

Similarity representation of pattern-information fMRI

XUE ShaoWei¹, WENG XuChu¹, HE Sheng^{1,2,3} & LI DianWen^{1,2*}

¹Center for Cognition and Brain Disorders, Hangzhou Normal University, Hangzhou 310015, China;

²Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments, Hangzhou 310015, China;

³Department of Psychology, University of Minnesota, Minneapolis, MN 55455, USA

Received November 14, 2012; accepted January 6, 2013; published online March 15, 2013

Representational similarity analysis (RSA) is a rapidly developing multivariate platform to investigate the structure of neural activities. Similarity/dissimilarity is the core concept of RSA, realized by the construction of a representational dissimilarity matrix, that addresses the closeness/distance for each pair of research elements (e.g., one minus the correlation between the brain responses to 2 different stimuli) and in turn, constitutes a multivariate pattern as its analytic foundation. This approach is also welcome for its sensitivity in detecting subtle differences of distributed experimental effects in the brain. Importantly, RSA is not only an experimental tool but a promising data-analytical framework that can integrate cross-modal imaging signals, explore brain-behavior link, and verify computational models according to measured neural activities. RSA substantiates its integrative power by relating similarity structure in one domain (e.g., stimulus features) to that in another domain (e.g., neural activities). This review summarizes dissimilarity/similarity definition of RSA, introduces how to derive the dissimilarity structure in neural response pattern, and carry out connectivity analysis based on RSA platform. Several recent advances are highlighted, such as the extraction of across-subjects regularity, cross-validation of brain reactivity in human beings and monkeys, the incorporation of computational models and behavioral profiles into RSA. Voxel receptor field modeling, another promising multivariate tool of pattern elucidation, is presented and compared. The application of RSA is expected to surge and extend in many fields of neuroscience, computation, psychology and medicine. We also discuss the limitations of RSA and some critical questions that need to be addressed in future research.

representational similarity analysis, pattern information, condition-rich experiments, decode

Citation: Xue S W, Weng X C, He S, et al. Similarity representation of pattern-information fMRI. *Chin Sci Bull*, 2013, 58: 1236–1242, doi: 10.1007/s11434-013-5743-0

It is believed that there is a correspondence between psychological representation in percept and neural representation in brain [1–3]. Conventional functional magnetic resonance imaging (fMRI) studies endeavor to localize neural populations that process specific categories of stimuli [4]. Despite these fruitful results, several issues remain unclear, for example, how a category-specific neural population handles within-category complexity; thus, neurons in the face area may “identify” tens of thousands of faces and those in animal recognition area may “recognize” a wide plethora of species in animal kingdom. In addition, it has been debated that whether there is a clear-cut boundary of neural segregation so that, for example, neurons in fusiform

face area (FFA) may or may not support neural computation in animal recognition. To answer these questions, the traditional approach relates peak coordinates to a category-specific neural response. Peak coordinates can be easier to understand, but assigning them to a neural representation can be a misleading form of reductionism that ignores abundant information that is inherent in real world [5].

In contrast to localizing a neural point or coordinate with some form of maxima statistic, there is a trend of research that explores the topology or the distribution of neural codes [6,7]. In accord, the studies decoding pattern information of fMRI signals have grown in popularity recently [8–10]. Pattern-information analysis treats brain activity across multiple voxels as a pattern, extracts information encoded in the pattern as representation [11], and reveals the relation

*Corresponding author (email: dwlee_ibru@yahoo.com.tw)

with richly structured sets of phenomena: perceptual and cognitive content, as well as behavior [12]. From the viewpoint of technology, this representation approach is in contrast to conventional activation-based univariate analysis using the general linear model (GLM) where spatially clustered voxels that may contribute to specific cognitive function are commonly averaged (e.g., smoothed) and examined on a voxel-by-voxel basis (i.e., massive univariate). Only the voxels with robust statistics are reported, resulting in a loss of fine-grained spatial-pattern information naturally embedded across voxels [13]. Pattern analysis aims to capture/characterize neural distribution by multivariate or network methodology, which allows relating multi-dimensional neural responses to within-category perceptual variability and enables to delineate the distributed neural responses for the mental processing of different perceptual categories (could be composed of common and unique components). Given its emphasis on “information” rather than “activation” in the neuronal system, pattern representation investigation is more sensitive to the subtle differences of a distributed experimental effect than is GLM [14].

