Concept Typicality Responses in the Semantic Memory Network

Andrea Santi[1,2], Ana Raposo[1], Sofia Frade[1], J. Frederico Marques[1]

[1] Universidade de Lisboa, Faculdade de Psicologia, Portugal, 1649-013 Lisboa

[2]University College London, Department of Linguistics, England, WC1N 1PF

Corresponding Author:

Andrea Santi

Department of Linguistics, University College London

Chandler House 2 Wakefield Street London UK, WC1N IPF

Email: a.santi@ucl.ac.uk

Phone: +44 (0) 207 679 4017

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.

Key words: concepts, typicality, fMRI, semantic memory

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 & 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 & Raposo, 2011; Marques, Raposo, & Almeida, 2013; Rosch & Mervis, 1975; Rosch, Simpson, & Miller, 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 & Mervis, 1975; Rosch et al., 1976). The correlation between feature sharing within a category and typicality has been referred to as feature sharedness (Raposo, Mendes, & Marques, 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, Ntourou, & Eubank, 2007; Larochelle & 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 & Storms, 2010; Rosch & Mervis, 1975; Smith, Shoben, & Rips, 1974). Typicality also predicts response times in naming tasks (Holmes & 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 & Meany, 1978). In the case of word frequency, there are conflicting results, with one study demonstrating a relationship with typicality (Holmes & Ellis, 2006) and another failing to (Mervis, Catlin, & Rosch, 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 & Pineau, 1994; Marques, 2007; McCloskey, 1980; Woollams, 2012; Woollams, Cooper-Pye, Hodges, & Patterson, 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) (Woollams, 2012; Woollams, Cooper-Pye, Hodges, & Patterson, 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, Desai, Graves, & Conant, 2010; Visser, Jefferies, Embleton, & Lambon Ralph, 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, Nestor, & Rogers, 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, Blank, Rosen, & Wise, 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 & 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 & 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, Embleton, Jefferies, Parker, & Lambon Ralph, 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 wide-spread 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 & 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 & Shelton, 1998; Mahon & Caramazza, 2008). Woollams et al. (2008) ran one of the largest studies of SD patients and investigated, in addition to typicality, domain category effects on naming, given previous inconsistent findings on the matter. Woollams et al. (2008) found significant though small domain effects. However, when including familiarity, frequency, age-of-acquisition and typicality in modeling naming accuracy, the effect disappeared. Neuroimaging studies that control for most of these factors have observed category effects in a subset of the ATL (Devlin et al., 2002), suggesting that only a limited section of the ATL demonstrates such effects. Category effects have more consistently been observed in patient populations other than SD, such as herpes simplex encephalitis patients presenting a category specific impairment (Warrington & Shallice, 1984). These effects are not all-or-none, but rather proportionally greater for one category over another. This proportional difference may be explained by category typicality. That is, typical members may be those that are spared and atypical members impaired. If an area is category specific, it should demonstrate a category effect when those categories are considered in terms of their gradation in typicality, given this is the fundamental defining property of categories. That is, there should be a correlation with concept typicality in the category that the region is selective for and no (or at least less) correlation with typicality of the alternative categories. These potential interaction effects between typicality and category have not yet been tested. If on the other hand, *Category Effects* are due to differences in processes or feature sharing engaged by the categories, then no such interaction with typicality would be expected.

In summary, the neurological studies that considered typicality in studying the neurobiology of semantic memory have been limited to the ATL. In order to better assess all regions potentially sensitive to typicality, we conducted an fMRI study, which allows the entire network to be observed. Unlike previous studies that used a picture-naming task, we used a category verification task with verbal stimuli in which concept typicality was parametrically manipulated. Category verification more tightly links to typicality and provides another task and modality to test

