NeuroImage 50 (2010) 1243–1250

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

An Algebra for the Analysis of Object Encoding

Christopher W. Tyler *, Lora T. Likova

Smith-Kettlewell Eye Research Institute, USA

ARTICLE INFO

Article history:
Received 15 June 2009
Revised 29 September 2009
Accepted 8 October 2009
Available online 16 December 2009

ABSTRACT

The encoding of the objects from the world around us is one of the major topics of cognitive psychology, yet the principles of object coding in the human brain remain unresolved. Beyond referring to the particular features commonly associated with objects, our ability to categorize and discuss objects in detailed linguistic propositions implies that we have access to generic concepts of each object category with well-specified boundaries between them. Consideration of the nature of generic object concepts reveals that they must have the structure of a probabilistic list array specifying the Bayesian prior on all possible features that the object can possess, together with mutual covariance matrices among the features. Generic object concepts must also be largely context independent for propositions to have communicable meaning. Although, there is good evidence for local feature processing in the occipital lobe and specific responses for a few basic object categories in the posterior temporal lobe, the encoding of the generic object concepts remains obscure. We analyze the conceptual underpinnings of the study of object encoding, draw some necessary clarifications in relation to its modality-specific and amodal aspects, and propose an analytic algebra with specific reference to functional Magnetic Resonance Imaging approaches to the issue of how generic (amodal) object concepts are encoded in the human brain.

This paper concerns the neural representation of generic object concepts in the human brain and a novel approach to how to study them by means of functional Magnetic Resonance Imaging (fMRI). We first develop the logic of object concepts and their role in human cognitive processing, and then provide an algebraic approach to identifying their corresponding neural signals and dissociating them from domain-specific activation within particular sensory domains.

'Concept' is a philosophical term with a long history, meaning much the same as 'idea', or the abstraction of something in the world into its mental representation. We will take the liberty of assigning a neural basis to this representation. We use the term 'generic object concept' in distinction from the concept of a particular exemplar of an object. The neural representation of a concept is generalized from any specific representation of the object, such as one view of it under one set of lighting conditions, etc., or from any possible input modality (visual, tactile, auditory, etc.) providing the information about the object. In this sense, it is 'amodal' or abstract. Thus, representing particular objects has a high degree of generalization to it, as discussed at length by Plato and his philosophical descendants. However, our goal is to move up one level to the category of generic object concepts, as reflected in the common nouns of our language. The neural representation for basic nouns, such as 'apple' or 'chair', is the referent of the corresponding generic object concept. As implied by Rogers and McClelland (2008), the linguistic denotations must, almost necessarily, be mediated by an abstract or non-linguistic representation of the underlying concept, i.e., a representation of the ***meaning*** of the words beyond their lexical roles.

In principle, the same sort of logic could be developed for the processes underlying any aspect of linguistic encoding, but we are not attempting to do so here. The restriction in the development of this logic here is that it is limited to the object (or concrete noun) aspect of linguistic meaning. We are not attempting to address other aspects of linguistic representation, such as object manipulations (e.g., verbs) or interrelationships (e.g., logical propositions). Thus, we are proposing an approach that identifies the brain processes involved in the generic (amodal) object encoding that is the referent of concrete nouns.

Representational specificity

Arguably, a principal function of the brain is to optimize the ability of the organism to deal with the structures and contingencies of events in the world. A major issue in cognitive science, therefore, is the neural specificity of the cognitive representations of the structures in the world, which largely consist of self-contained objects, both rigid and flexible. The neural specificity of the representations of such structures has been discussed in terms of ***domain specificity***, the concept that certain cortical regions are specialized for representing particular categories of cognitive structures, such as 'faces' or 'houses' (Kanwisher et al., 1997; Haxby et al., 2001; Ishai et al., 1999), which are commonly termed 'objects'. This then became elaborated into the issue of how these cognitive categories are used in the neural

* Corresponding author.
E-mail address: cwt@ski.org (C.W. Tyler).

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.10.091

processing underlying the generation of propositions for object manipulation (Fodor, 2001; Martin, 2007; Pernet et al., 2007; Moss et al., 2005).

There is, however, a core difference between the concept of an ***object representation***, which plays a role something like that of a noun in linguistic analysis, identifying a stable structure of the world, and a cognitive ***processing domain***, which has more the role of a syntactic structure or sentence form in which operations are performed on noun objects. Proponents of domain specificity tend to define and explore brain organization in terms of operations on selected object categories, such as the recall of objects from previously seen categories or the task of categorization into subordinate or superordinate categories (Rosch, 1975; Mervis and Rosch, 1981). By employing such tasks, these investigations confound the operational ***processes*** of object recognition with the (lower-level) cognitive ***representations*** on which the operations are performed. The results of such task analyses have to be critically examined in terms of the level of operation that is being analyzed in such experiments. For example, brain regions identified as constituting working memory are explicitly understood as not forming the representational or semantic memory of objects as a whole, but a temporary transactional 'stock exchange' in which tokens for the stored objects are involved in some process of sentence generation. Any brain area that is activated during such a processing task but not during a non-object-relevant task involving the same objects must, *a fortiori*, be excluded from consideration as part of the abstract concept of the object *per se.*

Conversely, studies in which objects are presented with no category related task (e.g., with only a one-back task as to whether the previous image is repeated, whatever it may have been) avoid activating specialized processing domains, and focus on the specificity of the cortical activation corresponding to the object ***representation*** *per se*. Generally speaking, the range of neurological and neuroimaging techniques has generated a picture of cortical processing in which cognitive ***representations*** are located in the posterior portions of the brain while the operational ***processing*** of these representations takes place in the more anterior regions. This posterior/anterior distinction is analogous to the representational/operational distinction that we are making between the philosophical concepts of category specificity and domain specificity. The function of a chair is for people to sit on. To study the brain representation of that function (involving chairs), one needs to look at the task interaction between chairs as objects and sitting as the function of those objects. But the structure of a chair is something with a horizontal seat, supporting legs and a back, and potentially, armrests. We can sit on many things that are not chairs (such as stools, couches, divans, ottomans, benches, loveseats, pews, prie-dieux, tree-stumps, ledges, rocks, etc.) but there is a structural essence to a chair that does not encompass all these other sitting devices. We would argue that the functional specificity of 'a sitting device' is thus an ***insufficient*** definition of a chair, and moreover that there is no operational task that restricts the definition to what we call a 'chair' (other than instructions how to build something of this structural description). Hence, the concept of a chair is, in fact, ***structural*** rather than functional and does not require an operational task to be included in its categorical definition. In particular, the concept of an object is not the word for the object, it is not the color of the object, the visual shape of the object, the sound of the word for the object, the tactile shape of the object, or the function of the object; it does not depend on which input modality is providing the information about the object. The concept of the object is an ***amodal structural representation*** that transcends all these particulars.

