Search for patterns of functional specificity in the brain: A nonparametric hierarchical Bayesian model for group fMRI data

Published in final edited form as:
Neuroimage. 2012 January 16; 59(2): 1348–1368. doi:10.1016/j.neuroimage.2011.08.031.

Search for Patterns of Functional Specificity in the Brain: A Nonparametric Hierarchical Bayesian Model for Group fMRI Data

Danial Lashkari[a], **Ramesh Sridharan**[a], **Edward Vul**[b], **Po-Jang Hsieh**[b], **Nancy Kanwisher**[b], and **Polina Golland**[a]

Danial Lashkari: danial@csail.mit.edu; Ramesh Sridharan: rameshvs@mit.edu; Edward Vul: evul@ucsd.edu; Po-Jang Hsieh: pjh@mit.edu; Nancy Kanwisher: ngk@mit.edu; Polina Golland: polina@csail.mit.edu

[a]Computer Science and Artificial Intelligence Lab., Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA

[b]Brain and Cognitive Science Dept., Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA

Abstract

Functional MRI studies have uncovered a number of brain areas that demonstrate highly specific functional patterns. In the case of visual object recognition, small, focal regions have been characterized with selectivity for visual categories such as human faces. In this paper, we develop an algorithm that automatically learns patterns of functional specificity from fMRI data in a group of subjects. The method does not require spatial alignment of functional images from different subjects. The algorithm is based on a generative model that comprises two main layers. At the lower level, we express the functional brain response to each stimulus as a binary activation variable. At the next level, we define a prior over sets of activation variables in all subjects. We use a Hierarchical Dirichlet Process as the prior in order to learn the patterns of functional specificity shared across the group, which we call functional systems, and estimate the number of these systems. Inference based on our model enables automatic discovery and characterization of dominant and consistent functional systems. We apply the method to data from a visual fMRI study comprised of 69 distinct stimulus images. The discovered system activation profiles correspond to selectivity for a number of image categories such as faces, bodies, and scenes. Among systems found by our method, we identify new areas that are deactivated by face stimuli. In empirical comparisons with perviously proposed exploratory methods, our results appear superior in capturing the structure in the space of visual categories of stimuli.

Keywords

fMRI; clustering; high level vision; category selectivity

1 Introduction

It is well-known that functional specificity at least partially explains the functional organization of the brain (Kanwisher, 2010). In particular, fMRI studies have revealed a

Correspondence to: Danial Lashkari, danial@csail.mit.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

number of regions along the ventral visual pathway that demonstrate significant selectivity for certain categories of objects such as human faces, bodies, or places (Kanwisher, 2003; Grill-Spector and Malach, 2004). Most studies follow the traditional confirmatory framework (Tukey, 1977) for making inference from fMRI data. This approach first hypothesizes a candidate pattern of functional specificity. The hypothesis may be derived from prior findings or the investigator's intuition. An experiment is then designed to enable detection of brain areas that exhibit the specificity of interest. Unfortunately, fMRI data is extremely noisy and the resulting detection map does not provide a fully faithful representation of actual brain responses. In order to confirm the hypothesis, it is common to consider detection maps across different subjects and look for contiguous areas located around the same anatomical landmarks. Anatomical consistency in the detected areas attests to the validity of the hypothesis.

The traditional confirmatory approach to fMRI analysis comes with fundamental limitations when employed to search for patterns of functional specificity. Consider again the case of visual category selectivity. The space of categories that could constitute a likely grouping of objects in the visual cortex is large enough to make brute force confirmatory tests for all likely patterns of selectivity infeasible. Instead, prior studies only focus on specific categories based on semantic classifications of objects. Yet, we cannot disregard the possibility that some cortical groupings may not exactly agree with our conceptual abstractions of object classes.

Another limitation of the traditional method is its reliance on spatial correspondence across subjects for validation. It is possible that the organization of category-selective areas varies across subjects relative to anatomical landmarks. Furthermore, it is likely that, instead of contiguous blob-like structures, category selectivity appears in distributed networks of smaller regions. Yet, most fMRI analysis techniques are based on the premise that functionally specific areas are relatively large and tightly constrained by the anatomical landmarks in all subjects.

Here, we present a model for group fMRI exploratory analysis that circumvents the limitations above, building on a previously demonstrated approach (Lashkari et al., 2010b). The key idea is to employ a rich experimental design that includes a large number of stimuli and an analysis procedure that automatically searches for patterns of specificity in the resulting fMRI data. To explicitly express these patterns, we define the *selectivity profile* of a brain area to be a vector that represents this area's selectivity to different stimuli in the experiment. We employ clustering to identify *functional systems* defined as collections of voxels with similar selectivity profiles that appear consistently across subjects. The method considers all relevant brain responses to the entire set of stimuli, and automatically learns the selectivity profiles of dominant systems from the data. This framework eliminates the need for spatial correspondences.

The method presented in this paper simultaneously estimates voxel selectivity profiles, system profiles, and spatial maps from the observed fMRI time courses. Moreover, the model refines the assumptions regarding the group structure of the mixture distribution by allowing variability in the size of systems and the parameters of fMRI signals such as the hemodynamic response function (HRF) across subjects. The model further enables the estimation of the number of systems from data. Since all variables of interest are treated as latent random variables, the method yields posterior distributions that encode uncertainty in the estimates.

1.1 Nonparametric Bayesian Model for Group Clustering

We employ Hierarchical Dirichlet Processes (HDP) (Teh et al., 2006) to share structure across subjects. In our model, the structure shared across the group corresponds to grouping of voxels with similar functional responses. The nonparametric Bayesian aspect of HDPs enables automatic search in the space of models of different sizes.

Nonparametric Bayesian models have been previously employed in fMRI data analysis, particularly in modeling the spatial structure in the significance maps found by confirmatory analyses (Kim and Smyth, 2007; Thirion et al., 2007b). The probabilistic model introduced in this paper is more closely related to recent applications of HDPs to DTI data where anatomical connectivity profiles of voxels are clustered across subjects (Jbabdi et al., 2009; Wang et al., 2009). In contrast to prior methods that apply stochastic sampling for inference, we take advantage of a variational scheme that is known to have faster convergence rate and greatly improves the speed of the resulting algorithm (Teh et al., 2008).

As before, this approach uses no spatial information other than the original smoothing of the data and therefore does not suffer from the drawbacks of voxel-wise spatial normalization.

1.2 FMRI Signal Model for Activation Profiles

The goal of this work is to employ clustering ideas to automatically search for distinct forms of functional specificity in the data. Consider a study of high level object recognition in visual cortex where a number of different categories of images have been presented to subjects. Within a clustering framework, each voxel in the image can be represented by a vector that expresses how selectively it responds to different categories presented in the experiment. We may estimate the brain responses for each of the stimuli using the general linear model for fMRI signals and perform clustering on the resulting response vectors. However, the results of such an analysis may yield clusters of voxels with responses that only differ in their overall magnitude (as one can observe, e.g., in the results of Thirion and Faugeras, 2004). The vector of brain responses, therefore, does not directly express how *selectively* a given voxel responds to different stimuli.

Unfortunately, fMRI signals do not come in well-defined units of scale, making it hard to literally interpret the measured values. Univariate confirmatory methods avoid dealing with this issue by only assessing voxel contrasts, differences in signal evaluated separately in each voxel. Others instead express the values in terms of the percent changes in signal compared to some baseline, but then there is no consensus on how to define such a baseline (Thirion et al., 2007a). There is evidence that not only the characteristics of the linear BOLD response vary spatially within the brain (e.g., Schacter et al., 1997; Miezin et al., 2000; Makni et al., 2008), but the neuro-vascular coupling itself may also change from an area to another (Ances et al., 2008). A wide array of factors can contribute to this within-subject, within-session variability in fMRI measurements, from the specifics of scanners to the local tissue properties and relative distances to major vessels. As might be expected, similar factors also contribute to within-subject, across-session, as well as across-subject variations in fMRI signals, although the latter has a more considerable extent likely due to inter-subject variability in brain function (Wei et al., 2004; Smith et al., 2005).