typicality effects. In terms of the ATL, and following prior studies, we might expect an increase in activation as concept typicality decreases (i.e. *Negative Typicality Effects*). Given the limitations of fMRI, this was not our main question of interest, however. In terms of relevant regions not previously tested for typicality, we expect Negative Typicality Effects in: (1) posterior temporal lobe, if it plays the same integrative role as proposed for the ATL, (2) inferior parietal lobe, if it plays either the same integrative function as the ATL or if it is involved in semantic control, (3) IFG if it is associated with semantic control. The reverse result, that is, more activation for more typical items (ie, *Positive Typicality Effects*), is expected in the right parietal cortex, an area that is engaged in similarity based categorization, as more typical exemplars are more similar to the prototype (Grossman et al., 2002). Lastly, we expect to find *Category Effects* in regions previously observed for Natural Kinds vs Artefacts (ie, medial ATL) and Artefacts vs Natural Kinds (ie, posterior temporal, fusiform gyrus, premotor), by collapsing over the superordinate categories belonging to these domains. Regions that are truly category specific would be expected to demonstrate a category effect when those categories are considered in terms of their gradation in typicality. That is, there should be a greater correlation with typicality in the category the region is specific to over the other.

2 Material & Methods

2.1 Participants

Twenty right-handed, healthy participants, native speakers of Portuguese (17 females, M=19.65 years, range: 18-29 years) took part in the study. Two participants were excluded from fMRI data analysis because of a corrupted file or too much motion (>6mm in one of the runs). All gave informed written consent to the experimental procedure, which was approved by the local ethics committee.

2.2 Materials

Two natural kind categories (mammals, fruit) and two of artefact categories (clothing, vehicles) were used. For each category, we selected 24 exemplars with

different levels of typicality and familiarity (see Table 1; see Appendix A for list of items). The typicality and the familiarity rates were obtained in a previous pre-test (n=24, n=29, respectively), in which a different group of participants judged on a 7 point scale how typical the exemplar was of a certain category (1= very atypical to 7=very typical) and how familiar that item was (1=very unfamiliar to 7=very familiar). The mean length was 6,08 (1,73) characters, which was matched across the four categories, as determined through a non-significant effect of category in a one-way ANOVA of character length ($p>0.05$). These items were used in the true condition, in which the exemplar was presented along with the category it belongs to. Additionally, 96 items from others categories (e.g., vegetables, kitchen tools) were selected to create false trials, in which the item did not belong to the category presented.

_____

Insert Table 1. here

_____

2.3 Procedure

Participants performed a category verification task in which they had to decide if an item belonged to a given category. Each trial started with a fixation cross for 500ms, followed by the visual presentation of the category name (750ms) and after an inter-stimuli interval of 200ms the target item was presented as text for 2050ms. During the presentation of the target, participants had to decide if that target belonged to the category previously presented, by pressing a button with their left index or middle finger. In half of the trials the target belonged to the category previously presented, while in the other half the target was presented from a different category. The prime category was always Mammals, Fruit, Clothing, or Vehicles, such that participants could not anticipate a "False" response. Trials were separated by a variable inter-trial interval (1500, 2000, 2500 and 3000ms) in order to optimize statistical efficiency (Dale, 1999). The task was divided into three blocks, each one with 64 trials. Overall, the session included 192 trials and lasted

approximately 19 minutes. Presentation and timing of stimuli were controlled using EPrime software (www.psnet.com).

2.4 fMRI Parameters

Scanning was conducted at Sociedade Portuguesa de Ressonância Magnética on a 3-Tesla Philips MR system (Philips Medical Systems, Best, NL) using a standard head coil. Functional data were acquired by using an echo-planar sequence (TR = 2000 ms, 34 bottom-up interleaved slices parallel to the AC-PC line, with isotropic voxels, 3mm thick, interslice gap of 0.5mm, 2mm×2mm in-plane resolution, FOV= 23cm×23cm, matrix size = 116×115). Acquisition covered the entire brain. Before functional data collection, three dummy volumes were discarded to allow for T1 equilibrium. High-resolution T1-weighted anatomical images were acquired for visualization.

