Concept typicality responses in the semantic memory network

Andrea Santi[a,b,*], Ana Raposo[a], Sofia Frade[a], J. Frederico Marques[a]

[a] Universidade de Lisboa, Faculdade de Psicologia, 1649-013 Lisboa, Portugal
[b] University College London, Department of Linguistics, WC1N 1PF England, UK

ARTICLE INFO

ABSTRACT

For decades concept typicality has been recognized as critical to structuring conceptual knowledge, but only recently has typicality been applied in better understanding the processes engaged by the neurological network underlying semantic memory. This previous work has focused on one region within the network – the Anterior Temporal Lobe (ATL). The ATL responds negatively to concept typicality (i.e., the more atypical the item, the greater the activation in the ATL). To better understand the role of typicality in the entire network, we ran an fMRI study using a category verification task in which concept typicality was manipulated parametrically. We argue that typicality is relevant to both amodal feature integration centers as well as category-specific regions. Both the Inferior Frontal Gyrus (IFG) and ATL demonstrated a negative correlation with typicality, whereas inferior parietal regions showed positive effects. We interpret this in light of functional theories of these regions. Interactions between category and typicality were not observed in regions classically recognized as category-specific, thus, providing an argument against category specific regions, at least with fMRI.

1. Introduction

Conceptual categories help us make sense of the world, quickly, knowledgably, pragmatically, and on some occasions unfortunately also inaccurately (e.g., stereotypes). A fundamental property of concepts is their graded category membership, where some members are more typical, or a better exemplar of the category, than others (Rosch and Mervis, 1975). For example, a robin is a more typical bird than is a penguin. Typicality has been explained by family resemblance (Barsalou, 1985; Marques and Raposo, 2011; Marques et al., 2013; Rosch and Mervis, 1975; Rosch et al., 1976), whereby a category member that both shares many features with other members and few features with members of other categories, is more typical. Typicality has been demonstrated for both natural and artificial categories at superordinate and basic levels (Rosch and Mervis, 1975; Rosch et al., 1976). The correlation between feature sharing within a category and typicality has been referred to as feature sharedness (Raposo et al., 2012). Indirect evidence for feature sharedness is provided by *typicality effects*: in category verification tasks, faster response times are observed for more typical items (e.g. deciding if a robin is a bird) than less typical items (e.g. deciding if a penguin is a bird; Casey, 1992; Hampton, 1979; Kiran et al., 2007; Larochelle and Pineau, 1994; Rosch, 1975). Presumably, feature comparisons across category members are required for task completion; hence, the more shared features, the faster the categorization time (e.g. Dry and Storms, 2010; Rosch

and Mervis, 1975; Smith et al., 1974). Typicality also predicts response times in naming tasks (Holmes and Ellis, 2006). The degree to which typicality is related to other measures that also predict such response times, such as frequency, familiarity, and age-of-acquisition, is variable. Typical items are generally familiar, but atypical items can either be familiar or not (Glass and Meany, 1978). In the case of word frequency, there are conflicting results, with one study demonstrating a relationship with typicality (Holmes and Ellis, 2006) and another failing to (Mervis et al., 1976). Critically, however, typicality provides predictive power above all three of these measures on task performance in healthy participants and patients with semantic impairments and to our knowledge there is no available counter-evidence to this effect (Barsalou, 1985; Casey, 1992; Hampton, 1979; Kiran et al., 2007; Larochelle and Pineau, 1994; Marques, 2007; McCloskey, 1980; Woollams, 2012; Woollams et al., 2008). These data demonstrate that typicality provides a unique and fundamental dimension to the organization of conceptual knowledge.

This evidence also seems to imply that concept typicality is critical to the neurological implementation of concepts. Yet, decades of studies investigating the neural substrates underlying semantic memory have ignored this factor, and it is only in recent years that typicality has been considered. In a series of recent studies, Woollams and colleagues, using a picture-naming task, have investigated the role of concept typicality in both Semantic Dementia (SD) patients and healthy participants following Transcranial Magnetic Stimulation (TMS)

* Correspondence to: Department of Linguistics, University College London, Chandler House, 2 Wakefield Street, WC1N 1PF London, UK
  E-mail address: a.santi@ucl.ac.uk (A. Santi).

(Woollams, 2012; Woollams et al., 2008). These studies have focused on a specific brain region, the Anterior Temporal Lobe (ATL), and its potential role in amodal feature integration. We will review these relevant studies before considering additional theoretical perspectives that predict that other cortical regions, notably the Inferior Frontal Gyrus, and Inferior Parietal Lobe, should also be sensitive to typicality (Binder et al., 2010; Visser et al., 2012). Moreover, we will consider how typicality might interact with categories, which would be insightful for understanding the nature of previously observed category effects (i.e. the observed difference in deficits and activation across category domains). This has not previously been explored. We will then present a functional Magnetic Resonance Imaging (fMRI) study to provide a novel wide angle view on the role of concept typicality within the entire semantic memory network and its interaction with the categories, Natural Kinds and Artefacts, using a category verification task while varying the degree of concept typicality.