This structural definition is somewhat contentious, as many have argued that concepts such as that of a 'chair' also have a functional component. While this may be the case for many objects of human construction, it is not the case for natural objects such as rocks or insects. In fact, our argument is that the functional role is superordinate to the structure, as in the concept of a 'vehicle'. A 'vehicle' has the functional definition of being a device for moving people or objects over the earth's surface, but it is specified at an abstract level that has few defined structural features, whereas exemplars such as a 'car' have wheels, an engine, an enclosing body, etc. Thus, the role of the functional definition applies at a more abstract level than the object concept to which we are referring.

Object representation

The issue of object representation, or the neural basis for the understanding of what an object is, is most commonly conceived as the concatenation of the multimodal properties that make up the object, in the form of the 'feature list' or the concatenated 'jet' of the object properties (Von der Marlsburg, 1973; Mervis and Rosch, 1981; Fodor, 1998; Pylyshyn, 2001; Murphy, 2004; Kapoor et al., 2009). For the ***generic object concept***, the feature list should not be regarded as absolute, but as ***probabilistic***, in that each feature is tagged with a probability of association with the specified object. In addition to its list of expected features, the generic object concept must incorporate estimates of the expected range of values for each feature in the list. In this sense, the expected range constitutes a probabilistic Bayesian prior on the possible values of features in this list (and even of the covariance matrices between items in the list), forming an elaborated structure that may be termed a 'list array'. Thus, for example, the prototype chair has four legs, but modern designs can have three, two or even a single preformed leg. However, if it has no legs (i.e., it is attached to some other structure such as a wall) it is no longer a 'chair' but becomes a 'seat'. Thus, there is a probability distribution on the feature of 'number of legs' that is not restricted to four, but has an inherent limit that can be expressed in its probability distribution.

This list array specification of the generic concept is well-conceived as the 'objects' that form the basis of 'object-oriented' computer language, in the sense that the feature list in the object concept is actually a set of parameters whose values remain to be specified. Thus, the list array is the ***primary*** specification of an object concept, while its role in the hierarchy of names and labels reflecting the relations among object concepts is a ***secondary*** stage of its specification. This hierarchical role, which is the relational aspect of object categorization, is derivative from its probabilistic list array, and thus, from the object concept. Although the idea of a probabilistic representation of the local features in visual processing is now well-recognized (e.g., Kersten et al., 2004), we are not aware of the specific proposal that the referent for ***object concepts*** (as amodally activated by the words for those objects) incorporates a Bayesian prior on all possibilities of each feature.

Moreover, there is an implicit understanding that, however strong the evidence is that some properties always co-occur, there is always some possibility that the characteristic concatenation can be violated. Such violations are the stuff of the carnival sideshow, precisely because the unexpected can occasionally occur. They can be built into the prior as a nonzero probability for all possible items in the list array. Even a canary could unexpectedly happen to be blue by some novel genetic mutation. Nothing in nature is absolutely prohibited, and Fodor's message is that this fact must be built in to our ability to conceptualize objects.

The nature of the concatenation process, however, is itself far from straightforward, and no convincing neural mechanism of combination of elementary features into the mid-levels of object structures has been found, even for such simple objects as letters and words (Tyler and Likova, 2007). Thus, the human ability to represent complex, three-dimensional objects (such as lawnmowers) remains unexplained. The probabilistic list array must also include some mechanism for encoding the three-dimensional configuration of the concatenated features, not just their probabilities of occurrence.

It is an important requirement that the ***core*** representations of the object concepts are largely context independent, making up the basic units of operation for later stages of cognitive processing such as their use in compound phrases and elaborated propositions, (cf. Fodor, 2001). Conversely, there is evidence (Freedman, 2001, 2006; Peelen, 2009) that the cortical activation patterns corresponding to particular objects are context dependent. These results do not, however, imply that the core object concepts are themselves context dependent. The argument may be expressed by analogy with a well-known form of context-dependence — color adaptation. If we view a patch of yellow color in the context of a green background, the patch may take on an orange hue. Nevertheless, this shift does not mean that our *concept* of yellow is exhibiting context dependence. We know that the appearance of the patch no longer matches our concept of yellow, but now matches our concept of orange. Thus, the percept of the same physical object is context dependent, but the abstract concept of 'yellow' against which the color is judged has not been affected by the context. In the same way, abstract concepts of objects need to maintain a high degree of context independence for propositions to have any communicable meaning, since the role of propositions is precisely to place objects (nouns) in an endless variety of contexts. (This is not to say that examples of context-dependence in core object concepts could not be found, but that their main character must exhibit context independence for widespread communication to be meaningful.)

Objects as species

To clarify the essence of the object concept, it may help to consider it in relation to the notion of biological species. Despite the best efforts of geneticists, the operational definition of most species is morphological — a set of organisms that all look approximately alike and can all interbreed.[1] Of course there are minor differences among members of the species, sometimes even quite substantial ones. For example, all humans are one species, and all dogs are one species, despite substantial differences in the dimensions of many of their features. We can test particular examples to verify the hypothesis that the morphological variants can interbreed, or to determine their genetic make-up, but in 99.99% of the cases the way we know that a dog is not, say, a baboon is based on our observation of its morphological features. These are statistical regularities that are based on centuries of cultural observation, and form a list array that is typically well-separated from that for any other species (in the *n*-dimensional space of possible biological features.) Even if it is born an albino, or loses one leg, or forgets how to bark, we still know that it is a dog from the concatenation of its other features.