2.5 Data Analysis

2.51 Behavioral Data Analysis

The response time (RT) data for True Targets responses were log transformed and analyzed with a mixed effects model using lme4 in R (Bates and Sarkar 2007). *Participants* and *Items* were treated as random effects and *Typicality, Category, Familiarity, Category*Typicality, Category*Familiarity,* were treated as fixed effects. The effect of Category was assessed with sum-coded contrasts. Both random slopes and intercepts were included for the *Participant* effects (excluding interactions) and intercepts only for the *Item* effects (Barr, 2013). Effects were tested for their significant contribution to the prediction of logRT. For continuous fixed effects variables, significance was assessed through treating the t-value as a z-statistic (Barr, 2013), while for polytomous variables (ie, Category and its interaction with familiarity and Typicality) it was assessed through a model comparison, with the only difference between models being the presence of the predictor or not. The same random effects structure was used for each model comparison.

2.52 fMRI Data Analysis

The fMRI data were preprocessed and statistically analyzed using Statistical Parametric Mapping Software (SPM12, Wellcome Institute of Cognitive Neurology, www.fil.ion.ucl.ac.uk) within Matlab (Mathworks Inc., Sherborn MA, USA). First we corrected for differences in slice acquisition timing by resampling all slices to the middle slice, these data were then corrected for motion across all sessions by aligning to the first session of three collected. In most functional runs the maximum movement was under 1mm in any direction, with exception of the y-plane where most runs had a maximum movement in the 1-3mm range. Exceptionally, one subject had a maximum movement in the y-plane that was >6mm in one run and that subject was eliminated from the fMRI data analysis. 11 runs (out of 57) had a maximum movement between 1-2mm in the z-plane and 1 run had a maximum movement around 1mm in the x-plane. The mean resliced functional data was coregistered to the participants' T1. The T1 was then segmented and normalization parameters provided. The parameters were applied to the preprocessed functionals and then spatially smoothed with an 8mm FWHM filter.

The data were modeled in two ways. The first model (*Category Typicality*) factored in a typicality parameter and in total included 15 regressors (1 prime, 1 ISI, 4 True Category Targets, 4 True Category Target Typicality Parameter, 4 True Category Target Familiarity Parameter, 1 False target) as well as 6 nuisance regressors for the motion parameters. Familiarity was included in the model to extract out these effects, as familiarity significantly contributed to the prediction of logRT in the behavioral data analysis[1]. To forecast the behavioral results *Familiarity* and *Typicality* along with *TargetCategory* uniquely and significantly predicted logRT in the current task. The second model (*Category*) did not consider the typicality parameter and included 7 regressors (1 prime, 1 ISI, 4 True Category Targets, 1 False target) as well as 6 nuisance regressors for the motion parameters. We modeled each of the true targets separately (mammals, vehicles, fruit, clothing), but

[1] Note that the typicality parameter was entered into the model before the familiarity one. In SPM the second parameter only accounts for variability that is not accounted for by the first parameter.

included only 1 regressor for all false targets, as we were not interested in the false targets.

The data were modeled using the canonical hemodynamic response function (HRF) for each subject. The contrast files for each category predictor and its typicality parameter (in the case of the *Category Typicality* model) relative to baseline (fixation cross) were stored for use in second-level analyses. Additionally for the *Category Typicality* model the overall typicality effect (assigning a 1 to all 4 category typicality parameters) relative to baseline was stored.

We entered the overall typicality effect of all true targets from the *Category Typicality* model into a second-level one-sample t-test. Here we were interested in two effects: (1) areas that demonstrated an increase in activation with decreasing typicality (*Negative Typicality Effects*; -1), and (2) areas that demonstrated an increase in activation with increasing typicality (*Positive Typicality Effects*; 1).

Additionally we entered each of the four individual category typicality contrasts into separate, second-level, one-sample t-tests. These results were used to mask the overall typicality effects to observe areas that demonstrate typicality effects (positive and negative) across all 4 categories.