Given the reasoning above, we aim to transform the brain responses into a space where they directly express their relative selectivity to different stimuli. Such a space allows us to compare voxel responses from different areas, and even from different subjects.

To achieve this goal, our framework includes a model for fMRI time courses that handles the ambiguity in fMRI measurements by introducing a voxel-specific amplitude of response. The model assumes that the response to each stimulus is the product of the voxel-specific amplitude of response and an activation variable. While the former encodes overall magnitude of signal, which may be a byproduct of physiological confounds such as the distance between the voxel and nearby veins, the latter measures the size of signal in the voxel in response to each stimulus when compared to others. Therefore, the activation profile of a voxel can be naturally interpreted as a signature of functional specificity: it describes the probability that any stimulus or task may activate that brain location.

The remainder of the paper is organized as follows. Section 2 presents a review of prior work on exploratory fMRI analysis. In Section 3, we describe the two main layers of the model, the fMRI signal model and the hierarchical clustering model, and discuss the variational procedure for inference on the latent variables of the model. We present the results of applying the algorithm to data from a study of human visual object recognition and compare them with results found by the finite mixture model clustering model (Lashkari et al., 2010b) and the tensorial group ICA (Beckmann and Smith, 2005) in Section 4. This is followed by discussion in Section 5 and conclusions in Section 6.

2 Prior Work on Exploratory fMRI Analysis

Early work on clustering of fMRI data typically employed fuzzy clustering, which allows soft cluster assignments, in simple fMRI studies of early visual areas (Baumgartner et al., 1997, 1998; Moser et al., 1997; Golay et al., 1998, and references therein). Baumgartner et al. (2000) reported superior performance of clustering compared to PCA and Moser et al. (1999) suggested that it can be used for removing motion confounds. Variants of fuzzy clustering (Chuang et al., 1999; Fadili et al., 2000; Jarmasz and Somorjai, 2003), *K*-means (Filzmoser et al., 1999), and other heuristic clustering techniques (Baune et al., 1999) have been applied to fMRI data, but little evidence exists for advantages of clustering beyond the experimental settings where they were first reported. A mixture model formulation of clustering has been employed in (Golland et al., 2007, 2008) to recover a hierarchy of large-scale brain networks in resting state fMRI.

Applying clustering directly to the time course data, as described above, may not be the best strategy when it comes to discovering task-related patterns. First, the high dimensionality of fMRI time courses makes the problem challenging since noise represents a large proportion of the variability in the observed signals. Second, the spatially varying properties of noise may increase the dissimilarity between the time courses of different activated areas. Third, in order to interpret the results, one must determine the relationship between the estimated cluster mean time courses and different experimental conditions, usually through a post hoc stage of regression or correlation.

Alternatively, some clustering methods use information from the experimental paradigm to define a measure of similarity between voxels, effectively projecting the original high-dimensional time courses onto a low dimensional feature space, and then perform clustering in the new space (Goutte et al., 1999, 2001; Thirion and Faugeras, 2003, 2004). The paradigm-dependent feature space represents the dimensions of interest in fMRI measurements. For instance, if the experiment involves a paradigm that is rich enough, we can simply cluster vectors of estimated regression coefficients for all stimuli in the experiment (Thirion and Faugeras, 2004; Lashkari et al., 2010b).

We previously demonstrated a clustering method that consists of two separate stages (Lashkari et al., 2010b). We first computed voxel selectivity profiles using regression estimates from the standard linear model and then clustered profiles from all subjects together. This analysis does not account for inter-subject variability and provides no obvious choice for the number of clusters. The unified model presented in this paper integrates the two steps into a single estimation procedure and incorporates model selection.

Independent component analysis (ICA) is another popular exploratory technique commonly applied to fMRI data. McKeown et al. (1998) employed a basic noiseless ICA algorithm for the analysis of fMRI data and demonstrated improved results compared to PCA (see also McKeown and Sejnowski, 1998; Biswal and Ulmer, 1999). Beckmann and Smith (2004) proposed a probabilistic formulation that includes Gaussian noise. When applied directly to fMRI time courses, interpretation of ICA components still requires relating the estimated component time courses to the experimental conditions. Balslev et al. (2002) provides an example of regression on component time courses to identify relevant systems.

Similar to standard confirmatory techniques, most extensions of exploratory methods to multisubject data rely on voxel-wise correspondence. Using this framework, Beckmann and Smith (2005) proposed a tensorial group factorization of the data within the ICA framework. This method factorizes the group fMRI data into a number of components. Each component is characterized by a time course and a group spatial map defined in the population template. The only across-subject variability assumed is the differences in the contribution of each group component to measurements in different individuals. In contrast, our approach to group analysis avoids making any assumptions about spatial correspondences of functional areas across subjects. Spatial maps for different clusters are defined in each subject's native space.

The model developed in the next section can be viewed as a Bayesian probabilistic extension of simple mixture model clustering that includes three important elements: 1) a nonparametric prior that enables estimation of the number of components, 2) a hierarchical structure that enables group analysis, and 3) an fMRI time course model that explicitly accounts for the experimental paradigm.

3 Methods

Consider an fMRI experiment with a relatively large number of different tasks or stimuli, for instance, a design that presents S distinct images in an event-related visual study. We let $\boldsymbol{y}_{ji}$ be the acquired fMRI time course of voxel i in subject j. The goal of the analysis is to identify patterns of functional specificity, i.e., distinct profiles of response that appear consistently across subjects in a large number of voxels in the fMRI time courses $\{\boldsymbol{y}_{ji}\}$. We refer to a cluster of voxels with similar response profiles as a functional system. Figure 1 illustrates the idea of a system as a collection of voxels that share a specific functional profile across subjects. Our model characterizes the functional profile as a vector whose components express the probability that the system is activated by the stimuli in the experiment.

To define the generative process for fMRI data, we first consider an infinite number of group-level systems. System k is assigned a prior probability π_k of including any given voxel. While the vector π is infinite-dimensional, any finite number of draws from this distribution will obviously yield a finite number of systems. To account for inter-subject variability and noise, we perturb the group-level system weight π independently for each subject j to generate a subject-specific weight vector β_j. System k is further characterized by a vector $[\varphi_{k1}, \cdots, \varphi_{kS}]^t$, where $\varphi_{ks} \in [0, 1]$ is the probability that system k is activated by stimulus s. Based on the weights β_j and the system probabilities φ, we generate binary activation variables $x_{jis} \in \{0, 1\}$ that express whether voxel i in subject j is activated by stimulus s.

So far, the model has the structure of a standard HDP. The next layer of this hierarchical model defines how activation variables x_{jis} generate observed fMRI signal values y_{jit}. If the voxel is activated ($x_{jis} = 1$), the corresponding fMRI response is characterized by a positive voxel-specific response magnitude a_{ji}; if the voxel is non-active ($x_{jis} = 0$) the response is assumed to be zero. The model otherwise follows the standard fMRI linear response model where the HRF is assumed to be variable across subjects and is estimated from the data.

Below, we present the details of the model starting with the lower level signal model to provide an intuition on the representation of the signal via activation vectors and then move on to describe the hierarchical clustering model. Table 1 presents the summary of all variables and parameters in the model; Figure 2 shows the structure of our graphical model.