Thus, the concept of an 'object' has the same 'quality' as that of a species. Just as species come in many variants, an object is something that comes in many versions, but all are identifiable as falling under the same descriptive feature list. There are even disjunctive species, in which the male and female have different morphologies. Indeed, most species go through different morphological varieties during the life cycle, so the morphological criterion itself has many variants for any one species. Thinking of the object concept in this way gives flesh to the idea that any one concept has many variants, and is not stereotypically fixed in one immutable form. Each variant is, however, discriminable from those of other concepts (except in extreme cases), forming an isolatable cluster of feature values in *n*-dimensional feature space (much as the galaxies form isolatable clusters of stars in 3-dimensional astronomical space).

Cortical network for representation of object concepts

The current understanding of object processing and recognition in both monkeys and humans is represented by a hierarchical scheme beginning with local orientationally-tuned units in primary visual cortex (V1), progressing through successive integration stages to a level of view-specific representations in posterior inferotemporal (IT) cortex and view-invariant object representations in more anterior regions of IT (Felleman & Van Essen, 1991; Grill-Spector and Malach, 2004; Serre et al., 2007; Martin, 2007; Rogers & McClelland, 2008; Mahon & Caramazza, 2008). Rather than being the simple hierarchy that is often discussed, however, this organization takes the form of a serial/parallel ***heterarchy*** (that is, a network of parallel, interlocking hierarchies). With level of the heterarchy, the representations progress from local features through broader object groupings to the most abstract concepts of the object categories.

In accord with this heterarchical scheme, most studies of object processing show a distributed coding of the components of objects at a relatively low level of representation (e.g., the color, motion and size of objects; e.g., Grill-Spector et al., 1999; Haxby et al., 2001). Such activation can account for some aspects of object processing, or for the differential task demands of dealing with objects, e.g., the tasks of recognizing, naming, evaluating or comparing objects, which progress along the ventral stream of elaborated aspects of object processing (e.g., Small et al., 1995; Puce et al., 1996; Martin et al., 1996; Moore and Price, 1999; Lerner et al., 2001; Tyler et al., 2005, 2006; Taylor et al., 2006; Borowsky et al., 2007). The core issue is whether the view-invariant representations in posterior IT can be said to be the true stage of object representation *per se* (i.e., the amodal 'object concept'), as suggested by Moore and Price (1999), Joseph (2001) and Pietrini et al. (2004), as opposed to being a concatenation of image-based representations (Tarr and Buelthoff, 1998; Vuilleumier et al., 2002; O'Toole et al., 2004). As developed above, the object concept may be best understood as a probabilistic feature list, or list array instantiating the Bayesian priors for the structural and function features constituting each category of object. This list array representing the object concept is thus hierarchically 'above' the representations of the individual features, and may be expected to be in higher functional areas, anatomically anterior to the locations of their cortical representations.

The alternative is that the abstract concept representation is distributed through the dimensions of feature representation in a way that is inaccessible to the current cortical mapping techniques. In other words, is the mere existence of correlations within the multidimensional feature network a sufficient code for the functional operation of the organism in terms of objects, or do they need a specific, higher-stage representation that draws together the cluster of correlations for further processing as a defined object? This is a restatement of the so-called 'binding problem' of how the diverse features of an object are held together through the range of conceptual manipulations that objects may be mentally subject to (Treisman & Gelade, 1980; Singer & Gray, 1995; Pylyshyn, 2001; Thiele & Stoner; 2003; Bartels & Zeki, 2006; Dong et al., 2008). Is it sufficient to abstract the feature properties (including potential uses and relations to other objects) and spread these abstract properties throughout the brain (distributed processing), or is there a central 'coordinator' that maintains the relationships among these features in a coordinated data structure (i.e., binding the processed features) for each type of object?

The apparent need for a 'binding' mechanism implies that, at some point in this sequence of progressive abstraction from the specific exemplars, we should expect a representation of the generic concept of that object's identity that is not defined by any one of its features or

[1] Note that, although the biological conceptual definition of species is based on the criterion of interbreeding, the practical specification in most situations is morphological even for the assessment of interbreeding — we infer the capability of interbreeding based on observation of morphology and an empirical knowledge base of past interbreeding studies. For the paleontology of fossil species, in particular, the interbreeding criterion for species is entirely inferential from the morphology.

particular views. This mechanism corresponds to the role of the higher association cortices proposed by Meyer & Damasio (2009), in coordinating the activity of the lower, modality-specific representations. In this sense, the ***coordinating*** function, if it is indeed a necessary component of object encoding, corresponds to its abstract, or generic, aspect. Thus, most studies of different kinds of objects and object representation modalities (e.g., Small et al., 1995; Puce et al., 1996; Martin et al., 1996; Moore and Price, 1999; Lerner et al., 2001; Taylor et al., 2006; Shapiro, Moo & Caramazza, 2006; Borowsky et al., 2007) show a distributed coding of the cue components of objects at a relatively low level of representation (e.g., the color, motion and size of objects), or to the differential task demands of dealing with objects (e.g., the tasks of recognizing, naming, drawing or manipulating objects). Such activation can account for some aspects of object processing, but it is still at an early stage in the stream of true object processing. In particular, based on the established principles of organization of the occipital cortex, the generic object concept should be expected to have its representation anterior to the modality-specific aspects of that object.

