration rules proposed in this paper. On the other hand, we wanted to estimate the degree of deterioration of semantic knowledge representation in the DAT group as compared to the controls group. A simple statistical procedure is proposed to estimate the percentage of semantic degradation in the DAT group. In particular, we investigated how the degree of semantic degradation in the DAT group varied as a function of the relevance integration rules discussed in this paper.

Simulating degraded semantic knowledge

To simulate the progressive nature of the disease and evaluate its impact on integrated relevance values, the (254×2619) relevance model matrix $\mathbf{K}$ was lesioned with varying levels of severity. A lesion consisted in removing a percentage of the 14,170 positive relevance entries in $\mathbf{K}$. Three aspects were systematically varied in a complete factorial design: (a) the size q of the concept description, at four levels: 2, 3, 5, and 10 features; (b) the proportion e of deleted positive relevance entries in $\mathbf{K}$, at 20 levels: .05, .10, .15, ..., 1.00; (c) the type of integration rule adopted: additive, multiplicative, and WTA. For each combination of factors q and e, 80 random concept descriptions were constructed as follows: First, 80 different concepts were randomly selected from the whole set of 254 concepts. Subsequently, for each selected concept, a concept description was constructed by randomly selecting q semantic features from the set of all the features that apply to the concept. Next, $\mathbf{K}$ was perturbed according to severity damage e. Random perturbation was carried out stochastically by setting to 0 the fraction e of all positive entries in $\mathbf{K}$. This yielded the new perturbed relevance matrix $\mathbf{K}^*$. Therefore, for each concept description, the corresponding integrated relevance was recomputed according to $\mathbf{K}^*$. Dependent variable was the proportion of integrated relevance loss[8] as a functions of q, e, and type of integration rule. Fig. 10 presents the averaged results of 500 simulations. Overall our results suggested that all three integration rules were sensitive to perturbed semantic representations. However, a dominance relation can be read from Fig. 10 as follows:

$$\text{Multiplicative} \succ \text{Additive} \succ \text{WTA},$$

where $X \succ Y$ denotes that X is more sensitive to semantic perturbation than Y. In general, the effect

[8] This proportion was computed as the average of the values

$$1 - [K_i^*/K_i],$$

across the 80 randomly selected concepts, where K_i^* and K_i denote the integrated semantic relevance computed according to the perturbed relevance matrix $\mathbf{K}^*$ and the original relevance matrix $\mathbf{K}$, respectively.