Currently, there are two main approaches to decoding pattern information in the brain: multivariate pattern analysis (MVPA) and representational similarity analysis (RSA). In general, MVPA uses pattern-classification algorithms that can extract diagnostic information from multi-dimensional space and separate data samples into different classes [15], i.e., MVPA determines whether the response pattern in a brain region contains the stimulus information in the experiment [16]. The logic behind is: if prediction of the stimuli from the measured/observed neural response patterns is significantly above chance level, the patterns provides information about the stimuli [17]. Due to its sensitivity in unveiling information stored in multivariate patterns of brain activity, MVPA has been developed extensively and rapidly in recent years. However, MVPA requires a priori definition of stimulus categories (typically 2–5 classes) [18] which allows to draw uni-directional conclusion from stimulus space to neural space, but not *vice versa*.

RSA is another promising tool that characterizes multivariate response patterns based on quantifying the strength of similarities among different neural patterns, rather than resorting to categorical judgment, and thus affords further analysis on the structure of representational spaces [11]. The distance in neural similarity naturally provides clustering information, enabling making inference from neural space to perceptual space. In addition, RSA is suitable for dealing with many condition-rich experiments without pre-assignment of stimulus categories. Further, RSA can serve as a general experimental and data-analytical framework for integrating several research domains including neural activity, behavioral experimentation and computational modeling [19]. In this review, we summarize major approaches and significant advances of RSA with a focus on the above three domains. Several important facets of

RSA are also highlighted, including the dissimilarity/similarity concept, investigation of dissimilarity structure in neural response pattern, connectivity analysis derived from RSA, extraction of across-subjects regularity, comparative cross-validation of human beings and monkeys, incorporation of computational models and behavioral profiles into RSA, and last, a comparison with voxel receptor field modeling. The limitations of RSA and some critical questions that need to be addressed in future research are also discussed.

1 Representing dissimilarity structure of response patterns to experimental stimuli

Representational similarity analysis characterizes representational content of stimuli-evoked brain activities by means of a representational dissimilarity matrix (RDM). In brief, each cell of RDM reflects the dissimilarity (e.g. one minus correlation coefficient) between the brain activity distributions associated with the correspondent pair of stimuli in one experiment and hence, reveals the degree to which each pair of stimuli is distinguished by neural response pattern (more detailed description below). This relationship between the stimulus property and brain-activity pattern constitutes what has been termed first-order isomorphism [20].

Many previous studies of brain representations have focused on first-order isomorphism in pre-defined brain regions or regions of interest (ROIs). An ROI is firstly demarcated by anatomical information or functional data from a separate block-localizer experiment, and then the spatially contiguous sets of voxels within the ROI are selected for pattern-information analysis. Brain activity evoked by stimuli are usually estimated under the GLM framework, and each stimulus (in opposition to each category of tens of stimuli commonly adopted in fMRI studies) is regarded as a predictor in the GLM model; then the estimated coefficient of GLM (beta value) for each predictor and each voxel is obtained and forms the basis for computation of the representational dissimilarities. In detail, the “distance” between two stimuli is defined by a measure of dissimilarity between two elicited activity patterns, which can be assessed in a variety of ways, including correlation or rank-correlation analysis, Euclidean and Manhattan distance measure, etc. A widely used measure of dissimilarity is correlation distance, i.e., one minus the correlation between patterns [21].