The question of whether the generic representations of different object classes would be partially localized or would be fully distributed at a finer-grained scale over the same location remains undetermined. Based on previous evidence, the most likely candidate for the highest cortical object coding are the middle regions of the temporal lobe (e.g., Piefke et al., 2003; Zahn et al., 2007; Kriegeskorte et al., 2007; Canessa et al., 2008). It is to be expected that relatively weak activation would be found in the region coding the abstract concept of each object, due to the large variety of objects that need to be coded. For example, if one is standing in front of an elephant, the local features of elephantness may fill a large portion of the visual field, but the ***concept*** of an elephant is only one of many wild animals, which are only a small subset of many types of live organism, which are again a small subset of many types of objects in the universe at large, which itself is only a subset of the things that can be described linguistically by nouns, all of which must be represented somehow in the brain. Thus, the abstract concept of "elephant" should activate only a very small subset of the activations for all possible objects and noun-predicates in general. If the specific object concepts are represented locally in mid-IT, or the relevant cortical substrate for abstract object concepts, the BOLD activation should be expected to be found in only a small local region or network, and one that may well be idiosyncratic for each individual brain (since it is unlikely that there is genetic coding for the wide variety of objects that we encounter in everyday life). Thus, any analysis of the localized representation of object types needs to be conducted on an individual subject basis in order to be seen effectively (Postle and D'Esposito, 1999; Wheaton et al., 2004; van Atteveldt et al., 2004; O'Toole et al., 2004).

An important study that focuses on this issue comes from the Haxby lab (Pietrini et al., 2004), in which they identify regions that are jointly activated by the same visual and tactile objects in sighted subjects and by the same tactile objects in blind subjects, and argue that these co-activated regions constitute the cortical basis for the abstract representation of object concepts *per se*. The joint activation was located in the posterior IT, overlapping in the region strongly activated by visual objects in the sighted subjects.

To emphasize this point, the temporal lobe activations are reproduced from Pietrini et al. (2004) in Fig. 1. The temporal lobe pattern for tactile object exploration shows two main activation sites (Fig. 1A), one labeled IT in the posterior temporal lobe and the other more ventral. The visual activation (Fig. 1B) spreads through a large region of the posterior temporal lobe, the overlap between the two forms of activation being shown by the yellow coloration in Fig. 1C. This pattern of activation is consistent with the general framework of other studies from the Haxby group, which is that "object categories are represented in distributed and overlapping representations" (Haxby et al., 2001; see also Beauchamp et al., 2003; Kriegeskorte et al., 2007). While the activation may be distributed on a local level, it is clearly restricted to predominantly one region of the posterior temporal cortex, supporting the concept of large-scale specialization of cortical function.

Even more abstract properties of the world, such as of the organization of scenes into 'frames' of expected groups of objects and maps of the layout of the cognitive universe may be expected to be even more anterior in the temporal lobe. In terms of the cortical organization of object representation, the serial/parallel heterarchy of object representation thus appears to have some degree of location specificity, both in terms of (horizontal) category specificity and of the (vertical) level of feature-to-concept specificity. The logic of the analysis isolating the abstract component of the activity is applicable to any level of object categorization.

The philosophy of fMRI studies

In probing the encoding of object concepts by means of the fMRI methodology, we need to keep in mind its strengths and weaknesses. Functional MRI has both the blessing and the curse that it probes activation throughout the brain for any controlled activity that can be conducted inside the scanner. This global purview is an advantage because it reveals significant increases in neural activation (or, strictly, in the consequent hemodynamic changes) in any part of the brain that is involved in the task imposed by the experimental design. Its disadvantage is that fMRI activation reflects all (metabolically reactive) aspects of the neural processing, which may include many aspects that are not intended in the experimental design, clouding the interpretation of the activation pattern. For example, if the subject is viewing an object having the form of a Necker cube relative to a blank fixation field, differential activation in a particular brain region may reflect the local luminance difference, the presence of oriented lines, the connectivity of the global structure, the two-dimensional shape, the perceived three-dimensional structure, the occurrence of the alternations, the transparency of the implied surfaces, the artificiality of the line drawing, the implication that it is an empty box, and so on. Potentially, each of these properties is encoded in a separate region of

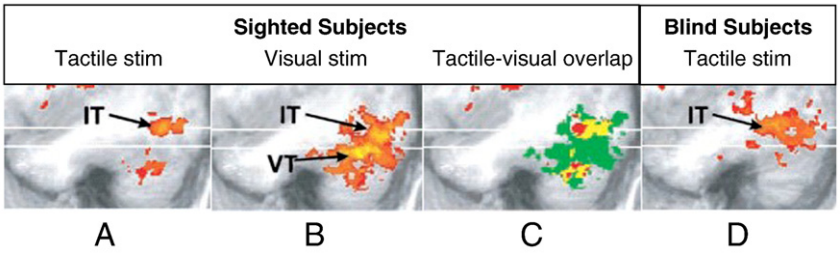

Fig. 1. Posterior inferotemporal (IT) activation by several categories of objects presented via the tactile (A) and visual (B) modalities to sighted subjects (with their overlap shown in C) and via the tactile modality in blind subjects (D). Note the remarkable consistency between the localization of the tactile activation in the sighted (A) and blind (D) subject groups. (Reconfigured from Pietrini et al., 2004.)

the cortex, resulting in the independent activation of large numbers of subregions of the overall network response.

The usual approach to this issue is to report all the regions activated and then provide a speculative assessment of what function each region is serving in the global task that was taking place. Such speculations should be recognized for what they are, since they are not directly tested in the experiment in which the activation is compared with observation of a blank field. This caution is even more strongly applicable when a task is included during the test periods, because the task tends to recruit more cortical functions (such as motor planning of the choice of which button to press) beyond those implied by the original observation state.

The specialized functions of particular brain areas thus cannot be unambiguously determined in any comparison in which large numbers of brain functions covary between the test and null intervals (or the test presentation and the preceding baseline level in event-related designs). The combinatorics of determining the functions of each activated area from multiple comparisons among different forms of such tests are similarly intractable. The only case in which an unambiguous interpretation can be given to a particular cortical activation is when it is derived from the comparison between two active conditions with the particular relationship that Condition A has one extra property that Condition B (or *vice versa*). Any less-controlled subtraction is just as difficult to interpret as an isolated condition, but the $(n+1)/n$ comparison makes an almost ironclad interpretation that the extra property must be what gave rise to the differential activity. It is often difficult to achieve an exact single-property increment (if, for example, the additional property gives rise to increased attention, or extra high-level processing), but the more restricted the single-property increment is, the more certain is the interpretation of the pattern of activation.