3.1 Model for fMRI Signals

Using the standard linear model for fMRI signals (Friston et al., 2007), we model measured signal $\boldsymbol{y}_{ji}$ of voxel i in subject j as a linear combination

$$\boldsymbol{y}_{ji}=\boldsymbol{G}_j\boldsymbol{b}_{ji}+\boldsymbol{F}_j\boldsymbol{e}_{ji}+\boldsymbol{\varepsilon}_{ji}, \quad (1)$$

where $\boldsymbol{G}_j$ and $\boldsymbol{F}_j$ are the stimulus and nuisance components of the design matrix for subject j, respectively, and ε_{ji} is Gaussian noise. To facilitate our derivations, we rewrite this equation explicitly in terms of columns of the design matrix:

$$\boldsymbol{y}_{ji}=\sum_s b_{jis}\,\boldsymbol{g}_{js}+\sum_d e_{jid}\boldsymbol{f}_{jd}+\boldsymbol{\varepsilon}_{ji}, \quad (2)$$

where $\boldsymbol{g}_{js}$ is the column of matrix $\boldsymbol{G}_j$ that corresponds to stimulus s and $\boldsymbol{f}_{jd}$ represents column d of matrix $\boldsymbol{F}_j$.

We devise a model that integrates this representation with binary *activation variables* $\boldsymbol{x}$ that connect the signal model with the hierarchical prior. If voxel i in subject j is activated by stimulus s, i.e., if $x_{jis} = 1$, its response takes positive value a_{ji} that specifies a voxel-specific *amplitude of response*; otherwise, its response remains 0. Using this parametrization, $b_{jis} = a_{ji}x_{jis}$. The response amplitude a_{ji} represents uninteresting variability in fMRI signal due to physiological reasons unrelated to neural activity (examples include proximity of major blood vessels).

To explicitly describe the properties of the hemodynamic response, we define $\boldsymbol{g}_{js} = \xi_{js} * \boldsymbol{h}_j$ where $\xi_{js} \in IR^T$ is a binary indicator vector that shows whether stimulus $s \in \mathscr{S}$ is present during the experiment for subject j at each of the T acquisition times, and $\boldsymbol{h}_j \in IR^L$ is a finite-time vector characterization of the hemodynamic response function (HRF) in subject j.

It is common in fMRI analysis to use a canonical shape for the HRF, letting $\boldsymbol{h}_j = \bar{\boldsymbol{h}}$ for all subjects j. However, prior research has demonstrated considerable variability in the shape of the HRF across subjects (Aguirre et al., 1998; Handwerker et al., 2004). We define $\boldsymbol{h}_j$ to be the shape of the HRF for subject j. It is a latent variable that is inferred from data. To simplify future derivations, we let Ξ_{js} be a $T \times L$ convolution matrix derived from the stimulus indicator vector ξ_{js} such that $\boldsymbol{g}_{js} = \Xi_{js}\boldsymbol{h}_j$. Here, we assume a shared HRF for all voxels in a subject since our application involves only the visual cortex. For studies that investigate responses of the entire brain or several cortical areas, the model can be easily generalized to include separate HRF variables for different areas (Makni et al., 2005).

With all the definitions above, our fMRI signal model becomes

$$\boldsymbol{y}_{ji}=a_{ji}\left(\sum_s x_{jis}\,\Xi_{js}\right)\boldsymbol{h}_j+\sum_d e_{jid}\boldsymbol{f}_{jd}+\boldsymbol{\varepsilon}_{ji}. \quad (3)$$

We use a simplifying assumption throughout that $\boldsymbol{\varepsilon}_{ji}\overset{i.i.d.}{\sim}\mathrm{Normal}(\mathbf{0},\lambda_{ji}^{-1}\mathbf{I})$. In the application of this model to fMRI data, we first apply temporal filtering to the signal to decorrelate the noise in the preprocessing stage (Burock and Dale, 2000; Bullmore et al., 2001; Woolrich et al., 2001). An extension of the current model to include colored noise is possible, although it has been suggested that noise characteristics do not greatly impact the estimation of the HRF (Marrelec et al., 2002).

3.1.1 Priors—We assume a multivariate Gaussian prior for $\boldsymbol{h}_j$, with a covariance structure that encourages temporal smoothness,

$$\boldsymbol{h}_j\sim\mathrm{Normal}\left(\bar{\boldsymbol{h}},\boldsymbol{\Lambda}^{-1}\right), \quad (4)$$

$$\boldsymbol{\Lambda}=\nu\mathbf{I}+\boldsymbol{\Delta}^t\boldsymbol{\Delta}, \quad (5)$$

where

$$\boldsymbol{\Delta}=\begin{pmatrix}1 & -1 & 0 & \cdots & 0 & 0\\ 0 & 1 & -1 & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & 1 & -1\end{pmatrix}, \quad (6)$$

and $\bar{\boldsymbol{h}}$ is the canonical HRF. The definition of the precision matrix above yields a prior that involves terms of the form $\sum_{l=1}^{L-1}(h_l-h_{l+1})^2$, penalizing differences between the values of the HRF at consecutive time points.

We assume the prior distributions on the remaining voxel response variables as follows. For the response magnitude, we assume

$$a_{ji}\sim\mathrm{Normal}_+\left(\mu_j^a,\sigma_j^a\right), \quad (7)$$

where Normal$_+$(η, ρ) is the conjugate prior defined as a normal distribution restricted to positive real values:

$$p(a)\propto e^{-(a-\eta)^2/2\rho},\ \text{for } a\geq 0. \quad (8)$$

Positivity of the variable a_{ji} simply reflects the constraint that the expected value of fMRI response in the active state is greater than the expected value of response in the non-active state. For the nuisance factors, we let

$$e_{jid}\sim\mathrm{Normal}\left(\mu_{jd}^e,\sigma_{jd}^e\right), \quad (9)$$

where Normal(η, ρ) is a Gaussian distribution with mean η and variance ρ. Finally, for the noise precision parameter, we assume

$$\lambda_{ji} \sim \text{Gamma}\ \left(\kappa_j, \theta_j\right), \quad (10)$$

where Gamma (κ, θ) is a Gamma distribution parametrized by shape parameter κ and scale parameter θ^{-1}:

$$p(\lambda) = \frac{1}{\theta^{-\kappa}\Gamma(\kappa)}\lambda^{\kappa-1}e^{-\theta\lambda}. \quad (11)$$

3.2 Hierarchical Dirichlet Prior for Modeling Variability across Subjects

Our model assumes a shared clustering structure in the fMRI activations ***x*** that allows inter-subject variability in the size of clusters across subjects. We further include a nonparametric prior to estimate the number of clusters supported by the observed data.

Similar to standard mixture models, we define the distribution of a voxel activation variable x_{jis} by conditioning on the system membership $z_{ji} \in \{1, 2, \cdots\}$ of the voxel and on the system probabilities of activation for different stimuli $\varphi = \{\varphi_{ks}\}$:

$$x_{jis} \mid z_{ji}, \boldsymbol{\phi} \overset{i.i.d.}{\sim} \text{Bernoulli}(\phi_{z_{ji}s}). \quad (12)$$

This model implies that all voxels within a system have the same prior probability of being activated by a particular stimulus *s*.

We place a Beta prior distribution on system-level activation probabilities φ:

$$\phi_{ks} \overset{i.i.d.}{\sim} \text{Beta}\left(\omega^{\phi,1}, \omega^{\phi,2}\right). \quad (13)$$

Parameters ω^{φ} control the overall proportion of activated voxels across all subjects. For instance, we can induce sparsity in the results by introducing bias towards 0, i.e., the non-active state, in the parameters of this distribution.

To capture variability in system weights, we assume:

$$z_{ji} \mid \boldsymbol{\beta}_j \overset{i.i.d.}{\sim} \text{Mult}(\boldsymbol{\beta}_j), \quad (14)$$

$$\boldsymbol{\beta}_j \mid \boldsymbol{\pi} \overset{i.i.d.}{\sim} \text{Dir}(\alpha\boldsymbol{\pi}), \quad (15)$$

where β_j is a vector of subject-specific system weights, generated by a Dirichlet distribution centered on the population-level system weights π. The extent of variability in the size of different systems across subjects is controlled by the concentration parameter α of the Dirichlet distribution. Finally, we place a prior on the population-level weight vector π that allows an infinite number of components:

$$\boldsymbol{\pi} \mid \gamma \sim \text{GEM}(\gamma), \quad (16)$$

where GEM(γ) is a distribution over infinitely long vectors $\pi = [\pi_1, \pi_2, \cdots]^t$, named after Griffiths, Engen and McCloskey (Pitman, 2002). Specifically,

$$\pi_k = \upsilon_k \prod_{k'=1}^{k-1} (1 - \upsilon_{k'}),$$

$$\upsilon_k \mid \gamma \overset{i.i.d.}{\sim} \text{Beta}\,(1, \gamma)\,. \quad (17)$$

It can be shown that the components of the generated vectors π sum to 1 with probability 1. With this prior over system memberships ***z*** = {z_{ji}}, the model in principle allows for an infinite number of functional systems; however, for any finite set of voxels, a finite number of systems is sufficient to include all voxels.