To investigate an interaction between typicality and category we ran a within subject analysis of variance (ANOVA) using the flexible factorial tool, whereby there was a subject factor (18 levels) and a category factor including each of the four category typicality contrasts from the *Category Typicality* model (four levels; Henson & Penny, 2003). The 3 contrasts of interest were: (1) a t-test of Natural kinds > Artefacts (1, -1, 1, -1), (2) a t-test of Artefacts > Natural Kinds (-1, 1, -1, 1), and (3) an F-test of effects of interest to plot contrast estimates for each category typicality effect.[2]

To study Category Effects (independently of typicality), as they have previously been identified in the literature, we entered all true targets from the

[2] While the flexible factorial tool can be used to plot the individual category contrast estimates, it is inappropriate for statistical analysis of typicality effects overall or for those of individual categories compared to baseline. This is because these contrasts involved a between subject error term and the flexible factorial uses a within subject error term.

Category model into a 4-way ANOVA (mammals, fruit, clothing, vehicles) using the flexible factorial tool, whereby there was a subject factor (18 levels) and a category factor (four levels). The two t-tests of interest were: (1) Natural kinds > Artefacts [1, -1, 1, -1] and (2) the reverse contrast of Artefacts > Natural kinds [-1, 1, -1, 1].

To study effects of Category with their typicality effects regressed out, we ran a within subject analysis of variance (ANOVA) using the flexible factorial tool, whereby there was a subject factor (18 levels) and a category factor including each of the four category contrasts (ie, not the parametric variation in typicality) from the *Category Typicality* model (four levels; Henson & Penny, 2003). The 3 contrasts of interest were: (1) a t-test of Natural kinds > Artefacts (1, -1, 1, -1), (2) a t-test of Artefacts > Natural Kinds (-1, 1, -1, 1).

Maps were thresholded at voxel-wise $p=0.001$, cluster size = 20 voxels, and then clusters that survived $p<0.05$ FDR were reported.

3 Results

3.1 Behavioral Results

Participants were very accurate in their responses across all categories with an average accuracy of 85% or above for all True Target categories (see Figure 1). We report here accuracy of True Targets only, because the fMRI data analysis focuses in the processes engaged during categorization and not those involved in rejecting an exemplar as a category member, since in this case it is more difficult to know what processes are being engaged. In addition to being accurate, the participants were also relatively quick to respond across all true target categories with an average of 991.67ms or lower (see Figure 2).

Insert Figure 1 & Figure 2 here

In predicting logRT, neither *Target Category* nor any of the interaction effects significantly contributed to explaining the variance. The factors that added

significantly to the prediction of logRT included *Typicality (B=-0.062 SE=0.009 t=-7.10 p<0.001)* and *Familiarity (B=-0.024 SE=0.009 t=-2.64 p=0.008)*[3].

3.2 fMRI Data

3.21 Negative Typicality Effects

Increasing activation with decreasing concept typicality demonstrated core activation in bilateral IFG and medially in supplementary motor cortex (see Table 2, Figure 3A). Additionally, consistent with the hypotheses mentioned at the outset, there was a small cluster of activation in an inferior temporal region encroaching on the left ATL. When masking these effects with the individual contrasts for a *Negative Typicality Effect* within each individual category (voxel-wise p<.05, see Figure 3C), activation is confined to bilateral IFG and supplementary motor cortex, however. The Negative Typicality estimates for each Category taken from the peak coordinate in the inferior temporal lobe (see Figure 3B), indicates that the Clothing category is the primary category not producing *Negative Typicality Effects* in the ATL. When looking at the same plot but using the coordinate from Woollams (2012) TMS study (-53, 4, -32), which is a bit anterior to ours, we find the mammal category to be the only category that demonstrates a significant Negative Typicality Effect.

Insert Table 2 & Figure 3 here

3.22 Positive Typicality Effects

Increasing activation with increasing typicality was demonstrated primarily in bilateral inferior parietal, posterior temporal, and precuneal regions (see Table 2 and Figure 4A). When masking with the individual contrasts for a *Positive Typicality Effect* within each category (voxel-wise p<.05, see Figure 4C), the activation is

[3] This analysis is based on RTs from both correct and incorrect responses to True Targets (only those trials where the participant did not respond were not included). However, the results are equivalent when only correct response RTs to True Targets are included.

confined to the right hemisphere. This indicates that activation in the left hemisphere is not present for all categories (at least not significantly).