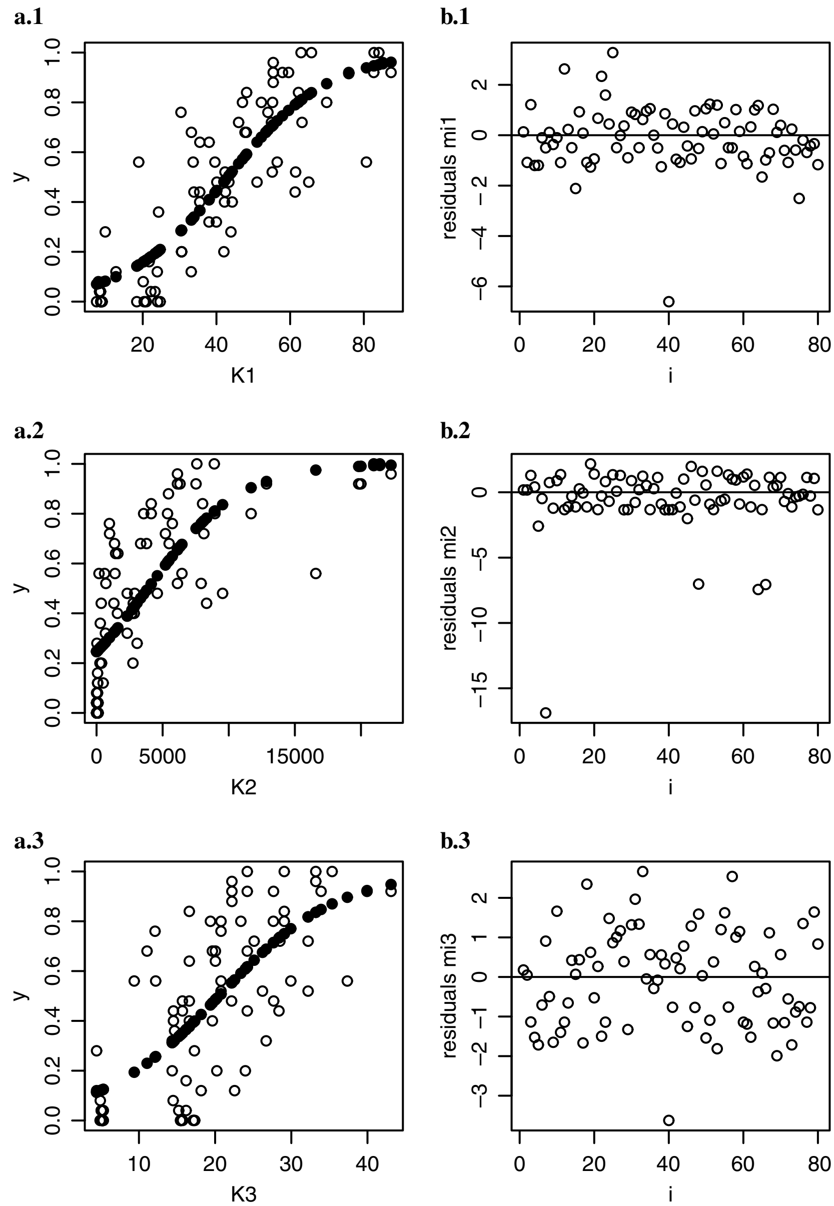

Fig. 8. Logit models for naming accuracy y_i. [Controls]. Panel (a.h): scatterplot of y_i on K^h where filled dots represent the model predictions π^h. Panel (b.h): residual plots of the linear predictor μ^h. Index value: $h = 1$ (additive), 2 (multiplicative), 3 (WTA).

of the perturbation was stronger in the multiplicative model. The effect increased with larger sizes q of concept description or more severe proportion e of random perturbation. Unlike, the multiplicative rule, the additive rule was linearly sensitive to random perturbation and its effect remained constant across q. As expected, the WTA rule was very robust to random perturbation and overall the effect decreased with larger values of q.

Estimating the amount of semantic memory damage

Another issue worth of investigation concerns the overall degree of deterioration of semantic knowledge in DAT patients. Here we will propose a statistical procedure to estimate the percentage of semantic degradation in the DAT group. In particular, we will investigate how semantic degradation varies as a function of the relevance integration rules proposed in this paper.