The single-property increment paradigm for the identification of differential aspects of object encoding is equally applicable to the analysis of more elaborated cognitive processes such as proposition generation. Again, the requirement is to compare two conditions in an $(n+1)/n$ paradigm, i.e., having just one additional aspect in one condition vs the other. This requirement is similar to that for nested designs in the statistical evaluation of model fitting, in which models can be compared through the addition of one parameter at a time to a baseline model. An example would be the generation of simple sentences with no adjectives (n) compared with the generation of sentence containing an adjective ($n+1$). The contrast of these two conditions would thus probe the specific process of ***adjectival modification of a noun*** in the overall context of full sentence generation, while controlling all other aspects of the sentence generation process.

One important aspect of the design for isolating the abstract object concept is to control for as many low-level cues as possible, since low-level cues will activate large extents of the early sensory cortex. It is well established that fMRI contrasts with appropriate null stimuli can, indeed eliminate much of the activation generated by object stimuli vs a blank field. Since the null stimuli do not contain perceptible object cues, the activity thus eliminated must relate to ancillary cues rather than the object-specific information. Balancing the activation for such ancillary cues relieves the paradigm of the burden of controlling it by other means.

In order to focus the activation on object properties *per se*, the null event types should not be an empty screen with a fixation point, but Fourier-matched, spatial-location matched noise images (such as those developed for a face-coding study by Chen et al., 2006). Two examples of such null-object images are shown in Fig. 2. They are generated by either randomizing the phase matrix (Fig. 2B) or setting imaginary values to zero in the Fourier plane (Fig. 2C) before transforming the 2D Fourier spectrum back to the image plane. These manipulations generate either random or symmetric-random null images, respectively, which are then windowed by a fourth-power Gaussian envelope matching the spatial extent of the original object image. Note that this method of symmetrizing the image (Fig. 2C) generates contour structure reminiscent of that in the original object images, in addition to the symmetry structure. This approach thus controls for two important mid-level cues – contour structure and symmetry – in addition to the low-level cues of local Fourier energy.

Specific hypotheses for the cortical representation of generic object concepts

Thus, the logic for the analysis of the cortical representation of generic object concepts is based on four straightforward hypotheses that are well-validated in the literature reviewed above.

1. That the object-specific property of object images is revealed by contrasting the object images with appropriately scrambled versions of the same images (which both activate the retinotopic occipital cortex). This contrast is known to give object-specific activation in lateral occipital and posterio-ventral temporal cortex, together with weak activation in more anterior sites.
2. That the semantic content of auditory words is revealed by contrasting the auditory presentation of auditory words versus auditory nonsense words (which themselves activate the auditory cortex in the superior temporal gyrus). This contrast is known to give semantic-specific activation in and near the posterior and anterior STS.
3. That the object-specific property of visually-presented words is revealed by contrasting the visual object words with visual nonsense words (which themselves activate the retinotopic occipital cortex and the ventral temporal visual word form area). This contrast is known to give object-specific activation in anterior inferotemporal and posterior frontal cortex.
4. The above three hypotheses lead to the contingent hypothesis that there are abstract object concepts or amodal semantic representations separate from the specific visual-modality and verbal-modality representations identified in (1–3).

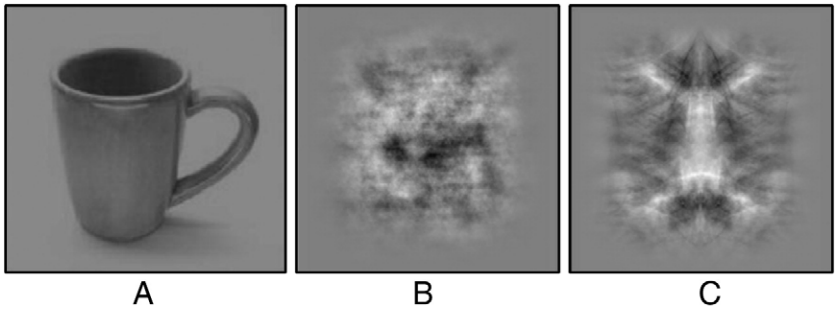

Fig. 2. Typical object (A), with random (B) and symmetric-random (C) forms of spatially-windowed null image matching the low-level properties of an object image. (Reconfigured from Chen et al., 2006)

An algebra of object processing

Given these assumptions, we may define an algebra for the analysis of the components of the object concept, under the usual assumptions of the General Linear Model (i.e., that the overall effect is the linear sum of the individual effects, if they were measured separately) and the block design paradigm (i.e., that the temporal response waveform for all events is specified by the convolution of the design matrix with the metabolic response function). Specifically, we are assuming that any interactions, such as intersensory inhibition, operate to the same extent under separate or combined stimulation conditions. We emphasize that all quantities specified in this algebra are object-processing activations $\boldsymbol{\beta}$, equated for the properties of the sensory input modality. Each $\boldsymbol{\beta}$ represents a matrix of time series across voxels, and collapses to a vector of response amplitudes if the time series are identical for the contrasted stimulus types.

Formally, the visual object (V) processing network is defined as difference between the activation produced by visual ***object*** stimuli ($\boldsymbol{\beta}_V$) and that produced by scrambled images of the same objects ($\boldsymbol{\beta}_{Vn}$):

$$V = \beta_V - \beta_{Vn} \tag{1}$$

Similarly, for auditory word (A) presentation, the semantic representation is obtained by subtraction of nonsense auditory words:

$$A = \beta_A - \beta_{An} \tag{2}$$

and for visual-word (W) presentation, the semantic representation is obtained by subtraction of nonsense visual words:

$$W = \beta_W - \beta_{Wn} \tag{3}$$

Single modality analysis

However, we need to avoid the confound that activation through each modality has been shown to generate weak percepts in the other modalities, known as imagery. We first develop the logic for avoiding this confound for the case of a single presentation modality. Visual imagery of objects is assumed to activate the same regions as the pictorial stimulation with objects (e.g., Thompson et al., 2001; Kosslyn and Thompson, 2003; Ganis et al., 2004; Slotnick et al., 2005), but at a reduced strength compatible with estimates of the perceived contrast of the visual imagery (Kosslyn, 1994). Similarly, the visual image of the object may be expected to be evoked when reading object words or hearing their names spoken, and the faint auditory image of a spoken name may be expected to be evoked when seeing objects or reading their visual names.