Insert Figure 4 here

Positive Typicality Effects within each category were investigated by plotting their contrast estimates at the peak within the left posterior temporal region (see Figure 4B), which indicates the effects were greatest for Mammals.

3.23 Category Effects

Category effects (independent of typicality) were observed in a variety of regions that are consistent with previous reports. Considering the Natural Kinds > Artefacts contrast (see Table 3, Figure 5), we found activation in medial anterior temporal cortex, as has been observed by Devlin et al. (2002), namely at coordinates (24, -8, -24) and (-30, 6, -18). We also observed activation in left prefrontal regions, which has been found to be activated when making typicality judgments of Natural Kinds, although in that study the activation was more superior and posterior, engaging primarily motor cortex (Grossman et al., 2006). The activations found in LIFG likely reflect differences across the categories in terms of difficulty (i.e., typicality). Areas more predominantly cited as category specific (i.e., fusiform gyrus) were not observed. However, activation of the fusiform gyrus has not always been reliably replicated, particularly when the stimuli are written words, involving categories other than mammals and tools (as our stimuli do; Devlin et al., 2005). Furthermore, when we run the same contrast defined by the typicality parameter over this thresholded map, we see no activated voxels.

Insert Table 3 and Figure 5 and 6 here

Regarding the contrast of Artefacts > Natural kinds (see Table 3 and Figure 6) we replicate previous findings of activation in postero-lateral temporal cortex (Devlin et al., 2002). Again, we did not see category effects in the fusiform gyrus, as has

been previously observed, but this might reflect the lexical stimuli and particular categories studied. Similarly to the results of the previous contrast, when we run the same contrast defined by the typicality parameter over this thresholded map, we see no activated voxels.

Looking at category effects with typicality regressed did not differ from the results from the overall Category Effects. The only minor change was that one of the clusters (coordinates: -16 2 -22) in the Natural Kinds > Artefacts map becomes marginally significant (p=.059).

Areas that do demonstrate a main effect of category [Natural kinds > Artefacts] defined by the typicality parameter include the Angular Gyrus, posterior temporal, inferior occipital and inferior frontal (see Figure 7). The posterior temporal activations are more posterior than in the model that does not consider typicality and the inferior frontal activation is more superior than in the model that does not consider typicality (compare Figure 5, 6 and 7). There were no significant voxels in the Artefacts > Natural kinds contrast (unless the p-voxelwise threshold was dropped dramatically p<.05).

Insert Figure 7 here

---

4 Discussion

In this fMRI study we explored concept typicality in a category verification task to further assess the conceptual semantic network for typicality and true category effects. We present results that are relevant to areas engaged in feature integration, semantic control, as well as category-related effects.

4.1 Typicality Effects

Negative Typicality Effects were primarily observed bilaterally in the IFG, and in supplementary motor cortex. Additionally, effects were observed in the left inferior anterior temporal lobe, albeit a bit posterior to that where typicality effects were previously observed using TMS (Woollams, 2012). The anterior temporal activation was not robust across all categories. The ATL is susceptible to magnetic

distortion (Lipschutz, Friston, Ashburner, Turner, & Price, 2001), however, there are reasons to believe this is not the case in the current study. The locations that typically gives rise to the greatest signal loss are those surrounding the sinuses (ie, medial ATL) and ear canal. Indeed, we found medial ATL activation in our category contrast. Thus, at most we might have some distortion in its localization, but signal should be sufficient to observe any effects. It would thereby seem the region is not equally sensitive to all categories. The category that had the smallest negative typicality effect was that of clothing. Thus, it seems that the ATL is increasingly engaged as concept typicality decreases and/or feature co-occurrence diminishes, though this may depend to some degree on the category considered, which poses a challenge to amodal accounts (Patterson et al., 2007).