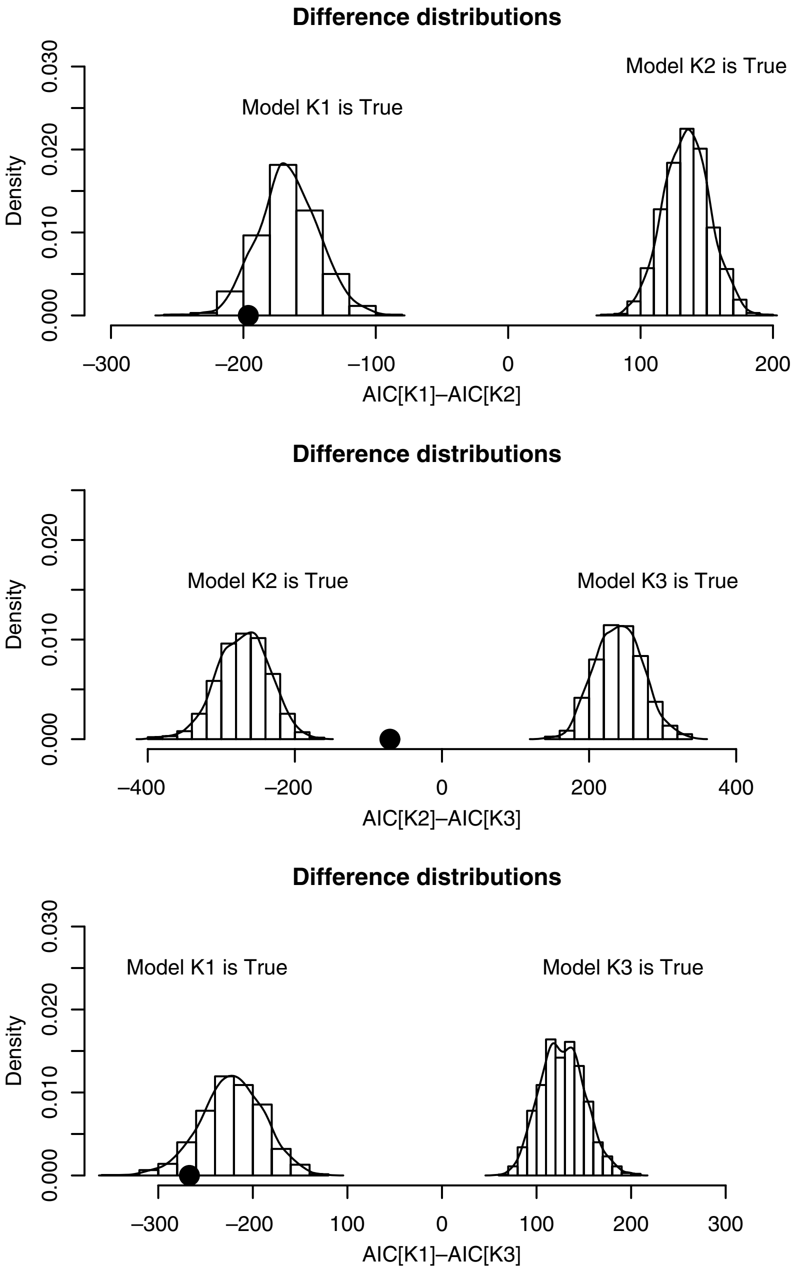

Fig. 9. Difference distributions obtained by the PBCM procedure applied to the data of the first study [Controls]. The solid dot indicates the observed AIC's difference computed on the original data set.

We implemented three general linear models corresponding to the three integration rules in which we considered also a group factor (DAT vs. controls) as an additional variable in the models. This allows a direct test for a possible group effect on naming accuracy. In the final logistic models the new linear predictor μ_i^h (for each integration rule $h = 1,2,3$) is defined as follows:

$$\mu_i^h = \beta_0^h + \gamma^h D_i^h + \beta_1^h K_i^h. \tag{7}$$

In Eq. (7), D_i^h is the dummy variable codifying the two groups of concepts, more precisely, D_i is coded 1 for DAT-concepts (the 80 concepts of Study 2) and 0 for control-concepts (the 80 concepts of Study 3). Therefore, for DAT the model becomes

$$\mu_i^h = (\beta_0^h + \gamma^h) + \beta_1^h K_i^h, \tag{8}$$

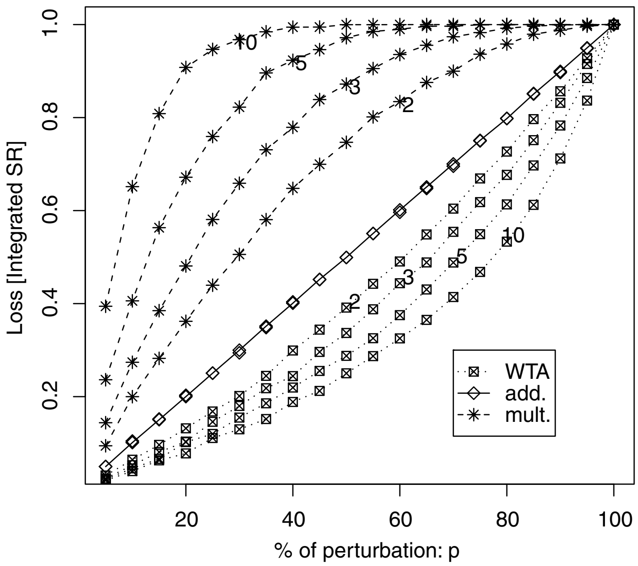

Fig. 10. *Perturbation graph*. Integrated semantic relevance loss (proportion) as a function φ of semantic knowledge deterioration (percentage) and type of relevance integration rule. For the multiplicative and WTA models the profiles are represented according to the description size factor ($q = 2, 3, 5$, and 10). For the additive rule a single averaged profile is depicted as this rule is unsensitive to q.