Direct visual inspection of RDMS may yield clustered structures in neural response patterns [22,23]. Nevertheless, the dissimilarity structures can be measured and analyzed by automated algorithms of multi-dimensional scaling (MDS) [24] and hierarchical clustering analysis [25]. MDS attempts to find a lower dimensional display space in which data points (experimental stimuli or associated dissimilarity indices to experimental stimuli in this case) can be organized in such a way that their pairwise spatial distances are

retained and approximately reflect their dissimilarity structure. In the rearrangement via MDS, data points of dissimilarity indices placed close together, i.e. shorter distance, denote similar response profiles to the correspondent stimuli. MDS is unsupervised by nature in that it does not presuppose dissimilarity structure and it aims to reveal the composition that dominates the representation of stimuli in a given ROI by data-driven processes. Hierarchical clustering analysis assists in identifying relatively homogeneous groups of stimuli, using an algorithm that starts with each stimulus in a separate cluster and combines clusters based on their closeness until only one is left [26]. For the experimental condition with a small number of stimuli, similarity-graph is often used to visualize the representation where each of the nodes and edges respectively indicates a stimulus and how these different stimuli are similar together. In summary, RSA starts with the construction of RDM which indicates the distance between neural response patterns and then discloses the underlying structure of RDM by various data re-organizing and clustering techniques, such as MDS and hierarchical clustering.

Exploration of similarity structure can be traced back in earlier functional imaging research. In 1998, Edelman et al. [2] pioneered the application of similarity analysis to fMRI activity patterns. Initially, however, this analytic scheme has not attracted much attention even after continuous publication of several well designed studies [27,28]. It is not until recently that applications of this method have gained momentum and developed rapidly. In 2011, Kravitz et al. [23] applied MDS to reconstruct and visualize representational dissimilarity structure derived from the distributed neural responses to 96 diverse real-world scenes. The authors found, contrary to previous studies [29,30], that representations in the ventral temporal parahippocampal place area (PPA) were characterized primarily by the spatial factor of expanse (open, closed) and in early visual cortex (EVC) primarily by distance (near, far), not by category or context. Kriegeskorte et al. [22] applied hierarchical clustering analysis of response patterns in human brain evoked by 92 ungrouped-object stimuli photos, and found that object representation was inherently categorical in inferotemporal cortex (IT): animate and inanimate objects form the two major clusters; faces and bodies form subclusters within the animate cluster.

2 Representational connectivity and similarity searchlight analysis

The concept of RSA can be extended to the so-called second-order isomorphism, in contrast to its first-order counterpart, which measures representational dissimilarity matrices between two given ROIs [31]. This approach yields a novel definition of connectivity where similar RDMs of two ROIs indicate substantial information exchange and hence,

neural interaction. Analogous to the concept of functional connectivity [32], this approach is termed “representational connectivity”. To quantify the match/mismatch between two RDMs, it commonly starts from assessing the discrepancy of the correlation distances ($1 - \text{correlation}$) of the two matrices. Because RDM is a mirror-symmetric matrix about a diagonal of zeros, only elements in the upper (or equivalently the lower) triangle of the matrix are used as a sample for calculation. It is noted that representational connectivity between two ROIs does not imply a direct structural connection. Nevertheless, representational connectivity promises a higher-level functional perspective to assess to what extent two regions represent the same information, which is in contrast to the perspective of first-order isomorphism that addresses the similarity in perceptual and neural spaces, and in single ROI [19].

Representational connectivity is usually combined with a similarity searchlight procedure to identify voxel clusters of shared representational dissimilarity structure across the brain, without pre-defined ROIs. The searchlight method was developed by Kriegeskorte et al. [33], and the main steps include: a sphere with a center at one voxel and user-specified radius, then move the sphere (or spotlight) throughout the brain with multivariate statistics (e.g., classification performance or indices of similarity) computed at each location, with the result stored at each voxel. In turn, the contiguous voxels identified by searchlight procedure provides analytic units for further similarity analyses, regional or inter-regional. Representational connectivity may also be further elaborated by using a broad range of network analysis approaches, such as graph theory [34].