This prior for activation variables corresponds to the stick-breaking construction of HDPs (Teh et al., 2006), which is particularly suited for the variational inference scheme that we discuss in the next section.

3.3 Variational EM Inference

Having devised a full model for the fMRI measurements in a multi-stimulus experiment, we now provide a scheme for inference on the latent variables from the observed data. Sampling schemes are most commonly used for inference in HDPs (Teh et al., 2006). Despite theoretical guarantees of convergence to the true posterior, sampling techniques generally require a time-consuming burn-in phase. Because of the relatively large size of our problem, we will use a collapsed variational inference scheme for inference (Teh et al., 2008), which is known to yield faster algorithms. Here, we provide a brief overview of the derivation steps for the update rules. Appendix A contains the update rules and more detailed derivations.

3.3.1 Formulation—To formulate the inference for system memberships, we integrate over the subject-specific unit weights β = {β_j} and introduce a set of auxiliary variables ***r*** = {r_{jk}} that represent the number of tables corresponding to system *k* in subject *j* according to the Chinese Restaurant Process formulation of HDP in (Teh et al., 2006). Appendix A provides some insights into the role of these auxiliary variables in our model; they allow us to find closed-form solutions for the inference update rules. We let ***u*** = {***x***, ***z***, ***r***, φ, π, υ, ***a***, ***e***, ***h***, λ} denote the set of all latent variables in our model. In the framework of variational inference, we approximate the model posterior on ***u*** given the observed data *p*(***u***|***y***) by a distribution *q*(***u***). The approximation is performed through the minimization of the Gibbs free energy function:

$$\mathscr{F}[q] = E[\log\, q(\boldsymbol{u})] - E[\log\, p(\boldsymbol{y}, \boldsymbol{u})]. \quad (18)$$

Here, and in the remainder of the paper, *E*[·] and *V* [·] indicate expected value and variance with respect to distribution *q*. We assume a distribution *q* of the form:

$$q(\boldsymbol{u}) = q(\boldsymbol{r}|\boldsymbol{z}) \left(\prod_k q(\nu_k)\right) \cdot \left(\prod_{k,s} q(\phi_{ks})\right) \cdot \prod_j \left\{ q(\boldsymbol{h}_j) \prod_i \left[q(a_{ji}) q(\lambda_{ji}) q(z_{ji}) \left(\prod_s q(x_{jis})\right) \left(\prod_d q(e_{jid})\right)\right]\right\}, \quad (19)$$

where we explicitly account for the dependency of the auxiliary variables ***r*** on the system memberships ***z***. Including this structure maintains the quality of the approximation despite the introduction of the auxiliary variables (Teh et al., 2007). We use coordinate descent to solve the resulting optimization problem. Minimizing the Gibbs free energy function in terms of each component of $q(\boldsymbol{u})$ while fixing all other parameters leads to closed form update rules, provided in Appendix A.

3.3.2 Initialization—

Iterative application of the update rules leads to a local minimum of the Gibbs free energy. Since variational solutions are known to be biased toward their initial configurations, the initialization phase becomes critical to the quality of the results. We can initialize the variables in the fMRI signal model by ignoring higher level structure of the model and separately fitting the linear model of Equation (3) to the observed signal in each subject, starting with the canonical form of the HRF. Note that these estimates are the same as the traditional GLM estimates used in most fMRI analyses. Our method begins with these estimates and modifies them according to the assumptions made in the model.

The standard least squares regression produces estimates for coefficients b_{jis} in Equation (3) that describe the contribution of each condition to signal in different voxels. In our model, we assume that these coefficients can be factored as $b_{jis} = a_{ji}x_{jis}$ to positive voxel-specific response amplitudes a_{ji} and activation variables x_{jis}. Therefore, for the initialization we let $E[a_{ji}] = \max_s \hat{b}_{jis}$ and $E[x_{jis}] = (\hat{b}_{jis} - \min_s \hat{b}_{jis}) / (\max_s \hat{b}_{jis} - \min_s \hat{b}_{jis})$, where $\hat{b}_{jis}$ is the least squares estimate based on the standard GLM. We initialize nuisance factors $\boldsymbol{e}_{ji}$ directly to the values of the nuisance regressor coefficients obtained via least squares estimation, and variance reciprocals of noise λ_{ji} to values found based on the estimated residuals.

To initialize system memberships, we introduce voxels one by one in a random order to the collapsed Gibbs sampling scheme (Teh et al., 2006) constructed for our model with each stimulus as a separate category and the initial $\boldsymbol{x}$ assumed known. In contrast to the initialization of the other variables, the initialization of system memberships has a random nature and we repeat it several times to find the configuration that yields the best Gibbs free energy.

The update rules for each variable usually depend only on the previous values of other variables in the model. The exception to this is the update for $q(x_{jis})$, which also depends on previous estimates of $\boldsymbol{x}$. Therefore, unless we begin by updating $\boldsymbol{x}$, the first variable to be updated does not need to be initialized. Due to the coupling of the initializations for $\boldsymbol{x}$ and $\boldsymbol{a}$, we can choose to initialize either one of them first and update the other next. By performing both variants and choosing the one that provides the lower free energy after convergence, we further improve the search in the space of possible initializations and the quality of the resulting estimates.

4 Results

This section presents the results of applying our method to data from an event-related visual fMRI experiment. We compare the results of our hierarchical Bayesian method with the finite mixture model (Lashkari et al., 2010b) and the tensorial group ICA of (Beckmann and Smith, 2005) in a high-level visual experiment.

4.1 Data

Ten subjects were scanned in an event-related experiment. Each subject was scanned in two 2-hour scanning sessions. During the scanning session, the subjects were presented with images from nine categories (animals, bodies, cars, faces, scenes, shoes, tools, trees, vases) in the event-related paradigm. Images were presented in a pseudo-randomized design generated by optseq (Dale, 1999) to optimize the efficiency of regression. During each 1.5s presentation, the image moved slightly across the field of view either leftward or rightward. Subjects were asked to indicate the direction of motion by pressing a button. Half of the image set was presented in the first session, and the other half was presented in the second session. Figure 3 shows the stimuli used in this study.

Functional MRI data were collected on a 3T Siemens scanner using a Siemens 32-channel head coil. The high-resolution slices were positioned to cover the entire temporal lobe and part of the occipital lobe (gradient echo pulse sequence, TR = 2s, TE = 30ms, 40 slices with a 32 channel head coil, slice thickness = 2mm, in-plane voxel dimensions = 1.6 × 1.6mm). The anatomical scans were obtained at an isotropic resolution of 1mm in all three directions, and were subsequently subsampled to an isotropic resolution of 2mm.

The data was first motion corrected separately for the two sessions (Cox and Jesmanowicz, 1999) and spatially smoothed with a Gaussian kernel of 3mm width. We then registered the two sessions to the subject's native anatomical space (Greve and Fischl, 2009). We used FMRIB's Improved Linear Model (FILM) to prewhiten the acquired fMRI time courses before applying the linear model (Woolrich et al., 2001).