Table 8
Final model

	Estimate	*SE*	*z*	*p*
Additive: K_i^1				
β_0^1	−2.940	0.121	−24.23	<.001
γ^1	−1.956	0.106	−18.29	<.001
β_1^1	0.068	0.002	26.31	<.001
Multiplicative: K_i^2				
β_0^2	−0.936	0.060	−15.58	<.001
γ^2	−1.113	0.090	−12.36	<.001
β_1^2	0.0002	0.00001	20.39	<.001
WTA: K_i^3				
β_0^3	−1.677	0.088	−19.03	<.001
γ^3	−1.826	0.103	−17.62	<.001
β_1^3	0.080	0.003	21.30	<.001

Parameter point estimates and standard errors for the additive, multiplicative, and WTA integration models fitted to the observed naming accuracy y_i.

and for controls

$$\mu_i^h = \beta_0^h + \beta_1^h K_i^h. \quad (9)$$

The parameters β_0 and β_1 are, respectively, the intercept and slope for the controls group; γ^h gives the difference in intercepts between the controls and DAT groups. Table 8 reports the best MLE parameter estimates and their standard errors for the three models. In accordance with the results of Study 2 and Study 3, the fit of the model based on the additive integration rule ($\mathrm{AIC}_1 = 932.05$, $R_1^2 = 0.678$, $r_1 = 0.840$) was better than the fit yielded by the multiplicative rule ($\mathrm{AIC}_2 = 1325.4$, $R_2^2 = 0.450$, $r_2 = 0.731$) and the max rule ($\mathrm{AIC}_3 = 1392.6$, $R_3^2 = 0.411$, $r_3 = 0.668$).

For each final model we estimated the percentage of semantic degradation in the DAT group by means of a simple statistical procedure (see Appendix A.3). According to this procedure we observed the following approximated percentages of semantic degradation: $\bar{p} = 37.5\%$ (additive relevance model), $\bar{p} = 22.5\%$ (multiplicative relevance model), and $\bar{p} = 67.5\%$ (WTA relevance model). Notice that the additive model predicts a degree of semantic degradation which is proportional to the amount of integrated relevance loss (see Fig. 10). In contrast, both the multiplicative and WTA models are non-linearly related to this loss. In particular, the multiplicative model infers a mild level of semantic degradation (about 20%), whereas the WTA model predicts a very severe deterioration (approximately 70%) in the DAT group. Given that the fit of the final model based on the additive integration rule was clearly better than the fit yielded by the other two models, we may conclude that a mild-severe damage of the semantic representation can be associated to the DAT group.

Discussion

Semantic degradation may be modelled assuming that damage reduces the connection strength between semantic features and concepts (this is a widely accepted assumption; e.g. see McLeod et al., 2000). As the weight of connections between semantic features and concepts may be a way of conceptualizing relevance, the more relevant a feature for a concept, the more probable that concept will be misnamed when the feature is damaged. Hence, the behavioral consequence of damage is expected to be proportional to the relevance of the lost/damaged feature. In sum, given random damage, the likelihood of correctly retrieving a concept will be reduced proportionally to the magnitude of the damage while the same qualitative pattern among controls and patients with DAT will be reproduced. And this is what we observed.

DAT patients with a general cognitive level similar to that of our patients, as measured by the MMSE, typically show impairment on semantic tasks. Degradation of conceptual knowledge follows the severity of the disease (Garrard, Lambon Ralph, Hodges, & Patterson, 2001) and it is the contention of Gainotti, Silveri, Daniele, and Giustolisi (1995) that DAT causes widespread damage to the temporal lobes and consequently impairment of semantic knowledge. An interesting issue is the difference between DAT patients and controls in terms of the binding mechanism underlying naming to description. Results show that, rather than causing a qualitative change in the binding mechanism, the semantic impairment of patients with DAT seems to reduce the relevance values of semantic features. More precisely, when compared to controls, patients with DAT require greater values of integrated relevance to correctly retrieve concept names. This result is consistent with those neuropsychological findings that explain the degenerative semantic disorder as a widespread neuronal damage (e.g. Hodges & Patterson, 1995) that, in our terms, is likely to affect relevance values of semantic features.