Connolly et al. [35] have combined RSA, searchlight analysis, and cluster algorithm to explore distributed response to the stimuli images of six animal species without external assumptions. The authors discovered that the largest cluster was formed by voxels in lateral occipital complex (LOC) region, with their RDMs reflecting biological class structure of the stimuli. Mur et al. [36] extended this methodology to perform a connectivity analysis of 96 object images from a wide range of categories, including faces and places, and also humans and animals. They found functional similarities between EVC and IT cortices, and between bi-hemispheric FFAs and PPAs. Good discrimination of preferred from non-preferred stimuli was also demonstrated based on single-image activation of category-selective regions in FFA and PPA across a wide range of ROI sizes.

It is noteworthy that representational connectivity is very different from the rationales of conventional connectivity analyses, such as Kalman filter [37], structural equation model [38], Pearson correlation [39], mutual information [40], or granger causality analysis [41] where the strength in the relation between selected time courses is the main focus. Frequently, the surrogate time courses can be derived either from the averaged temporal series of the voxels in ROIs or from the temporal series of the voxels with peak statistics in

ROIs (indicating the voxels with most confident/reliable neural responses). Representation connectivity, on the contrary, emphasizes the similarity in spatial pattern and can be a complementary approach to unveil neural interaction at the scale of distributed neural network. In summary, representation connectivity analysis replaces neural node with a distributed pattern and connectivity strength with RDMS similarity.

3 Performing across-subject decoding of fine-grained neural representations

What makes different individuals' representations of the outside world alike or different? To investigate this question demands across-subject decoding, where visualizing and comparing neural similarity space by MDS and hierarchical clustering analysis are commonly used [23,35]. However, classification by intuitive visualization or averaging alone might not be self-evident at a finer grained scale and not sufficient to judge whether an attempt to decode the stimuli across subjects succeeds or fails. There is a need for new approach to further bridging representational structures of different subjects so as to accommodate population-level variation and regularity.

In 2012, Raizada et al. [42] have developed an analytic pipeline named DEMOSS (DEcoding by Matching Of Similarity Spaces) to seek to decide whether different subjects' representations are alike. The authors first employed the classic dataset from Haxby et al. [43] which reliably activates object recognition area in ventral temporal (VT) cortex, and then calculated the neural similarities between eight stimulus conditions for each of the six subjects by spatial correlation. One subject was selected for producing neural similarity matrices from all possible stimulus label-permutations ($8! = 40320$ labels); the similarity matrix which had the highest correlation with the average matrix from the other subjects was the winner to provide multivariate neural decoding for the group-level commonality. The success of this across-subject decoding strategy demonstrates that neural similarity space captures a representational scheme that is shared across individuals, even at the finer-grained level of multiple subcategories.

In contrast, it would also be worthwhile to investigate individual differences in fine-grained neural representations. Polk et al. [44] studied the similarity between twin pairs in the distribution of neural responses to faces, houses, pseudowords and chairs in the ventral stream. The authors found that face and place-related responses within face and place selective regions, respectively, were significantly more similar for monozygotic than for dizygotic twins [45]. Variability in the similarity structures of brain activity patterns may account for inter-individual difference in performing a given neuropsychological task or reflect differences in the status of the subjects. Importantly, one potential

direct application of RSA is in the domain of diagnosis, which may serve as a vehicle for distinguishing patient and control populations.

The theoretical implication for the extraction of across-subject neural representation is profound to the philosophy of mind. It has long been asked, for example, why different individuals share similar mental representations? Whether different species of mammalian decode the real world in a similar way? The study of neural similarity by Raizada et al. [42] and that by Kriegeskorte et al. [17] provide preliminary clues to these, and perhaps other more fundamental questions of human nature. The information revealed by the scatter or trace in the neural similarity space may offer evolutionary trajectories from neuroscientific account. Further, the framework of RSA may help to shape theoretical questions based on something measurable and real in the brain.