Negative Typicality Effects in bilateral IFG is consistent with previous work that suggests that the role of this region is in semantic control (Novick, Trueswell, & Thompson-Schill, 2005). When typicality is low it requires more semantic control to perform a category judgment. The less typical the item the more taxing it would be to retrieve the item due to the low co-occurrence of the atypical features. Alternatively, within the context of the category verification paradigm, the prime may trigger a prediction for a target (ie, a prototype) and the more the target diverges from the prediction, the more the "atypical" features need to be suppressed in order to decide if the item belongs to the category (ie, by focusing on the features that are typical of that category).

Interestingly, the bilateral inferior parietal and posterior temporal regions were the main regions demonstrating *Positive Typicality Effects*. When looking for effects across all categories, activation was only observed in the right Inferior Parietal lobe. This finding is consistent with the region performing similarity categorizations, which has previously been reported (Grossman et al., 2002). The more typical the exemplar, the more similar it is to a potential prototype, used for categorization judgments, and hence the greater the activation observed. This does not provide evidence against feature integration also occurring here. The *Positive Typicality Effects* in the left hemisphere is consistent with the proposal that this region performs feature integration, if the region is additionally sensitive to the

relation between integrated features and a prototype, or the like, for categorization (ie, more typical). The effect in the left inferior parietal lobe was not reliable across all categories, but was strongest for Mammals. It seems difficult to align these results with a semantic control account as proposed by some authors (eg, Visser et al., 2012). If that were the case, then one would expect that as typicality decreases there would be greater activation, as was observed in the IFG (ie, a *Negative Typicality Effect*). Alternatively, the relation between the right inferior parietal lobe and left IFG could be that the degree of match between target and prototype, which is established in the parietal lobe is inversely proportional to an error signal sent to the IFG to suppress irrelevant features (ie, atypical ones).

Critically, the activation in the ATL is much reduced in size and strength compared to the other regions activated, such as the IFG and inferior parietal lobe. These results point to a limited role for the ATL in concept representation. That said, it may house critical pathways between the IFG and inferior parietal lobe that when disrupted (virtually or naturally) impairs performance preferentially on low typicality items because of the IFG's role either in retrieving the concept or in updating predictions. Further methods, such as MEG, that are better at anatomically localizing cortical activation to the ATL should be used.

4.2 Category-specific effects

We replicated previous findings for Natural Kinds > Artefacts as well as Artefacts > Natural Kinds. However, these areas did not show any significant activation when investigating category effects defined along their gradation in membership (i.e., typicality). Given the nature of categories, it would seem that any category specific area should be more sensitive to typicality of its category over typicality of another category (ie, demonstrate an interaction between category and typicality). This is distinct from an overall effect of typicality, which does not distinguish among categories. Although there was also an absence of overall typicality effects within regions sensitive to categories. Not surprisingly then, the category effect maps were unaffected when looking at those same effects with typicality regressed out. Most critically, however, as mentioned, when categories are defined by their typicality, they do not demonstrate differences in activation in these areas. This

suggests that those activations are not really category-specific. If they were, then they should have been modulated by the degree of typicality (i.e., the degree to which an item belongs to that category). This is moreover consistent with previous results showing that not all tasks demonstrate these category effects (i.e., living effect was not observed with perceptual tasks). A remaining explanation for these category differences is provided by Devlin et al. (2002): Activation differences may be due to differences in the recruitment of processes across the categories. Anteromedial temporal lobes show greater activity when more detailed conceptual information is required for the task (Tyler et al., 2004), with natural kinds requiring more fine-detailed information.

When investigating category effects as defined by typicality we did find a distributed activation for the [Natural Kinds > Artefacts] contrast and no activation for the reverse contrast. Interestingly, the [Natural Kinds > Artefacts] demonstrated activation in the left inferior parietal region similar in position to that observed for *Positive Typicality Effects.* Thus, while this region may be engaged in feature integration it also seems affected by the domain category.