Conclusions and possible extensions

The main purpose of this article was to develop and test three different integration models based on the hypothesis of semantic relevance (Sartori & Lombardi, 2004). This hypothesis reflects the idea that concepts are described by a set of semantic features, as other concept representation models do (Durrant-Peatfield, Tyler, Moss, & Levy, 1997; Humphreys & Forde, 2001), but it also maintains that each of the constituent semantic features has an associated relevance weight which is believed to represent the level of informativeness of the semantic feature for the concept. In particular, when a feature for a given target concept has both high distinctiveness (Devlin et al., 1998) and high dominance (Ashcraft, 1978), then it shows much greater semantic relevance for that concept.

No previous empirical investigations were available on how the semantic relevance of a set of features is combined in the process of naming to description. In this study three integration models of semantic features were developed and evaluated: (1) the additive rule model, in which integration reflects a linear combination of dominance values of features; (2) the multiplicative rule model, in which the integrated relevance value is defined as the multiplicative combination of the relevance values in the concept description; and (3) the max rule model, which mimics a sort of winner-take-all integration process. All three models predict that errors may arise because of low integrated relevance of semantic features. The results of the three experiments based on the naming-to-description task provide a clear support for the additive integration rule.

A related issue—still empirically unexplored—regards the modulation of different task-instruction in a naming-to-description task. In our studies the participants were required to give the concept name that corresponded to a verbally presented set of features. In contrast to this standard naming-to-description task in which no emphasis was put on the way participants were to process the set of features, a variant could explicitly require to verify that all features in the description must conjunctively belong to the target concept (we call this sort of task the *conjunctive naming-to-description* task). In this case we expect that, when a concept description contains irrelevant features for the target concept, then a multiplicative-type rule would work better than a standard additive rule. As an example, let us consider two features f_1 and f_2 with corresponding relevance values $k_{i1} = 0.05$ and $k_{i2} = 7.5$ (for a concept c_i), respectively. The results of the integration rules are: $7.55 = 0.05 + 7.5$ for standard additive, $0.375 = 0.05 \times 7.5$ for multiplicative, and $7.5 = \max\{0.05, 7.5\}$ for max rule. Clearly, the multiplicative rule better reflects the fact that a description containing features f_1 and f_2 is (conjunctively) inconsistent for c_i. In addition, in a *disjunctive naming-to-description* task in which emphasis is on the information that at least some features in the concept description set belong to c_i, then integration rules based on either an additive property or a winner-take-all mechanism would probably provide a better account for naming accuracy. Therefore, we caution to add that the predictive superiority of the additive rule observed in standard naming-to-description tasks does not necessarily imply superiority in the other possible extensions of the same experimental paradigm. In this view, our relevance integration models resemble other concept representation models suggested by those researchers who advocate that psychologically plausible semantic models should characterize features in a context-sensitive manner (e.g., Barsalou, 1982; Tversky, 1977. Tversky (1977), for example, claims that feature salience plays an important role and that it is not fixed, but varies with context. A future avenue of research could therefore be the exploration of the effect of contextual factors in the binding mechanism of concept retrieval.

Appendix A

A.1. A GLM model for naming accuracy

On the basis of both the definitions of integration rules and their corresponding assumptions, three generalized linear models (GLM, Nelder & Wedderburn, 1972) of naming accuracy, each corresponding to a different integration rule h ($h = 1, 2, 3$), were defined as follows. Each model consists of three components:

(1) The binomial random component, specifying the conditional distribution of the response variable y_i, *naming accuracy*,[9] given the predictor K_i^h.
(2) The linear predictor

$$\mu_i^h = \beta_0^h + \beta_1^h K_i^h$$

on which the expected value π_i^h of y_i depends.
(3) The invertible link function

$$\phi(\pi_i^h) = \log\left(\frac{\pi_i^h}{1-\pi_i^h}\right),$$

which transforms π_i^h to the linear predictor, and where

$$\pi_i^h = \frac{1}{1+\exp[-(\beta_0^h + \beta_1^h K_i^h)]} \tag{10}$$

is the inverse link function, $\pi_i^h = \phi^{-1}(\mu_i^h)$, (also called mean function) representing a linear logistic function.