4 Identifying semantic dissimilarity structure in human brain matching monkey, behavior, and models

The comparison between human and monkey brains is important from an evolutionary perspective and has enjoyed a long history. One of the key questions is: do humans and monkeys see the world similarly? To this end, Kriegeskorte et al. [22] have estimated blood-oxygen-level dependent responses patterns elicited in human IT and neural activities recorded from monkey IT neurons by the same 92 object images (different face, body, animal, plant and artifact stimuli photographs), and then used RSA to investigate their neural similarity structures. Despite different imaging modalities and species, the authors found a high degree of match between the similarity structures of human and monkey, which were characterized by major distinctions between animate and inanimate stimuli and, within the animate domain, between faces and bodies. These studies evoke the view that the monkey brain might provide an appropriate model of the human brain not only for early sensory processing, but also at higher level of categorical object representations [18]. RSA thus offers an attractive way to decode representation content from different species and modalities of brain activity signals.

The representational dissimilarity matrix can serve as a platform that relates different representation contents from a variety of sources, especially neural and behavioral ones, which underlines the important issue of brain-behavior link. This platform based on similarity measure naturally provides clustering information, along a fine continuum of distance, which enables cross-modal comparison. Connolly et al. [35] conducted an fMRI study with a selection of photographs from 6 species and then asked the participants to rate their proximity at a post-scan session. Using behavioral judgments as target similarity structures, the authors applied the searchlight technique to map the relationship between

neural similarity and target behavioral similarity throughout the brain. The brain clusters with matched neural and behavioral similarity matrices were mainly situated in the ventral visual stream, encompassing lateral occipital complex (LOC) region. The close correspondence between the RDM of a brain region and the behavioral dissimilarity matrix would suggest that the brain region play a role in determining the measured behavioral hierarchy.

Obviously, many other forms of behavioral profiles can be related to the similarity feature of distributed brain response patterns, such as success and failure, or reaction time in performing a task. Raizada et al. [46] labeled desirable behavioral output as +1 and -1, voxels time course as input, and then applied a linear regression model to extract the brain voxels with significant power to discriminate the output labels. In contrast to ROI analysis frequently used in the field of pattern decoding in neuroimaging research, the authors used the entire brain at once as a single, large ROI, i.e. brain-wide multivariate pattern. Surprisingly, the results were quite robust and similar to other classifiers such as support vector machine and regularized logistic regression. Although the linear regression study by Raizada et al. [46] does not seem to be RSA at first glance, the neural pattern that is maximally discriminative according to a linear model actually reflects the voxels that embrace labeling information, and hence a binary pattern matching.

Computational models can be a rigorous alternative to produce target similarity structure to relate to brain activity data. The experiment stimuli may first be processed to obtain representations using models under testable hypotheses, and then their internal representations are handled in the same way as the measured neural representations described above. Here, computational models simulate certain aspects of central information processing during an experiment [47] and thus account for information represented by the relevant brain regions' response patterns. Rejection of hypotheses or selection of better models can be achieved by similarity assessment between simulation results and measured neural activities.

5 Comparison with voxel receptive-field modeling

Voxel receptive-field modeling is another attractive method, similar to RSA that integrates computational models of brain information processing into pattern-information analysis. Generally, the analytic method can be divided into two steps. First, a computational model is trained to best fit the pre-selected stimuli features and the associated brain response patterns (hence the name "receptive field"). The stimuli features here are important intermediates to define stimuli space where all the elements of the stimuli set can be projected to, which enables finer description of experimental material, resembling the strategy of categorization/

sub-categorization adopted in RSA. Then, the power of generalization can be assessed by the match/mismatch between the brain response pattern predicted by the model and that elicited by a subset of the unused stimuli [48]. Both voxel receptive-field modeling and RSA are based upon brain response patterns estimated for each single stimulus and both methods sample the stimuli space more richly [49]; however, the former uses computational models to predict response patterns, whereas the latter predicts response-pattern dissimilarities.

Although not as prevalent as RSA, several well performed studies that use receptive-field models have successfully identified and constructed brain response patterns to novel stimuli [48,50]. For example, Kay et al. [29] presented each subject with a large set of natural scene photographs while fMRI data were recorded in early visual cortex, and the authors tuned the receptive field model that best fit the stimuli features of space, orientation and spatial frequency of natural images to the multi-dimensional neural representation in V1, V2 and V3. The authors demonstrated that it was possible to decode subjects' brain activity to determine what specific image was seen, highlighting a possibility to reconstruct visual experience from the measured brain activity.