Under the assumption of linear independence, the differential response for any stimulation modality is the sum of the activations in its own modality representation and the modality-independent activation for the object concept *per se*, together with the imagery-level of activations in object coding areas for each other modality. Thus, the differential response to visual stimulation for each object category is the sum of the activations in visual object areas (***V***), the putative object concept *per se* (***C***), and the imagery-level (***im***) of activations in other object coding areas (object word forms, object names, etc., designated as ***im*Other***).

$$R_V = V + im{*}Other + C \tag{4}$$

The specifically visual component of this response complex can be identified by lowering the contrast of the visual stimulation to a level that matches the subjects' estimates of the perceived contrast of imagined visual objects ($im{*}\mathbf{V}$). This contrast reduction should not affect the activation levels in the other modalities, since the visual-word form, auditory name and concept of an object do not change when its contrast is reduced. For example, the differential response $\boldsymbol{R}_{V\_low}$ for the low contrast visual stimulus matching the imagined image contrast would be given by modifying Eq. (4):

$$R_{V\_low} = im{*}V + im{*}Other + C \tag{5}$$

The contrast manipulation thus allows us to differentiate the visual activation from these ancillary components of the object response network. Basic algebra derives the specifically ***visual*** component of the object processing (***V***) from the difference between Eqs. (4) and (5):

$$V = \left(R_V - R_{V\_low}\right) / (1 - im) \tag{6}$$

and the total ancillary activation as

$$C = R_{V\_low} - im{*}\left(R_V - R_{V\_low}\right) / (1 - im) - im{*}Other \tag{7}$$

This relation derives the activation pattern for the generic object concept ***C*** in terms of the two levels of visual response, but shows that it is still confounded by an imagery component from the other modalities represented by the ***Other*** activation patterns. (Note that these relations assume that the experiment is designed to equate the task and attentional components of object processing between the high and low contrast conditions.)

Triple modality analysis

We have thus made the first step toward the goal of identifying the pure object concept, but we need a technique to eliminate the imagery-level activations of the other sensory modalities. To identify the activation pattern corresponding to the ancillary or imagery-level activation of associated sensory modalities, such as the auditory name (***A***) and visual word-form (***W***) activations from a visual presentation of the object, we need to provide manipulations for those activation modalities by presenting a) the sound and b) the word form for the same objects alone. These will activate their respective modalities at full strength, but will also activate the visual object modality at imagery-level strength, which we have designated as *im* in eq. 4. Including these two ancillary modalities, therefore, for the two visual object conditions we have:

$$R_V = V + im{*}A + im{*}W + C \tag{8}$$

$$R_{V\_low} = im{*}V + im{*}A + im{*}W + C \tag{9}$$

Then for auditory stimulation (relative to nonsense word sounds), we should fully activate the auditory semantic regions, have reduced (imagery-level) activation in the visual object regions and the visual word-form region, and have full activation in the hypothesized object concept region.

$$R_A = A + im{*}V + im{*}W + C \tag{10}$$

Finally, for visual word stimulation (relative to nonsense words), we should fully activate the visual word-form region, have reduced (imagery-level) activation in the auditory semantic regions and the visual object regions, and again have full activation in the hypothesized object concept region.

$$R_W = W + im{*}V + im{*}A + C \tag{11}$$

From these four equations for the full three-modality model (eq 8-11), we can add the index j to derive the algebraic expression for the

activation corresponding to the abstract object concept C_j for each object category *j*.

$$C_j = \frac{im * \left(R_{V,j} + R_{A,j} + R_{W,j}\right) - (2 * im + 1) * R_{V\_low}}{(im - 1)} \quad (12)$$

Thus, an experimental paradigm based on the algebraic approach is able to provide objective measures for all the terms needed to isolate the representation of the abstract object concept.

An issue in the practical application of the proposed algebra for isolating the abstract object representations is its sensitivity to the assumptions of linearity and of the stability of the imagery factor across modalities and experimental conditions. These equations are proposed in the context of the fact that fMRI has a relatively low signal/noise ratio rarely exceeding a factor of 10. In this regard, we note that the main subtractions are the modality-specific contrasts of eqs 1-3, while any differences among the imagery factors across modalities are likely to be small, given that imagery is typically weaker than direct sensory stimulation. The resulting confidence interval on significant differences in fMRI signals thus provides a cushion against minor violations of the assumptions of linearity and of the stability of the constants in the equations. The value of the imagery parameter can be specified on the basis of psychophysical studies, such as Kosslyn, 1994, and thus based on either the average published value or based on individual measurements in the subjects of the fMRI study.

At the risk of redundancy, we describe a specific experiment that would take advantage of this algebra. Consider the abstract concept of the category 'clock'. In Damasio's terms, this would be the organizational concatenator that ties together the individual features of a clock, such as the rotary dial, the hands, the Arabic or Roman numerals, the setting knobs, the (optional) alarm bell(s), and so on. Each of these items evokes a visual image, a text–image, the sound of its name and other sounds, such as the sound of the ticking and the sound of the alarm, together with a sense of the movement, of the hands, the way it defines time, the urgency of the alarm and need to switch it off, and the configuration of its parts. The abstract concept of the clock is the neural representation that ties all these elements together into a coherent object. The algebra is designed to isolate this abstract level of the cortical activation from the specific sensorimotor components.

It should be recognized that, although in its current form this algebraic approach controls for contamination by the modality-specific activation of the major sensory input modalities from the activation pattern for the abstract concept of the object category, it does not exclude ancillary activation, such as possible motor component, or embodied cognition (in visual object studies Martin et al., 1996; Chao et al., 2002; Beauchamp and Martin, 2007 and Weisberg et al., 2007, among others, and in linguistic studies by Grossman et al., 2002; Hauk et al., 2004; Tettamanti et al., 2005; Kemmerer et al., 2008, and Raposo et al., 2009, among others). The question then would be, which parts of the cortical activation isolated form the 'embodied cognition' aspect? Given the well-known specialization of the brain for motor representation in the motor and premotor cortices of the frontal lobes, together with the cerebellum, it would be reasonable to assume that activation of these motor-specific cortical regions would form the embodied cognition component of an abstract concept. Our novel method provides one step ahead from the current fMRI methodology of studying the generic object concept domain, but further development would be needed to disentangle all possible aspects.