Finally, in order to asses goodness-of-fit, we use maximized log likelihood

$$\log L = \sum_{i=1}^{I} [y_i \log p_i^h + (1-y_i)\log(1-p_i^h)],$$

where p_i^h denotes the fitted probabilities.

Implementing the PBCM procedure

In order to apply the parametric bootstrap to our data we implemented the following steps.

(1) We generated 1000 bootstrap samples from the original accuracy sample. More precisely, for each observed accuracy y_i, the resampling was done by drawing an integer number between 0 and N from a binomial distribution, $n_i^* \sim \text{Bin}(y_i, N)$, therefore the observed proportion y_i was replaced with $y_i^* = n_i^*/N$.
(2) Each of these nonparametric samples were used to estimate the parameters for the three integration relevance models. In particular for each model h ($h = 1, 2, 3$), and for each bootstrap sample m ($m = 1, 2, \ldots, 1000$) we obtained the new estimates for the intercept parameter (β_0^h) and the steepness parameter (β_1^h).
(3) Each of these new pair was then used once to generate data by first computing the probability $p_i^h[m]$ of correctly retrieving the name of the target concept c_i according to the logistic model h. Next, sampling was done by drawing an integer number between 0 and N from a binomial distribution, $n_i^h[m] \sim \text{Bin}(p_i^h[m], N)$ and a new simulated value was obtained by $y_i^h[m] = n_i^h[m]/N$. In this way, 1000 different new data sets were generated according to model h.
(4) Finally, each of the $n = 3000 (= 3 \times 1000)$ generated data set was fit by all three models, and the difference in the AIC criterion was calculated for each pair of different models. The latter step yielded three distributions of AIC differences each associated with a different pair (h, h') of integration rule models.

A.2. A Bayesian model for multiple cues integration

A simple Bayesian model for multiple cues can be defined as follows. Let A_i be the set of features corresponding to a description of concept c_i. Using the Bayes' rule and under the assumption of independence of features the Bayesian model for multiple cues is

$$P(c_i|A_i) = \frac{P(c_i)\prod_{j\in J_i} P(f_j|c_i)}{\sum_{i'=1}^{I} P(c_{i'})\prod_{j\in J_i} P(f_j|c_{i'})}, \tag{11}$$

where J_i denotes the set of indices associated with the features in A_i. According to a feature-to-listing task paradigm, the parameters of the Bayesian model can be estimated as follows:

$$p(c_i) = 1/I, \tag{12}$$
$$p(f_j|c_i) = x_{ij}/N. \tag{13}$$

Therefore, by simple algebra we can approximate $P(c_i|\, A_i)$ using the following equation:

$$p(c_i|A_i) = \frac{\prod_{j\in J_i} x_{ij}}{\sum_{i'=1}^{I}\prod_{j\in J_i} x_{i'j}}. \tag{14}$$

A.3. A simple procedure to estimate the level of semantic damage

For each final model we estimated the percentage of semantic degradation according to the following simple statistical procedure. First, we computed the proportion of integrated relevance loss in the DAT group using the estimated intercept values, β_0^h (for controls) and $\beta_0^h + \gamma^h$ (for patients with DAT), and the following equation:

$$l_{\text{DAT}}^h = 1 - \frac{|\beta_0^h|}{|\beta_0^h + \gamma^h|}, \quad h = 1, 2, 3.$$

For example, to compute the integrated relevance loss for the additive relevance model we have

$$l_{\text{DAT}}^1 = 1 - \frac{|-2.940|}{|-2.940 + (-1.956)|} = 0.399$$

(see Table 8). This proportion can be represented in the y-axis of the perturbation graph (Fig. 10). Next, we derived the approximate amount of semantic deterioration in the DAT

[9] Naming accuracy y_i is defined as the proportion of subjects who correctly retrieve the name of concept c_i given A_i.

group by averaging the two consecutive percentages, p_1 and p_2 (with $p_1 < p_2$) in the perturbation graph (Fig. 10, *x*-axis), such that $l^h_{\mathrm{DAT}} \in [\varphi_q(p_1), \varphi_q(p_2)]$, where q denotes the size of the concept description in the naming-to-description task. So for example, in Fig. 10, we have that $0.399 \in [\varphi_3(35), \varphi_3(40)]$, where $p_1 = 35$ and $p_2 = 40$ represent the two consecutive percentages in the perturbation graph satisfying the interval condition. This yields 37.5% of semantic deterioration.