4.23 Conclusions

Studying concept typicality in fMRI has provided a wide angle view on how the semantic system is organized and affected by typicality and categories. The results have implications for category specific deficits, and the role of semantic processing in the IFG, inferior parietal lobe and the pan-modal role of the ATL in concepts. While our findings present weak evidence for category specificity, they demonstrate that the IFG and inferior parietal lobe have distinct processing roles. While activation within the IFG is consistent with a role in semantic control (among similar competing theories), the positive typicality effect in the inferior parietal lobule does not seem compatible with such a role and we suggest a role in feature integration (across semantically similar features) for similarity based categorization.

Acknowledgments

This research was funded by Foundation for Science and Technology of Portugal Project Grant PTDC/PSI-PCO//118148/2010.

References

Barr, D. J., Levy, R., Scheepers, C., Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629-654.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2010). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*, 2767-2796.

Caramazza, A., & Shelton, J. R. (1998). Domain specific knowledge systems in the brain: the animate-inanimate distinction. *Journal of Cognitive Neuroscience, 10*, 1-34.

Casey, P. J. (1992). A reexamination of the roles of typicality and category dominance in verifying category membership. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 823-834.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage, 12*, 478-484.

Devlin, J. T., Rushworth, M. F. S., & Matthews, P. M. (2005). Category-related activation for written words in the posterior fusiform is task specific. *Neuropsychologia, 43*, 69-74.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Moss, H. E., Fadili, M. J., & Tyler, L. K. (2002). Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia, 40*, 54-75.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., Matthews, P. M., & Tyler, L. (2000). Susceptibility-induce loss of signal: comparing PET and fMRI on a semantic task. *NeuroImage, 11*(6), 589-600.

Dry, M. J., & Storms, G. (2010). Features of graded category structure: Generalizing the family resemblance and polymorphous concept models. *Acta Psychologica, 133*, 244-255.

Glass, A. L., & Meany, P. J. (1978). Evidence for two kinds of low-typical instances in categorization task. *Memory and Cognition, 6*, 622-628.

Grossman, M., Smith, E. E., Koenig, P., Glosser, G., DeVita, C., Moore, P., & McMillan, C. (2002). The neural basis for categorization in semantic memory. *NeuroImage, 17*, 1549-1561.

Henson, R.N.A., Penny, W.D. (2003). ANOVAs and SPM. Technical report, Wellcome Department of Imaging Neuroscience.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*, 441-461.

Holmes, S. J., & Ellis, A. W. (2006). Age of acquisition and typicality effects in three object processing tasks. *Visual Cognition, 13*, 884-910.

Kiran, W., Ntourou, K., & Eubank, M. (2007). The effectof typicality on online category verification of inanimate category exemplars in aphasia. *Aphasiology, 21*(9), 844-866.

Larochelle, S., & Pineau, H. (1994). Determinants of response times in the semantic verification task. *Journal of Memory and Language, 33*, 796-823.
Lipschutz, B., Friston, K. J., Ashburner, J., Turner, R., & Price, C. J. (2001). Assessing study-specific regional variations in fMRI signal. *NeuroImage, 13*, 392-398.
Mahon, B. Z., & Caramazza, A. (2008). Concepts and categories: A cognitive neuropsychological perspecive. *Annual Reviews of Psychology, 60*, 15.11-15.25.
Marques, J. F. (2007). The general/specific breakdown of semantic memory and the nature of superordinate knowledge: Insights from superordinate and basic level feature norms. *Cognitive Neuropsychology, 24*, 879-903.
Marques, J. F., & Raposo, A. (2011). Structural dimensions of object pictures: Organization and relation to object decision and naming. *Visual Cognition, 19*(6), 705-729.
Marques, J. F., Raposo, A., & Almeida, J. (2013). Structural processing and category-specific deficits. *Cortex, 49*, 266-275.
McCloskey, M. E. (1980). The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning and Verbal Behavior, 19*, 485-502.
Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms and word frequency. *Bulletin of the Psychonomic Society, 7*, 283-284.
Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 5*, 263-281.
Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*(12), 976-987.
Raposo, A., Mendes, M., & Marques, J. F. (2012). The hierarchical organization of semantic memory: executive function in the processing of superordinate concepts. *NeuroImage, 59*(2), 1870-1878.
Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology-General, 136*(3), 451-469.
Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology-General, 104*, 192-233.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structural of categories. *Cognitive Psychology, 7*, 573-605.
Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 491-502.
Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*(12), 2400-2406.
Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model of semantic association. *Psychological Review, 81*, 214-241.