Conclusion

The introductory part of this paper discusses the nature of objects and object coding in the brain, with particular emphasis on the difference between the coding structure *per se* and in the use to which it is put in information retrieval and the propositional logic of sentence construction. We propose a Bayesian probabilistic view of object concepts, both particular and generic, that are the referents of the nouns we use to communicate with each other about objects. The success of this communication ability implies that the referents have a tangible basis that can be sought in the human brain by functional imaging methods, when corrected for the intrasensory and cross-sensory aspects of the patterns of activation by means of an appropriate algebraic manipulation.

The algebraic treatment of the intrasensory and cross-sensory activations from this study provides the ability to determine how the representations of abstract concepts are organized within the cortical manifold. The variance of the estimate may be derived as the sum of the variances of the four parameters in Eq. 12 measured by fMRI. The imagery terms can be specified by means of psychophysical studies, such as Kosslyn, 1994, based on either the average published value or individual measurements in the subjects of the fMRI study. The result will be a statistical parametric map of the cortical representations of the modality-independent, or abstract, object concept for any object category, in order to evaluate the hypothesis that these representations are localized at the mid-to-anterior temporal-lobe level of the object processing heterarchy. This novel approach to the identification of the generic component of the overall object representation can thus provide a significant advance in our understanding of the organization of object encoding, both in terms of it neural organization and the philosophical analysis of the nature of human thought.

Acknowledgment

Thanks to Zlatko Minev for help with the formulation of the equations.

References

Bartels, A., Zeki, S., 2006. The temporal order of binding visual attributes. Vision Research 46, 2280–2286.
Beauchamp, M.S., Martin, A., 2007. Grounding object concepts in perception and action: evidence from fMRI studies of tools. Cortex 43, 461–468.
Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A., 2003. FMRI responses to video and point-light displays of moving humans and manipulable objects. J. Cogn. Neurosci. 15, 991–1001.
Borowsky, R., Esopenko, C., Cummine, J., Sarty, G.E., 2007. Neural representations of visual words and objects: a functional MRI study on the modularity of reading and object processing. Brain Topogr. 20, 89–96.
Canessa, N., Borgo, F., Cappa, S.F., Perani, D., Falini, A., Buccino, G., Tettamanti, M., Shallice, T., 2008. The different neural correlates of action and functional knowledge in semantic memory: an FMRI study. Cereb. Cortex 18, 740–751.
Chao, L.L., Weisberg, J., Martin, A., 2002. Experience-dependent modulation of category-related cortical activity. Cereb. Cortex 12, 545–551.
Chen, C.C., Kao, C., Tyler, C.W., 2006. Face configuration processing in the human brain: the role of symmetry. Cereb. Cortex 7, 1423–1432.
Dong, Y., Mihalas, S., Qiu, F., von der Heydt, R., Niebur, E., 2008. Synchrony and the binding problem in macaque visual cortex. J. Vision 8, 1–16.
Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1, 1–47.
Fodor, J.A., 1998. Concepts: Where Cognitive Science Went Wrong. Oxford University Press, Oxford.
Fodor, J.A., 2001. The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology. MIT Press, Cambridge, MA.
Ganis, G., Thompson, W.L., Kosslyn, S.M., 2004. Brain areas underlying visual mental imagery and visual perception: an fMRI study. Brain Res. Cogn. Brain Res. 20, 226–241.
Grill-Spector, K., Malach, R., 2004. The human visual cortex. Annu. Rev. Neurosci. 27, 649–677.
Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., Malach, R., 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 24, 187–203.
Grossman, M., Koenig, P., DeVita, C., Glosser, G., Alsop, D., Detre, J., Gee, J., 2002. Neural representation of verb meaning: an fMRI study. Hum. Brain Mapp. 15, 124–134.
Hauk, O., Johnsrude, I., Pulvermüller, F., 2004. Somatotopic representation of action words in human motor and premotor cortex. Neuron 41, 301–307.
Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430.

Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V., 1999. Distributed representation of objects in the human ventral visual pathway. Proc. Natl. Acad. Sci. U. S. A. 96, 9379–9384.
Joseph, J.E., 2001. Functional neuroimaging studies of category specificity in object recognition: a critical review and meta-analysis. Cogn. Affect. Behav. Neurosci. 1, 119–136.
Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.
Kapoor, A., Urtasun, R., Darrell, T., 2009. Probabilistic kernel combination for hierarchical object categorization. EECS Tech. Rep. 2009–2016.
Kemmerer, D., Castillo, J.G., Talavage, T., Patterson, S., Wiley, C., 2008. Neuroanatomical distribution of five semantic components of verbs: evidence from fMRI. Brain Lang. 107, 16–43.
Kersten, D., Mamassian, P., Yuille, A., 2004. Object perception as Bayesian inference. Annu. Rev. Psychol. 55, 271–304.
Kosslyn, S.M., 1994. Image and Brain: The Resolution of the Imagery Debate. MIT Press, Cambridge. MA.
Kosslyn, S.M., Thompson, W.L., 2003. When is early visual cortex activated during visual mental imagery? Psychol. Bull. 129, 723–746.
Kriegeskorte, N., Formisano, E., Sorger, B., Goebel, R., 2007. Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc. Natl. Acad. Sci. U. S. A. 104, 20600–20605.
Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., Malach, R., 2001. A hierarchical axis of object processing stages in the human visual cortex. Cereb. Cortex 11, 287–297.
Mahon, B.Z., Caramazza, A., 2008. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. J. Physiol. Paris 11, 585–597.
Martin, A., 2007. The representation of object concepts in the brain. Annu. Rev. Psychol. 58, 25–45.
Martin, A., Wiggs, C.L., Ungerleider, L.G., Haxby, J.V., 1996. Neural correlates of category-specific knowledge. Nature 379, 649–652.
Mervis, C.B., Rosch, E., 1981. Categorization of natural objects. Annu. Rev. Psychol. 32, 89–113.
Meyer, K., Damasio, A., 2009. Convergence and divergence in a neural architecture for recognition and memory. Trends Neurosci. 32, 376–382.
Moore, C.J., Price, C.J., 1999. Three distinct ventral occipitotemporal regions for reading and object naming. NeuroImage 10, 181–192.
Moss, H.E., Rodd, J.M., Stamatakis, E.A., Bright, P., Tyler, L.K., 2005. Anteromedial temporal cortex supports fine-grained differentiation among objects. Cereb. Cortex 15, 616–627.
Murphy, G.L., 2004. The Big Book of Concepts. MIT Press, Cambridge, MA.
O'Toole, A.J., Jiang, F., Abdi, H., Haxby, J.V., 2004. Partially distributed representations of objects and faces in ventral temporal cortex. J. Cogn. Neurosci. 17, 580–590.
Pernet, C., Schyns, P.G., Demonet, J.F., 2007. Comments and controversies. Specific, selective or preferential: comments on category specificity in neuroimaging. NeuroImage 35, 991–997.
Piefke, M., Weiss, P.H., Zilles, K., Markowitsch, H.J., Fink, G.R., 2003. Differential remoteness and emotional tone modulate the neural correlates of autobiographical memory. Brain 126, 650–656.
Pietrini, P., Furey, M.L., Ricciardi, E., Gobbini, M.I., Wu, W.H., Cohen, L., Guazzelli, M., Haxby, J.V., 2004. Beyond sensory images: object-based representation in the human ventral pathway. Proc. Natl. Acad. Sci. 101, 5658–5663.
Postle, B.R., D'Esposito, M., 1999. "What"-Then-Where" in visual working memory: an event-related fMRI study. J. Cogn. Neurosci. 11, 585–597.
Puce, A., Allison, T., Asgari, M., Gore, J.C., McCarthy, G., 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. J. Neurosci. 16, 5205–5215.
Pylyshyn, Z.W., 2001. Visual indexes, preconceptual objects, and situated vision. Cognition 8, 127–158.
Raposo, A., Moss, H.E., Stamatakis, E.A., Tyler, L.K., 2009. Modulation of motor and premotor cortices by actions, action words and action sentences. Neuropsychologia 47, 388–396.
Rogers, T.T., McClelland, J.L., 2008. Precis of semantic cognition: A parallel distributed processing approach. Behav. Brain Sci. 31, 689–749.
Rosch, E., 1975. Cognitive representation of semantic categories. J. Exp. Psychology 104, 192–233.
Serre, T., Oliva, A., Poggio, T., 2007. A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. U. S. A. 104, 6424–6429.
Shapiro, K.A., Moo, L.R., Caramazza, A., 2006. Cortical signatures of noun and verb production. Proc. Nat. Acad. Sci. 103, 1644–1649.
Singer, W., Gray, C.M., 1995. Visual feature integration and the temporal correlation hypothesis. Annu. Rev. Neurosci. 18, 555–586.
Slotnick, S.D., Thompson, W.L., Kosslyn, S.M., 2005. Visual mental imagery induces retinotopically organized activation of early visual areas. Cereb. Cortex 15, 1570–1583.
Small, S.L., Hart, J., Nguyen, T., Gordon, B., 1995. Distributed representations of semantic knowledge in the brain. Brain 118, 441–453.
Tarr, M.J., Buelthoff, H.H., 1998. Image-based object recognition in man, monkey and machine. Cognition 67, 1–20.
Taylor, K.I., Moss, H.E., Stamatakis, E.A., Tyler, L.K., 2006. Binding crossmodal object features in perirhinal cortex. Proc. Natl. Acad. Sci. U. S. A. 103, 8239–8244.
Tettamanti, M., Buccino, G., Saccuman, M.C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S.F., Perani, D., 2005. Listening to action-related sentences activates fronto-parietal motor circuits. J. Cogn. Neurosci. 17, 273–281.
Thiele, A., Stoner, G., 2003. Neuronal synchrony does not correlate with motion coherence in cortical area MT. Nature 421, 366–370.
Thompson, W.L., Kosslyn, S.M., Sukel, K.E., Alpert, N.M., 2001. Mental imagery of high- and low-resolution gratings activates area 17. NeuroImage 14, 454–464.
Treisman, A., Gelade, G., 1980. A feature-integration theory of attention. Cognitive Psychology 12, 97–136.
Tyler, C.W., Likova, L.T., 2007. Crowding: a neuro-analytic approach. J. Vision 7, 1–9.
Tyler, C.W., Baseler, H.A., Kontsevich, L.L., Likova, L.T., Wade, A.R., Wandell, B.A., 2005. Predominantly extra-retinotopic cortical response to pattern symmetry. Neuro-Image 24, 306–314.
Tyler, C.W., Likova, L.T., Kontsevich, L.L., Wade, A.R., 2006. The specificity of cortical area KO to depth structure. NeuroImage 30, 228–238.
van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. Neuron 43, 271–282.
Von der Marlsburg, C., 1973. Self-organization of orientation-sensitive cells in the striate cortex. Kybernetik 14, 85–100.
Vuilleumier, P., Henson, R.N., Driver, J., Dolan, R.J., 2002. Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. Nat. Neurosci. 5, 491–499.
Weisberg, J., van Turennout, M., Martin, A., 2007. A Neural System for Learning about Object Function. Cerebral Cortex 17, 513–521.
Wheaton, K.J., Thompson, J.C., Syngeniotis, A., Abbott, D.F., Puce, A., 2004. Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. NeuroImage 22, 277–288.
Zahn, R., Moll, J., Krueger, F., Huey, E.D., Garrido, G., Grafman, J., 2007. Social concepts are represented in the superior anterior temporal cortex. Proc. Natl. Acad. Sci. U. S. A. 104, 6430–6435.