References

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*, 409–429.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*, 154–179.

Ashcraft, M. H. (1978). Property norms for typical and atypical items from 17 categories: a description and discussion. *Memory & Cognition, 6*, 227–232.

Barsalou, L. W. (1982). Context-dependent and context-independent information in concepts. *Memory & Cognition, 10*, 82–93.

Chertkow, B., & Bub, D. (1990). Semantic memory loss in dementia of Alzheimer's type. What do various measures measure? *Brain, 113*, 397–417.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General, 132*, 163–201.

Cree, G. S., McNorgan, C., & McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 32*, 643–658.

Davies, C. A., Mann, D. M. A., Sumpter, P. Q., & Yates, P. O. (1987). A quantitative morphometric analysis of the neuronal and synaptic content of frontal and temporal cortex in patient with alzheimer's disease. *Journal of Neurological Sciences, 78*, 151–164.

Dell'Acqua, R., Lotto, L., & Job, R. (2000). Naming time and standardized norms for the Italian PD/DPSS set of 266 pictures: direct comparisons with American, English, French, and Spanish published databases. *Behaviour Research Methods, Instruments and Computers, 32*, 588–612.

Devlin, J. T., Gonnerman, L. M., Andersen, E. S., & Seidenberg, M. S. (1998). Category-specific semantic deficits in focal and widespread brain damage: a computational account. *Journal of Cognitive Neuroscience, 10*, 77–94.

Durrant-Peatfield, M. R., Tyler, L. K., Moss, H. E. & Levy, J. P. (1997). The distinctiveness of form and function in category structure: a connectionist model. In: *Proceedings of the nineteenth annual conference of the cognitive science society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology, 18*, 500–549.

Estes, W. K. (1993). Models of categorization and category learning. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.), *The psychology of learning and motivation: Categorization by humans and machines* (pp. 15–56). San Diego: Academic Press.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: modality specificity and emergent category specificity. *Journal of Experimental Psychology: General, 4*, 339–357.

Folstein, M. F., Folstein, S. E., & Mc Hugh, P. R. (1975). "Mini Mental State": A practical method for grading the cognitive state of patients for clinician. *Journal of Psychiatric Research, 12*, 189–198.

Forde, E. M. E., & Humphreys, G. W. (2000). Category-specific recognition impairments: a review of important case studies and influential theories. In G. Cohen, R. Johnston, & K. Plunkett (Eds.), *Exploring cognition: damaged brains and neural networks (Readings in cognitive neuropsychology and connectionist modelling)*. Psychology Press and Open University.

Friedman, D., Massaro, D. W., Kitzis, S. N., & Cohen, M. M. (1995). A comparison of learning models. *Journal of Mathematical Psychology, 39*, 164–178.

Gainotti, G., Silveri, M. C., Daniele, A., & Giustolisi, L. (1995). Neuroanatomical correlates of category-specific semantic disorders: a critical survey. *Memory, 3*, 247–264.

Garner, W. R., & Morton, J. (1969). Perceptual independence: definitions, models, and experimental paradigms. *Psychological Bulletin, 72*, 233–259.

Garrard, P., Lambon Ralph, M., Hodges, J. R., & Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelations: analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology, 18*, 125–174.

Gigerenzer, G., Hoffrage, U., & Kleinblting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.

Hachinski, V. G., Iliff, L., Du Boulay, G. H., McAllister, V. L., Marshall, J., Ross Russell, R. W., et al. (1975). Cerebral blood flow in dementia. *Archieve of Neurology, 32*, 633–637.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98*, 74–95.

Hodges, J. R., & Patterson, K. (1995). Is semantic memory consistently early in the course of Alzheimers disease? Neuroanatomical and diagnostics implications. *Neuropsychologia, 33*, 441–459.

Humphreys, G. W., & Forde, E. M. E. (2001). Hierarchies, similarity and interactivity in object recognition: "Category-specific" neuropsychological deficits. *Behavioral and Brain Sciences, 24*, 453–509.

Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22–44.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 4*, 463–470.

Lambon Ralph, M. A., Graham, K. S., Ellis, A. W., & Hodges, J. R. (1998). Naming in semantic dementia—what matters? *Neuropsychologia, 36*, 775–784.