We created a mask for the analysis in each subject using an omnibus *F*-test that determines whether any stimulus regressors significantly explain the variations in the measured fMRI time course ($p = 10^{-6}$). This step essentially removed noisy voxels from the analysis and only retained areas that are relevant for the experimental protocol at hand. Since the goal of the analysis is to study high level functional specificity in the visual cortex, we further removed from the mask the set of voxels within early visual areas. Furthermore, we included the average time course of all voxels within early visual areas as a confound factor in the design matrix of Equation (3). This procedure selected between 2700 to 6800 voxels for different subjects and a total of 50435 voxels for all 10 subjects. Our method works directly on the temporally filtered time courses of all voxels within the mask.

4.2 Comparison and Evaluation

We compare our results with those of the finite mixture model of (Lashkari et al., 2010b) and tensorial group ICA (Beckmann and Smith, 2005). Below, we first describe the parameters and settings used with each of these methods and then introduce measures employed in our evaluation of their results.

4.2.1 Nonparametric Hierarchical Model—For HDP scale parameters, we use $\alpha = 100$, $\gamma = 5$. We show in Section 4.5 that the results are not sensitive to this specific choice. We also set $\omega^{\varphi,1} = \omega^{\varphi,2} = 1$ for the nonparametric prior to assume a uniform prior on activation probabilities. For the signal model, we use $\nu = 100$, and estimate the remaining hyperparameters of the fMRI signal model as follows. Like the initialization procedure of Section 3.3.2, we begin by applying least squares regression based on the standard GLM model of the signal. For each subject, we create empirical distributions for the estimated values of the fMRI signal variables a_{ji} and e_{jid} from the GLM estimates, and λ_{ji} from the residuals. Based on the assumptions made in Section 3.1.1, we fit the prior models to these empirical distributions and find maximum likelihood estimates of the hyperparameters $\mu_j^a, \sigma_j^a, \mu_{jd}^e, \sigma_{jd}^e$, κ_j, and θ_j.

We run the algorithm 20 times with different initializations for system memberships and choose the solution that yields the least Gibbs free energy function.

4.2.2 Finite Mixture Model—When evaluating the finite mixture model, we apply the standard regression analysis to find regression coefficients for each stimulus at each voxel and use the resulting vectors as inputs for clustering. Like ours, this method is also initialized with 20 random sets of parameters and the best solution in terms of log-likelihood is chosen as the final result.

In (Lashkari et al., 2010b), we provided an approach to quantifying and validating the group consistency of each profile found by the finite mixture model. We use this method to provide an ordering of the resulting systems in terms of their consistency scores. We define the consistency scores based on the correlation coefficients between the group-wise profiles with the selectivity profiles found in each subject. We first match group-wise profiles with the set of profiles found by the algorithm in each individual subject's data separately. We employ the Hungarian algorithm (Kuhn, 1955) to find the matching between the two sets of profiles that maximizes the sum of edge weights (correlation coefficients in this case).[1] The consistency score for each system is then defined as the average correlation coefficient between the corresponding group-wise profile and its matched counterparts in different subjects.

As we discuss in Section 5, basic model selection schemes for finite mixture model fail to provide a reasonable choice for the number of clusters. However, the results remain qualitatively similar when we change the number of clusters (Lashkari et al., 2010b). Given that we expect at least 3 or 4 areas selective for faces, scenes, and bodies, we choose $K = 15$ clusters to allow for several novel likely systems. Among the resulting profiles, we select systems whose consistency scores are significant at threshold $p = 10^{-3}$ based on the group-wise permutation test. We demonstrated in (Lashkari et al., 2010b) that the finite mixture modeling results are qualitatively insensitive to changes in the numbers of clusters.

4.2.3 Tensorial Group ICA—Tensorial group ICA requires spatial normalization of the functional data from different subjects to a common spatial template. We employ FMRIB's nonlinear image registration tool[2] (FNIRT) to register the structural image from each subject to the MNI template (T1 image of MNI152). As an initialization for this registration, we use FMRIB's linear image registration tool[3] (FLIRT). We create a group mask for the ICA analysis defined as the union of the masks found for different subjects by the *F*-test procedure above. We use the Melodic[4] implementation of the tensorial group ICA provided within the FSL package. Since the experiment includes a different number of runs for each subject, we cannot directly apply the ICA algorithm to the time courses. Instead, we use vectors of estimated regression coefficients for the 69 stimuli at each voxel as the input to ICA.

As implemented in the Melodic package, the tensorial group ICA employs the automatic model selection algorithm of Minka (2001) to estimate the number of independent components (Beckmann and Smith, 2004).

ICA provides one group spatial map for each estimated component across the entire group. In contrast, our method yields subject-specific maps in each subject's native space. In order to summarize the maps found by our method in different subjects and compare them with their ICA counterparts, we apply the same spatial normalization described above to spatial maps of the discovered systems. We then average these normalized maps across subjects to produce a group summary of the results.

4.2.4 Classification and Consistency Scores—As a quantitative way to evaluate the specificity patterns found by different methods, we define a classification score for each set of system (or component) profiles that measures how well they encode information about

[1] We use the open source matlab implementation of the Hungarian algorithm available at http://www.mathworks.com/matlabcentral/fileexchange/11609.

[2] http://www.fmrib.ox.ac.uk/fsl/fnirt/index.html

[3] http://fsl.fmrib.ox.ac.uk/fsl/flirt/

[4] http://www.fmrib.ox.ac.uk/fsl/melodic/index.html

stimulus categories. Each system activation profile in our model represents the probabilities that different stimuli activate that system. Therefore, the brain response to stimulus s can be summarized based on our results in terms of a vector of activations $[E[\varphi_{1s}], \cdots, E[\varphi_{Ks}]]^t$ that it induces over the set of all functional systems. Similarly, finite mixture profiles and ICA component profiles can be used as stimulus representations, which may in turn be used to classify stimuli. We consider all distinct binary classification problems involving all pairs of the first 8 categories (we do not include the 5 vase images in this analysis since there are fewer samples from this category). We apply 8-fold cross validation with linear SVM classifiers trained on the profiles and define the average classification accuracy for all 28 binary classification problems on the test data as the classification score.

We also use the consistency scores, which were defined above for finite mixture modeling, in our sensitivity and reproducibility analyses. In each of these cases, we aim to quantify the similarity of two sets of system (or component) profiles, e.g., when assessing the consistency of the results across two subgroups of data. In each case, we apply the Hungarian algorithm to find the one-to-one matching that maximizes the pairwise correlation coefficients and define the average correlation coefficient between matched profiles to be the consistency score of the results.

To test statistical significance of the consistency scores, we create a permutation-based null distribution for the pairwise correlation coefficients between two groups of matched profiles. For each sample, we randomly permute the S components of all profiles independently of each other. We then apply the matching, calculate pairwise correlation coefficients between matched profiles, and compute their average, i.e., the consistency score. We create 10, 000 samples of the consistency score in this way and use this empirical distribution to evaluate the significance of the average consistency score of the original result.

Systems or components found by the methods discussed in this paper do not come in a unique order or with unique labels. As mentioned earlier, for the finite mixture model we use the consistency scores to create a ranking that allows us to focus on the more relevant systems. For the two other methods, we use similar measures that capture the *variability in the size of systems across subjects* to provide an ordering of the profiles for their visualization. In tensorial group ICA results, variable c_{jk} expresses the contribution of component k to the fMRI data in subject j. Similarly, variable $E[n_{jk}]$ in our results denotes the number of voxels in subject j assigned to system k. We define this measure for system (component) k as the standard deviation of values of $E[n_{jk}]$ (or c_{jk}) across subjects when scaled to have unit average. We rank our profiles based on this measure in ascending order and label them accordingly.

4.3 System Functional Profiles

We apply our method to the real data from the visual experiment. In the data from ten subjects, the method finds 25 systems. Figure 4 presents the posterior activation probability profiles of these 25 functional systems in the space of the 69 stimuli presented in the experiment. We compare these profiles with the ones found by the finite mixture model and the group tensorial ICA, presented in Figure 5. ICA yields ten components. The profiles in Figure 4 and Figure 5 are presented in order of consistency. In the results from all methods, there are some systems or components that mainly contribute to the results of one or very few subjects and possibly reflect idiosyncratic characteristics of noise in those subjects.