6 Future perspectives and conclusion

Strong evidence has suggested that neuronal population codes of perceptual, cognitive and motoric representations are distributed in nature, which motivates a new research direction addressing "pattern" in contrast to seeking a coordinate with peak statistics in the brain, with the latter having dominated the neuroimaging field for decades. Most previous research about RSA has been restricted to visual object recognition [51]. Recently, RSA has been recognized as a powerful tool and has been applied to a broad range of fields including auditory objects recognition [51], memory [52], language [53], emotion [54], number cognition [55], and olfaction [56]. For example, Xue et al. [57] found that the degree of consistency in neural response patterns across repeated presentations of a stimulus was positively associated with later memory for that stimulus. Gilbert et al. [58] found successful prospective memory was associated with greater encoding-retrieval similarity by comparing the similarity of brain activity patterns in encoding and retrieval trials in prospective memory paradigm. Previous studies have reported that more intense neural activity was related to better language ability [59]. However, Raizada et al. [46] found the average intensity of fMRI activation was the same for /ra/ as it was for /la/, whereas the behavioral difference to /ra/ and /la/ identification is reflected in the distinctness in the spatial activation patterns.

With the rapid development of RSA, further potential applications could include the detection of individual dif-

ferences (in contrast to the group effect derivation), may apply to diverse domains beyond perception and cognition (e.g. multivariate biomarkers for brain diseases), and may integrate cross-modal neuroimaging techniques (e.g., structural MRI, diffusion MRI, EEG) from the multivariate perspective. Abnormal neural synchronization in neuropsychiatric condition could be detected by different degrees of representational dissimilarity. Besides, RSA can be useful for facilitating condition-rich experiments, i.e., designs that distinguish many conditions. In conventional experimental designs, stimuli are commonly grouped and useful information is missed after dimensionality reduction of the stimuli space into very few categories. An ungrouped event design would enable us to map stimuli situated at higher dimension spread within a given brain region as a pattern, for example through the construction of feature space, and can be further examined by theoretical models.

However, several limitations deserve consideration when applying RSA and interpreting the results. RSA is based on dissimilarity distance between two activity patterns evoked by stimuli. The dissimilarity distance can be assessed in a variety of ways. The choice of a pattern-dissimilarity metrics requires appropriate caution about their influences on the outcome. In addition, RSA uses a randomization test and random-effects inference as a statistical harness on the relatedness between two RDMs; more statistical methods remain to be developed to accommodate various analytic parameters, different definition of metrics, reliability and stability assessment, and different similarity measures from multivariate brain-activity data.

This review summarizes several technical consideration, recent advances and potential application of RSA. As a promising analytic platform, RSA can contribute to integrate several closely-related fields, such as neural activity, behavioral experimentation and computational modeling, so as to help us to elucidate the correspondence between the representations in mind and brain, to discover both the contents and rules of neuronal population computation.

We thank Dr. Carl M. Gaspar, Shao Hanyu and Pan Liping for helpful discussions and comments on a draft of this manuscript. This work was supported by the National Natural Science Foundation of China (31070905) and National Social Science Foundation of China (11AZD119).