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

Marques, J. F. (2005). Naming from definition: the role of feature type and feature distinctiveness. *The Quarterly Journal of Experimental Psychology: Section A, 58*, 603–611.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1990). Models of integration given multiple sources of information. *Psychological Review, 97*, 22–252.

Mechelli, A., Sartori, G., Orlandi, P., & Price, C. J. (2006). Semantic relevance explains category effects in medial fusiform gyri. *Neuroimage, 30*, 992–1002.

Medin, D. L. (1983). Structural principles in categorization. In T. J. Tighe & B. E. Shepp (Eds.), *Perception cognition, & development: Interactional analyses* (pp. 203–230). Hillsdale, NJ: Erlbaum.

Medin, D. L., & Schaffer, M. M. (1978). A context theory of classification learning. *Psychological Review, 85*, 207–238.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadan, E. M. (1984). Clinical diagnosis of Alzheimer disease: Report of the NINCDS-ADRDA work group under the auspices of the Department of Health and Human Services Task Force on Alzheimer disease. *Neurology, 34*, 939–944.

McLeod, P., Shallice T., & Plaut, D. C. (2000). Attractor dynamics in word recognition: converging evidence from errors by normal subjects, dyslexic patients and a connectionist model. *Cognition, 74*, 91–114.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126*, 99–130.

Murdock, B. B. (1993). TODAM2: a model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review, 100*, 183–203.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A (General), 135*, 370–384.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 104–114.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 87–108.

Randall, B., Moss, H. E., Rodd, M. G., & Tyler, L. K. (2004). Distinctiveness and correlation in conceptual structure: behavioral and computational studies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 393–406.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). The structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychological Review, 111*, 205–235.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*, 192–233.

Rosch, E. (1977). Human Categorization. In N. Warren (Ed.), *Advances in Cross-Cultural Psychology*. London: Academic Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27–48). Hillsdale NJ: Erlbaum.

Salton, G. (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. Reading, MA: Addison-Wesley.

Salton, G. (1991). Developments in automatic text retrieval. *Science, 253*, 974–979.

Salton, G., & Buckley, C. (1988). Term weighting approach in automatic text retrieval. *Information Processing and Management, 24*, 513–523.

Sartori, G., & Lombardi, L. (2004). Semantic relevance and semantic disorders. *Journal of Cognitive Neuroscience, 16*, 439–452.

Sartori, G., Lombardi, L., & Mattiuzzi, L. (2005a). Semantic relevance best predicts normal and abnormal name retrieval. *Neuropsychologia, 43*, 754–770.

Sartori, G., Polezzi, D., Mameli, F., & Lombardi, L. (2005b). Feature type effects in semantic memory: an event related potentials study. *Neuroscience Letters, 390*, 139–144.

Sartori, G., Mameli, F., Polezzi, D., & Lombardi, L. (2006). An ERP study of low and high relevance semantic features. *Brain Research Bulletin, 69*, 182–186.

Selfridge, O. G. (1959). Pandemonium: a paradigm for learning. *In Mechanisms of thought processes* (pp. 511–526). London: Her Majesty's Stationery Office.

Shepard, R. N. (1957). Stimulus and response generalization: a stochastic model, relating generalization to distance in psychological space. *Psychometrika, 22*, 325–345.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*, 1317–1323.

Silveri, M. C., & Gainotti, G. (1988). Interaction between vision and language in category-specific semantic impairment. *Cognitive Neuropsychology, 5*, 677–709.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: the early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1411–1436.

Smits, T., Storms, G., Rosseel, Y., & De Boeck, P. (2002). Fruits and vegetables categorized: an application of the generalized context model. *Psychonomic Bulletin and Review, 9*, 836–844.

Storms, G., De Boeck, P., & Ruts, W. (2001). Categorization of unknown stimuli in well-known natural concepts: a case study. *Psychonomic Bulletin & Review, 8*, 377–384.

Tversky, A. (1977). Features of similarity. *Psychological Review, 34*, 327–352.

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

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology, 48*, 28–50.

Witten, I. H., & Bell, T. C. (1991). The zero-frequency problem: estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions on Information Theory, 37*, 1085–1094.