Qualitatively, we observe that the category structure is more salient in the results of the nonparametric model. Most of our systems demonstrate similar probabilities of activation for images that belong to the same category. This structure is present to a slightly lesser extent in the results of the finite mixture model, but is much weaker in the ICA results.

More specifically, we identify systems 2, 9, and 12 in Figure 4 as selective for categories of bodies, faces, and scenes, respectively (note that animals all have bodies). Among the system profiles ranked as more consistent, these profiles stand out by the sparsity in their activation probabilities. Figure 5 shows that similarly selective systems 1 (faces), 2 (bodies), 3 (bodies), and 5 (scenes) also appear in the results of the finite mixture model. The ICA results include only one component that seems somewhat category selective (component 1, bodies). As discussed in Section 1, previous studies have robustly localized areas such as EBA, FFA, and PPA with selectivities for the three categories above. Automatic detection of these profiles demonstrates the potential of our approach to discover novel patterns of specificity in the data.

Inspecting the activation profiles in Figure 4, we find other interesting patterns. For instance, the three non-face images with the highest probability of activating the face selective system 9 (animals 2, 5 and 7) correspond to the three animals that have large faces (Figure 3). Beyond the three known patterns of selectivity, we identify a number of other notable systems in the results of Figure 4. For instance, system 1 shows lower responses to cars, shoes, and tools compared to other stimuli. Since the images representing these three categories in our experiment are generally smaller in terms of overall pixel size and overall image intensity, this system appears selective to lower level features (note that the highest probability of activation among shoes corresponds to the largest shoe 2). The correlation coefficient between this profile and the sum of the intensity values of the 69 images is 0.48, where a correlation value of 0.35 is in this case significant at $p = 0.05$ with Bonferroni corrections for 25 profiles. System 3 and system 8 are less responsive to faces compared to all other stimuli.

To quantify how well each set of profiles encodes the category information in images, we compute the classification scores of the three methods. For our method, the finite mixture model, and tensorial group ICA, the score, which represents average classification accuracy in binary category classification tasks, is equal to 0.95 ± 0.16, 0.97 ± 0.13, and 0.68 ± 0.31, respectively. We also apply the finite mixture model with $K = 30$ and find the classification score of the results to be 0.96 ± 0.13. The average classification score of our method is significantly greater than that of ICA with $p = 10^{-4}$ based on a nonparametric permutation test. This suggests that while nonparametric and finite mixture models yield similar classification performance for encoding category information, they both show much higher performance than that of ICA.

We investigate the spatial properties of the detected systems in the next section.

4.4 System Spatial Maps

For each system k in our results, vector $\{q(z_{ji}=k)\}_{i=1}^{V_j}$ describes the posterior membership probability for all voxels in subject j. We can represent these probabilities as a spatial map for the system in the subject's native space. Figure 6 (top) shows the membership maps for the systems 2 (bodies), 9 (faces), and 12 (scenes). For comparison, Figure 6 (bottom) shows the significance maps found by applying the conventional confirmatory t-test to the data from the same subject. These maps present uncorrected significance values $-\log_{10}(p)$ for each of the three standard contrasts of bodies-objects, faces-objects, and scenes-objects, thresholded at $p = 10^{-4}$ as is common practice in the field. While the significance maps appear to be generally spatially larger than the systems identified by our method, close inspection reveals that the system membership maps include the peak voxels for their corresponding contrasts. Figure 7 illustrates the fact that voxels within our system membership maps are generally associated with high significance values for the contrasts that correspond to their respective selectivity. The figure also clearly shows that there is considerable variability across subjects in the distribution of significance values.

Our method calculates the spatial maps in each subjects native anatomical space while we have to normalize the data before applying ICA so it finds a group map in the population template. As a result, we cannot directly compare the spatial properties of maps found by the two methods. To make an indirect comparison, we normalize the system probability maps of different subject to the population template and then average them to find the proportion of subjects whose system maps includes any given voxel in their system maps. Figure 9 compares this group-average of spatial maps for the body-selective system 2 with the group-level spatial map of component 1 found by ICA. Although both maps cover the same approximate anatomical areas, our group map includes very few voxels with values close to 1 suggesting that areas associated with body-selectivity do not have high voxel-wise overlap across subjects. In other words, the location of body-selective system 2 varies across subjects but generally remains at the same approximate area. This result, which agrees with the findings previously reported in the literature (Spiridon et al., 2006), does not appear in the ICA map that includes large areas with a maximum value of 1.

Figure 9 presents average normalized spatial maps for two other selective systems 9 and 12. These maps clearly contain previously identified category selective areas, such as FFA, OFA, PPA, TOS, and RSC (Kanwisher and Yovel, 2006; Epstein et al., 2007). We also examine the spatial map for system 1, which we demonstrated to be sensitive to low-level features. As Figure 10 (left) shows, this system resides mainly in the early visual areas. Figure 10 (right) shows the spatial map for system 8, which exhibits reduced activation to faces and shows a fairly consistent structure across subjects. To the best of our knowledge, selectivity similar to that of system 8 has not been reported in the literature so far.

4.5 Sensitivity Analysis

We test the sensitivity of our method to different initializations for system memberships and also to perturbations in values of HDP scale parameters.

Figure 11 (left) shows the histogram of classification scores for the results found from 20 different initializations of the algorithm. We observe that the performance of all different initializations is very similar to the results presented in Section 4.3 that achieve the least Gibbs free energy. In Figure 11 (right), we show the correlation coefficients between profiles found by each initialization and their matching profiles from the best results in terms of the Gibbs free energy. Please note that in cases where the numbers of systems in the two results are not equal, we assumed a correlation coefficient of zero for the systems that do not have a match. These zero correlation coefficients explain why the average consistency scores are less than the consistency scores of most profiles in the results. Nevertheless, the average consistency scores are significant for all 20 results at $p = 10^{-4}$. This analysis confirms that the results of our algorithm are robust across different initializations of the algorithm.

Most hyperparameters of our model have intuitive interpretations in terms of the parameters of fMRI signal model and are selected based on the GLM estimates from the data. Selection of HDP scale parameters, on the other hand, is not as straightforward. For the results presented in this section so far, we chose $(\gamma, \alpha) = (5, 100)$. To investigate how sensitive the results are with respect to this specific choice, we run the algorithm with a few different other choices for the HDP scale parameters. We first perturb each parameter slightly around the choice $(\gamma, \alpha) = (5, 100)$ by varying the values of γ and α by ±1 and ±10, respectively.

We then increase the range of change by roughly dividing or multiplying each parameter by 2. Table 2 presents the resulting number of systems and classification scores for all these changes of HDP scale parameters. We observe that the category information remains at similar levels when we change the parameters. To directly assess the similarity of the results to the ones presented in Section 4.3, Figure 12 reports the consistency scores when matching system profiles found with different HDP scale parameters to the results reported in Section 4.3. All average consistency scores are significant with $p = 10^{-4}$. This analysis confirms that our results are insensitive to the choice of HDP scale parameters.

Varying initializations or model parameters in Figures 11 (right) and 12, although the profiles *on average* remain consistent with the system profiles presented in Figure 4, we observe some degree of variation in consistency scores. If we investigate the results more closely, we find an interesting structure in these variations: the labeling of systems, which is based on the consistency of their sizes across subjects (Section 4.2.4), is highly correlated with their consistency across different initializations or model parameters. Figure 13 presents the correlation coefficients between each system profile of Figure 4 and the profiles matched to it in the results from 20 different initializations or from 8 different configrations of HDP scale parameters. The figure provides another way for examining the consistency scores in Figures 11 (right) and 12. Here, we can see that higher ranked profiles are generally more consistent. The ranking of voxels has correlation coefficients −0.74 and −0.65 with the average consistency of match profiles (blue squares in Figure 13) across different initializations and model parameters (both significant with $p = 10^{-4}$ in a permutation test). This result suggest that the systems that are more relevant in our analysis, i.e., the ones that appear more consistently across subjects, remain more consistent in the results found with different initializations and model parameters.