- 1 Decharms R C, Zador A. Neural representation and the cortical code. *Annu Rev Neurosci*, 2000, 23: 613–647
- 2 Edelman S. Representation is representation of similarities. *Behav Brain Sci*, 1998, 21: 449–467
- 3 Tong F, Pratte M. Decoding patterns of human brain activity. *Annu Rev Psychol*, 2011, 63: 483–509
- 4 Logothetis N K. What we can do and what we cannot do with fMRI. *Nature*, 2008, 453: 869–878
- 5 Carlson T A, Schrater P, He S. Patterns of activity in the categorical representations of objects. *J Cogn Neurosci*, 2003, 15: 704–717
- 6 Formisano E, De Martino F, Bonte M, et al. “Who” is saying “what”? brain-based decoding of human voice and speech. *Science*, 2008, 322: 970–973
- 7 Poldrack R A. Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, 2011, 72: 692–697
- 8 Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. *Nat Neurosci*, 2005, 8: 679–685
- 9 Haynes J D, Rees G. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci*, 2005, 8: 686–691
- 10 Friston K J. Modalities, modes, and models in functional neuroimaging. *Science*, 2009, 326: 399–403
- 11 Haxby J V. Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage*, 2012, 62: 852–855
- 12 Raizada R D S, Kriegeskorte N. Pattern-information fMRI: New questions which it opens up and challenges which face it. *Int J Imag Syst Tech*, 2010, 20: 31–41
- 13 Mur M, Bandettini P A, Kriegeskorte N. Revealing representational content with pattern-information fMRI—an introductory guide. *Soc Cogn Affect Neur*, 2009, 4: 101–109
- 14 Yang Z, Fang F, Weng X. Recent developments in multivariate pattern analysis for functional MRI. *Neurosci Bull*, 2012, 28: 399–408
- 15 Li S. Multivariate pattern analysis in functional brain imaging. *Acta Physiol Sin*, 2011, 63: 472–476
- 16 Norman K A, Polyn S M, Detre G J, et al. Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci*, 2006, 10: 424–430
- 17 Kriegeskorte N, Simmons W K, Bellgowan P S F, et al. Circular analysis in systems neuroscience: The dangers of double dipping. *Nat Neurosci*, 2009, 12: 535–540
- 18 Kriegeskorte N. Relating population-code representations between man, monkey, and computational models. *Front Neurosci*, 2009, 3: 363–373
- 19 Kriegeskorte N, Mur M, Bandettini P. Representational similarity analysis—connecting the branches of systems neuroscience. *Front Syst Neurosci*, 2008, 2: 1–28
- 20 Hatfield G. Representation and constraints: The inverse problem and the structure of visual space. *Acta Psychol*, 2003, 114: 355–378
- 21 Kiani R, Esteky H, Mirpour K, et al. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol*, 2007, 97: 4296–4309
- 22 Kriegeskorte N, Mur M, Ruff D A, et al. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 2008, 60: 1126–1141
- 23 Kravitz D J, Peng C S, Baker C I. Real-world scene representations in high-level visual cortex: It’s the spaces more than the places. *J Neurosci*, 2011, 31: 7322–7333
- 24 Borg I, Groenen P J F. *Modern Multidimensional Scaling: Theory and Applications*. 2nd ed. New York: Springer-Verlag, 2005
- 25 Von Luxburg U. A tutorial on spectral clustering. *Stat Comput*, 2007, 17: 395–416
- 26 Lemay M A, Grill W M. Modularity of motor output evoked by intraspinal microstimulation in cats. *J Neurophysiol*, 2004, 91: 502–514
- 27 Hanson S J, Matsuka T, Haxby J V. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: Is there a “face” area? *NeuroImage*, 2004, 23: 156–166
- 28 Otoole A J, Jiang F, Abdi H, et al. Partially distributed representations of objects and faces in ventral temporal cortex. *J Cogn Neurosci*, 2005, 17: 580–590
- 29 Kay K N, Naselaris T, Prenger R J, et al. Identifying natural images from human brain activity. *Nature*, 2008, 452: 352–355
- 30 Walther D B, Caddigan E, Li F, et al. Natural scene categories revealed in distributed patterns of activity in the human brain. *J Neurosci*, 2009, 29: 10573–10581
- 31 Shepard R N, Chipman S. Second-order isomorphism of internal representations: Shapes of states. *Cogn Psychol*, 1970, 1: 1–17
- 32 Friston K J. Functional and effective connectivity in neuroimaging: A synthesis. *Hum Brain Mapp*, 1994, 2: 56–78
- 33 Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. *Proc Natl Acad Sci USA*, 2006, 103: 3863–3868