Tyler, L. K., Stamatakis, E. A., Bright, P., Acres, K., Abdallah, S., Rodd, J. M., & Moss, H. (2004). Processing objecs at different levels of specificity. *Journal of Cognitive Neuroscience, 16*(3), 351-362.

Visser, M., Jefferies, E., Embleton, K. V., & Lambon Ralph, M. A. (2012). Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: Distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *Journal of Cognitive Neuroscience, 24*(8), 1766-1778.

Visser, M., Embleton, K. V., Jefferies, E., Parker, G. J., & Lambon Ralph, M. A. (2010). The inferior, anterior temporal lobes and semantic memory clarified: novel evidence from distortion-corrected fMRI. *Neuropsychologia, 48*(6), 1689-1696.

Warrington, E. K., & Shallice, T. (1984). Category-specific semantic impairments. *Brain, 107*, 829-853.

Woollams, A. M. (2012). Apples are not the only fruit: the effects of concept typicality on semantic representation in the anterior temporal lobe. *Frontiers in Human Neuroscience, 6*, 1-9.

Woollams, A. M., Cooper-Pye, E., Hodges, J. R., & Patterson, K. (2008). Anomia: A doubly typical signature of semantic dementia. *Neuropsychologia, 46*, 2503-2514.

Figures Captions

Figure 1. Mean Percent Correct +/- standard error of the mean (sem) for each true target category.

Figure 2. Mean Response Time +/- standard error of the mean (sem) for each target category.

Figure 3. A. Areas activated by a *Negative Typicality Effect* (increasing activation with decreasing typicality) p-voxelwise<.001, k=20m FDRc=48, overlaid on canonical rendered brain . B. Contrasts estimates of *Negative Typicality Effects for* each Category with 90% confidence intervals (C.I.) taken at the ATL peak coordinate (-38, -16, -30) (taken from the Flexible Factorial model). C. Areas demonstrating a *Negative Typicality effect* (increasing activation with decreasing typicality) p-voxelwise<.001, k=20, FDRc=48, masked with each category's *Negative Typicality Effect* (p-voxelwise<.05), overlaid on canonical rendered brain.

Figure 4. A. Areas activated by a *Positive Typicality Effect* (increasing activation with increasing typicality) p-voxelwise<.001, k=20, FDRc=40, overlaid on canonical rendered brain. B. Contrast estimates of *Positive Typicality Effects* of each category and 90% confidence intervals (C.I.) at left inferior parietal peak coordinate (-66, -24, 18) (taken from the flexible factorial model). C. Areas activated by a *Positive Typicality Effect* (increasing activation with increasing typicality) p-voxelwise<.001, k=20, FDRc=40, masked by each category's *Positive Typicality effect* (p-voxelwise<.05), overlaid on canonical rendered brain.

Figure 5. Areas demonstrating activation for the Natural kind > Artifacts contrast at p-voxelwise <.001, k=20m FDRc=61, overlaid on canonical rendered brain

Figure 6. Areas demonstrating activation by the Artifact > Natural Kinds contrast at p-voxelwise<.001, k=20, FDRc=95,overlaid on canonical rendered brain.

Figure 7. A. Areas activated by the contrast Natural Kinds > Artifacts, when those categories are defined by their typicality parameter. Maps are thresholded at voxelwise p<.001, k=20, FDRc=74 and overlaid on canonical rendered brain. B. Each category's typicality contrast estimate at inferior parietal peak coordinate (-58, -42, 34) in map above.