4.6 Reproducibility Analysis

In this section, we validate our results based on their reproducibility across subjects. We split the ten subjects into two groups of five and apply the analysis separately to each group. The method finds two sets of 17 and 23 systems in the two groups.

Figure 14 shows the system profiles in both groups of subjects matched with the top 13 consistent profiles of Figure 4. Visual inspection of these activation profiles attests to the generalization of our results from one group of subjects to another. Figure 15 reports correlation coefficients for pairs of matched profiles from the two sets of subjects for all three methods: our Bayesian nonparametric method, the finite mixture-model, and the group tensorial ICA. Average consistency scores for both nonparametric and finite mixture models are significant at $p = 10^{-4}$. In contrast, the p-value for the average consistency score of ICA profiles is only 0.05. This result suggests that, in terms of robustness across subjects, our unified model is more consistent than tensorial group ICA and is comparable to the finite mixture model. We note that due to the close similarity in the assumptions of our model and the finite mixture model, we do not expect a significant change in the robustness of the results when comparing the two models.

5 Discussion

The nonparametric nature of the model developed in this paper represents an important advantage over the finite mixture models (Golland et al., 2007; Lashkari et al., 2010b). The nonparametric construction enables the estimation of the number of systems from the data. In our experience, both basic model selection schemes such as BIC (Schwarz, 1978) and computationally intensive resampling methods such as that of Lange et al. (2004) yield monotonically increasing measures for the goodness of the finite mixture model up to cluster numbers in several hundreds. In contrast, our nonparametric method automatically finds the number of components within the expected range based on prior information. The estimates depend on the choice of HDP scale parameters α and γ. The results provide optimal choices within the neighborhood of model sizes allowed by these parameters. We also showed in our sensitivity analysis in Section 4.5 that the results remain fairly consistent as we change the HDP scale parameters.

Like the finite mixture model, the proposed hierarchical Bayesian model avoids making assumptions about the spatial organization of functional systems across subjects. This is in contrast to tensorial group ICA, which assumes that independent components of interest are in voxel-wise correspondence across subjects. Average spatial maps presented in the previous section clearly demonstrate the extent of spatial variability in functionally specific areas. This variability violates the underlying ICA assumption that independent spatial components are in perfect alignment after spatial normalization. Accordingly, ICA results are sensitive to the specifics of spatial normalization. In our experience, changing the parameters of registration algorithms can considerably alter the profiles of estimated independent components.

As mentioned earlier, Makni et al. (2005) also employed an activation model similar to ours for expressing the relationship between fMRI activations and the measured BOLD signal. The most important distinction between the two models is that the amplitude of activations in the model of Makni et al. (2005) is assumed to be constant across all voxels. In contrast, we assume a voxel-specific response amplitude that allows us to extract activation variables as a relative measure of response in each voxel independently of the overall magnitude of the BOLD response.

A more subtle difference between the two models lies in the modeling of noise in time courses. Makni et al. (2005) assume two types of noise. First, they include the usual time course noise term ε_{jit} as in Equation (3). Moreover, they assume that the regression coefficients b_{is} are generated by a Gaussian distribution whose mean is determined by whether or not voxel i is activated by stimulus s, i.e., the value of the activation variable x_{is}. This model assumes a second level of noise characterized by the uncertainty in the values of the regression coefficients conditioned on voxel activations. Our model is more parsimonious in that it does not assume any further uncertainty in brain responses conditioned on voxel activations and response amplitudes.

We emphasize the advantage of the activation profiles in our method over the cluster selectivity profiles of the finite mixture modeling in terms of *interpretability*. Our definition of a classification score uses the fact that vectors formed by concatenating components of different system profiles that correspond to the same stimulus can be used as a representation for the stimulus. In the case of our fMRI signal model, this representation has an intuitive interpretation as the probability that the stimulus can activate a given system. In contrast, the finite mixture modeling of (Lashkari et al., 2010b) defines the system profiles as vectors of unit length. As a result, it is not straightforward how we can interpret different components of each profile vector.

We note at that a preliminary version of the model demonstrated in this paper was presented elsewhere (Lashkari et al., 2010a).

6 Conclusion

In this paper, we developed a nonparametric hierarchical Bayesian model that allows us to infer patterns of functional specificity that consistently appear across subjects in fMRI data. The model accounts for inter-subject variability in the size of functionally specific systems.

It enables estimation of the number of systems from the data. In addition, we endow the model with a layer that explicitly connects fMRI activations to the observed time courses. We derived a variational inference algorithm for fitting the model to the data from a group of subjects. Most notably, the method does not require spatial alignment of the functional data across the group in order to perform group analysis.

We apply our method to an fMRI study of visual object recognition that presents 69 distinct images to ten subjects. The algorithm successfully discovers system activation profiles that correspond to well-known patterns of category selectivity along with a number of novel systems. These systems include one that is deactivated by face images. We showed that the results of our method are not sensitive with respect to changes in the initialization and model parameters and are reproducible across different groups of subjects.

Acknowledgments

This work was supported in part by the McGovern Institute Neurotechnology Program, NIH grants NIBIB NAMIC U54-EB005149 and NCRR NAC P41-RR13218 to PG, NEI grant 13455 to NGK, NSF grant CAREER 0642971 to PG and IIS/CRCNS 0904625 to PG and NGK, in addition to an MIT HST Catalyst grant and an NDSEG fellowship to EV.

Appendix

A Derivations of the Update Rules

In this section, we derive the Gibbs free energy cost function for variational inference and derive the update rules for inference using the variational approximation.

A.1 Joint Probability Distribution

Based on the generative model described in Section 3, we form the full joint distribution of all the observed and unobserved variables. For each variable, we use $\omega^{\cdot}$ to denote the natural parameters of the distribution for that variable. For example, the variable e_{jid} is associated with natural parameters $\omega_{jd}^{e,1}$ and $\omega_{jd}^{e,2}$.

A.1.1 fMRI model—Given the fMRI model parameters, we can write the likelihood of the observed data ***y***:

$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{a},\boldsymbol{\lambda},\boldsymbol{e},\boldsymbol{h})=\prod_{j,i}\sqrt{\frac{\lambda_{ji}^{T_j}}{2\pi}}\exp\left\{-\frac{\lambda_{ji}}{2}\left\|\boldsymbol{y}_{ji}-\sum_{d}e_{jid}\boldsymbol{f}_d-a_{ji}\sum_{s}x_{jis}\Xi_{js}\boldsymbol{h}_j\right\|^2\right\}. \quad \text{(A.1)}$$

We now express the priors on the parameters of the likelihood model defined in Section 3.1.1 in the new notation. Specifically, for the nuisance parameters ***e***, we have

$$p(e_{jid})=\mathrm{Normal}(\mu_{jd}^{e},\sigma_{jd}^{e}) \quad \text{(A.2)}$$

$$\propto \exp\left\{-\frac{1}{2}(\omega_{jd}^{e,2})e_{jid}^{2}+\frac{1}{2}(\omega_{jd}^{e,1})e_{jid}\right\}, \quad \text{(A.3)}$$

where $\omega_{jd}^{e,2}=(\sigma_{jd}^{e})^{-1}$ and $\omega_{jd}^{e,1}=\mu_{d}^{e}(\sigma_{jd}^{e})^{-1}$.

With our definition of the Gamma distribution in Equation (11), the natural parameters for the noise precision variables **λ** are $\omega_{jm}^{\lambda,1}=\kappa_{jm}$ and $\omega_{jm}^{\lambda,2}=\theta_{jm}$.