- 34 Liang X, Wang J, He Y. Human connectome: Structural and functional brain networks (in Chinese). *Chin Sci Bull (Chin Ver)*, 2010, 55: 1565–1583
- 35 Connolly A C, Guntupalli J S, Gors J, et al. The representation of biological classes in the human brain. *J Neurosci*, 2012, 32: 2608–2618
- 36 Mur M, Ruff D A, Bodurka J, et al. Categorical, yet graded–single-image activation profiles of human category-selective cortical regions. *J Neurosci*, 2012, 32: 8649–8662
- 37 Aqil M, Hong K S, Jeong M Y, et al. Detection of event-related hemodynamic response to neuroactivation by dynamic modeling of brain activity. *NeuroImage*, 2012, 63: 553–568
- 38 Pearl J. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press, 2000
- 39 Stigler S M. Francis Galton’s account of the invention of correlation. *Stat Sci*, 1989, 4: 73–79
- 40 Wells III W M, Viola P, Atsumi H, et al. Multi-modal volume registration by maximization of mutual information. *Med Image Anal*, 1996, 1: 35–51
- 41 Granger C W J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 1969, 37: 424–438
- 42 Raizada R D S, Connolly A C. What makes different people’s representations alike: Neural similarity space solves the problem of across-subject fMRI decoding. *J Cogn Neurosci*, 2012, 24: 868–877
- 43 Haxby J V, Gobbini M I, Furey M L, et al. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 2001, 293: 2425–2430
- 44 Polk T A, Park J, Smith M R, et al. Nature versus nurture in ventral visual cortex: A functional magnetic resonance imaging study of twins. *J Neurosci*, 2007, 27: 13921–13925
- 45 Mahon B Z, Caramazza A. What drives the organization of object knowledge in the brain? *Trends Cogn Sci*, 2011, 15: 97–103
- 46 Raizada R D S, Tsao F M, Liu H M, et al. Linking brain-wide multivoxel activation patterns to behaviour: Examples from language and math. *NeuroImage*, 2010, 51: 462–471
- 47 Zou Q, Wang Z, Luo S W, et al. A computational coding model for saliency detection in primary visual cortex. *Chin Sci Bull*, 2012, 57: 3943–3952
- 48 Mitchell T M, Shinkareva S V, Carlson A, et al. Predicting human brain activity associated with the meanings of nouns. *Science*, 2008, 320: 1191–1195
- 49 Kriegeskorte N. Pattern-information analysis: From stimulus decoding to computational-model testing. *NeuroImage*, 2011, 56: 411–421
- 50 Naselaris T, Prenger R J, Kay K N, et al. Bayesian reconstruction of natural images from human brain activity. *Neuron*, 2009, 63: 902–915
- 51 Giordano B L, McAdams S, Zatorre R J, et al. Abstract encoding of auditory objects in cortical activity patterns. *Cereb Cortex*, 2012, doi: 10.1093/cercor/bhs162
- 52 Jenkins L J, Ranganath C. Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. *J Neurosci*, 2010, 30: 15558–15565
- 53 Lee Y S, Turkeltaub P, Granger R, et al. Categorical speech processing in Broca’s area: An fMRI study using multivariate pattern-based analysis. *J Neurosci*, 2012, 32: 3942–3948
- 54 Peelen M V, Atkinson A P, Vuilleumier P. Supramodal representations of perceived emotions in the human brain. *J Neurosci*, 2010, 30: 10127–10134
- 55 Blair K P, Rosenberg-Lee M, Tsang J M, et al. Beyond natural numbers: Negative number representation in parietal cortex. *Front Syst Neurosci*, 2012, 6: 1–17
- 56 Howard J D, Plailly J, Grueschow M, et al. Odor quality coding and categorization in human posterior piriform cortex. *Nat Neurosci*, 2009, 12: 932–938
- 57 Xue G, Dong Q, Chen C, et al. Greater neural pattern similarity across repetitions is associated with better memory. *Science*, 2010, 330: 97–101
- 58 Gilbert S J, Armbruster D J N, Panagiotidi M. Similarity between brain activity at encoding and retrieval predicts successful realization of delayed intentions. *J Cogn Neurosci*, 2012, 24: 93–105
- 59 Crinion J, Price C J. Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain*, 2005, 128: 2858–2871

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.