Most researchers agree that the representation of concepts activates the same brain regions that are engaged during perception and action with that same concept (Patterson et al., 2007). Thus, in representing a bird we activate the same motion, sound, and shape areas that are engaged when perceiving birds. Further, there is compelling evidence that sensory areas feed into a pathway running from posterior in the temporal lobe to anterior aspects (Scott et al., 2000), where the most anterior regions are critical to the representation of specific, unique concepts (Tyler et al., 2004). This proposal has gained support from both neuroimaging data in healthy participants as well as from studies with SD patients. SD patients have relatively constrained damage to the bilateral ATL, along with selective deficits to conceptual knowledge, which are most pronounced for specific concepts. For example, SD patients demonstrate greater accuracy in naming a picture of a horse as "animal" than its more specific basic-level name "horse" (Rogers and Patterson, 2007). Interestingly, this performance pattern is opposite to that of healthy participants, who are more accurate on the specific (basic) than the general level (Rogers and Patterson, 2007). Some have proposed that the ATL works as a semantic hub, which amodally integrates conceptual information and forms abstractions or generalizations across categories (for review, see Patterson et al., 2007). According to this view, general categories are more robust to damage given that the features have greater co-occurrence and as such are less susceptible to damage.

Woollams (2008; 2012) has recently argued that these *specificity effects* can be reframed in terms of typicality. That is, rather than specific concepts being impaired, it is the atypical features or the weaker co-occurrence of features of atypical concepts that is susceptible to damage. The data demonstrate that SD patients have better picture naming performance with more typical items than less typical items (Woollams, 2012; Woollams et al., 2008). Moreover, typicality uniquely predicted naming accuracy when frequency, familiarity, age-of-acquisition, and domain category were included in the model. Thus, typicality effects in naming deficits cannot simply be reduced to other factors.

In addition to studying typicality in SD patients, the same researcher has carried out similar studies with healthy participants through application of TMS to the ATL offline, likewise, using a picture-naming task. Woollams (2012) demonstrated more performance interference for atypical than typical items. Hence, disruption to activation in the ATL, either through a temporary, virtual or natural, degenerative, lesion provides greater impairment to performance on the atypical items. This finding has been used to argue that the weaker the co-occurrence of features (ie, the less typical), the greater the (need for) activation of the ATL (ie, a negative relationship between typicality and activation). These results have been further used to support the claims that the ATL is an amodal semantic hub that represents concepts through the co-occurrence of constituent features.

Despite these studies, amongst others, providing support for the ATL as being an amodal center in forming conceptual generalizations,

challenges to this perspective have been presented. Neuroimaging data have overwhelmingly provided evidence for amodal integration lying outside the ATL, such as posteriorly in the temporal lobe and/or inferior parietal lobe (Binder et al., 2010). A recent meta-analysis (Binder et al., 2010) demonstrated that the core semantic regions includes the posterior to anterior middle temporal gyrus, posterior superior temporal gyrus, inferior parietal lobe and inferior frontal gyrus (IFG). Both the inferior parietal and superior-middle temporal activation were interpreted as heteromodal integration cortices. The IFG was interpreted in terms of processing "efficiency", but not necessarily storing semantic representations.

A limitation of fMRI is that it is particularly susceptible to signal distortion and dropout in the ATL, particularly in its inferior part (eg, Devlin, et al., 2000; Visser et al., 2010). This is the location of Woollam's TMS application. Absence of inferior ATL activation in the meta-analysis therefore does not imply the region is not engaged in amodal feature integration. However, the finding of activation outside of the ATL in tasks requiring amodal integration is consistent with *other* regions playing an integrative role in binding features.

Indeed those advocating a role for an amodal store in the ATL have also recently demonstrated that posterior aspects of the temporal lobe may also serve feature integration functions (Visser et al., 2012). This distortion-corrected fMRI study (providing sensitivity to detect inferior ATL activation) investigated regions activated during both word and picture versions of the Pyramids and Palm Trees test. They found widespread activation of the anterior (including inferior portions) and posterior temporal lobe as well as inferiorly in the parietal lobe and frontal lobe for both modalities. However, they alternatively claimed that the inferior parietal lobe serves a role in executive functioning, similar to that which has been proposed for the IFG, rather than feature integration. They argue that the IFG and inferior parietal lobe engage in semantic control, but do not provide a permanent store for conceptual representations. Semantic control includes processes to manipulate and actively store online conceptual representations. An area engaged in semantic control would be expected to demonstrate *Negative Typicality Effects* similarly to the ATL, that is, greater activation the lower the typicality of the item, as the atypical items would also be more difficult to *retrieve* given their lower feature co-occurrence. Thus, it is of interest to investigate typicality in regions outside the ATL that have likewise been claimed to serve a feature integration role and/or semantic control, that is the inferior parietal lobe, posterior temporal lobe and IFG. The reverse result of more activation, the more typical the item (ie, *Positive Typicality Effects*), would be consistent with an area that engages in similarity based categorization (ie, typical exemplar more similar to prototype), which has been observed in the right parietal cortex (Grossman et al., 2002). *Positive Typicality Effects* would not necessarily be incompatible with an integration function, but one that differs in nature from the ATL and potentially more concerned with the overall configuration of features, where more prototypical configurations generate greater activation due to greater similarity to a prototype in line with a proposal made by Grossman et al. (2002). Thus we will run the fMRI study that parametrically varies typicality in order to see which regions are sensitive to typicality. FMRI, unlike the patient and TMS studies mentioned previously in assessing typicality, provides whole-brain maps to see the entire semantic memory network's response to typicality. A second intriguing question this study will address is whether typicality interacts with domain categories. This is particularly relevant to understanding the category effects previously observed in both healthy and patient populations. Within healthy subjects, some cortical regions demonstrate distinct activation patterns for specific categories (or domains) over others (e.g., animals vs tools) (Chao and Martin, 2000), including the anterior medial temporal lobes (Devlin et al., 2002). This complements the category specific deficits observed in patients with damage to like regions (Caramazza and Shelton, 1998; Mahon and Caramazza, 2008). Woollams et al. (2008) ran one of the