The distribution over the activation heights ***a*** is given by

$$p(a_{jim})=\text{Normal}_{+}(\mu_{jm}^{a},\sigma_{jm}^{a}) \quad \text{(A.4)}$$

$$\propto \exp\left\{-\frac{1}{2}(\omega_{jm}^{a,2})a_{jim}^{2}+\frac{1}{2}(\omega_{jm}^{a,1})a_{jim}\right\}, a_{jim}\geq 0 \quad \text{(A.5)}$$

We have $\omega_{jm}^{a,2}=(\sigma_{jm}^{a})^{-1}$ and $\omega_{jm}^{a,1}=\mu_{jm}^{a}\left(\sigma_{jm}^{a}\right)^{-1}$.

The distribution Normal$_{+}$(η, ρ^{-1}) is a member of an exponential family of distributions and has the following properties:

$$p(a)=\sqrt{\frac{2\lambda}{\pi}}\left[1+\text{erf}\left(\sqrt{\frac{\rho}{2}}\eta\right)\right]^{-1}e^{-\rho(a-\eta)^{2}/2}, \quad \text{(A.6)}$$

$$E[a]=\eta+\sqrt{\frac{2}{\pi\lambda}}\left[1+\text{erf}\left(\sqrt{\frac{\rho}{2}}\eta\right)\right]^{-1}e^{-\rho\eta^{2}/2}, \quad \text{(A.7)}$$

$$E[a^{2}]=\eta^{2}+\rho^{-1}+\eta\sqrt{\frac{2}{\pi\lambda}}\left[1+\text{erf}\left(\sqrt{\frac{\rho}{2}}\eta\right)\right]^{-1}e^{-\rho\eta^{2}/2}. \quad \text{(A.8)}$$

A.1.2 Nonparametric Hierarchical Joint Model for Group fMRI Data—

The voxel activation variables x_{jis} are binary, with prior probability φ_{ks} given according to cluster memberships. Since φ ~ Beta(ω$^{\varphi,1}$, ω$^{\varphi,2}$), the joint density of ***x*** and φ conditioned on the cluster memberships ***z*** is defined as follows:

$$p(\boldsymbol{x},\boldsymbol{\phi}|\boldsymbol{z})=\prod_{j,k,s}\left[\frac{\Gamma(\omega^{\phi,1}+\omega^{\phi,2})}{\Gamma(\omega^{\phi,1})\Gamma(\omega^{\phi,2})}\phi_{ks}^{\omega^{\phi,1}-1+\sum_{i,s}x_{jis}\delta(z_{ji},k)}\times(1-\phi_{ks})^{\omega^{\phi,2}-1+\sum_{i,s}(1-x_{jis})\delta(z_{ji},k)}\right].$$

We assume a hierarchical Dirichlet process prior over the functional unit memberships, with subject-level weights **β**. We use a collapsed variational inference scheme (Teh et al., 2008), and therefore marginalize over these weights:

$$p(\boldsymbol{z}|\boldsymbol{\pi},\alpha)=\int_{\beta}p(\boldsymbol{z}|\boldsymbol{\beta})p(\boldsymbol{\beta}|\boldsymbol{\pi},\alpha), \quad \text{(A.9)}$$

$$=\prod_{j=1}^{J}\left[\frac{\Gamma(\alpha)}{\Gamma(\alpha+N_{j})}\prod_{k=1}^{K}\frac{\Gamma(\alpha\pi_{k}+n_{jk})}{\Gamma(\alpha\pi_{k})}\right], \quad \text{(A.10)}$$

where K is the number of non-empty functional units in the configuration and $n_{jk}=\sum_{i=1}^{N_j}\delta(z_{ji},k)$. To provide conjugacy with the Dirichlet prior for the group-level functional unit weights π, we prefer the terms in Equation (A.10) that include weights to appear as powers of π_k. However, the current form of the conditional distribution makes the computation of the posterior over π hard. To overcome this challenge, we note that for $0 \le r \le n$, we have $\sum_{r=0}^{n}\begin{bmatrix} n \\ r \end{bmatrix}\vartheta^r=\Gamma(\vartheta+n)/\Gamma(\vartheta)$, where $\begin{bmatrix} n \\ r \end{bmatrix}$ are unsigned Stirling numbers of the first kind (Antoniak, 1974). The collapsed variational approach uses this fact and the properties of the Beta distribution to add an auxiliary variable $\boldsymbol{r} = \{r_{ji}\}$ to the model:

$$p(\boldsymbol{z},\boldsymbol{r},|\boldsymbol{\pi},\alpha) \propto \prod_{j=1}^{J}\prod_{k=1}^{K}\begin{bmatrix} n_{jk} \\ r_{jk} \end{bmatrix}(\alpha\pi_k)^{r_{jk}}, \quad \text{(A.11)}$$

where $r_{jk} \in \{0, 1, \cdots, n_{ji}\}$. If we marginalize the distribution (A.11) over the auxiliary variable, we obtain the expression in (A.10).

A.2 Minimization of the Gibbs Free Energy

Let $\boldsymbol{u} = \{\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{r}, \varphi, \pi, \upsilon, \boldsymbol{a}, \boldsymbol{e}, \boldsymbol{h}, \lambda\}$ denote the set of all unobserved variables. In the framework of variational inference, we approximate the posterior distribution $p(\boldsymbol{u}|\boldsymbol{y})$ of the hidden variables $\boldsymbol{u}$ given the observed $\boldsymbol{y}$ by a distribution $q(\boldsymbol{u})$. The approximation is performed through minimization of the Gibbs free energy function in Equation (18) with an approximate posterior distribution $q(\boldsymbol{u})$ of the form in Equation (19). We derive a coordinate descent method where in each step we minimize the function with respect to one of the components of $q(\cdot)$, keeping the rest constant.

A.2.1 Auxiliary variables—

Assuming that all the other components of the distribution q are constant, we obtain:

$$\mathscr{F}[q(\boldsymbol{r}|\boldsymbol{z})]=E_{\boldsymbol{z}}\left[\sum_{\boldsymbol{r}} q(\boldsymbol{r}|\boldsymbol{z})\left(\log\ q(\boldsymbol{r}|\boldsymbol{z})+\ -\sum_{j,k}\left\{\log\begin{bmatrix} n_{jk} \\ r_{jk} \end{bmatrix}+r_{jk}\,E[\log(\alpha\pi_k)]\right\}\right)\right]+\text{const.} \quad \text{(A.12)}$$

The optimal posterior distribution on the auxiliary variables takes the form

$$q^*(\boldsymbol{r}|\boldsymbol{z})=\prod_j\prod_k q(r_{jk}|\boldsymbol{z}). \quad \text{(A.13)}$$

Under q^*, we have for the auxiliary variable $\boldsymbol{r}$:

$$q(r_{jk}|z)=\frac{\Gamma(\tilde{\omega}_{jk}^{r})}{\Gamma(\tilde{\omega}_{jk}^{r}+n_{jk})}\begin{bmatrix} n_{jk} \\ r_{jk} \end{bmatrix}(\tilde{\omega}_{jk}^{r})^{r_{jk}}. \quad \text{(A.14)}$$

This distribution corresponds to the probability mass function for a random variable that describes the number of tables that n_{jk} customers occupy in a Chinese Restaurant Process with parameter $\tilde{\omega}_{jk}^{r}$ (Antoniak, 1974). The optimal value of the parameter $\tilde{\omega}_{jk}^{r}$ is given by

$$\log\ \tilde{\omega}_{jk}^{r}=E[\log(\alpha\pi_k)]=\log\ \alpha+E[\log\ \upsilon_k]+\sum_{k'<k}E[\log(1-\upsilon_{k'})]. \quad \text{(A.15)}$$

As a distribution parameterized by log $\tilde{\omega}^r_{jk}$, Equation (A.14) defines a member of an exponential family of distributions. The expected value of the auxiliary variable r_{jk} is therefore:

$$E[r_{jk}|\boldsymbol{z}]=\frac{\partial}{\partial \log \tilde{\omega}^r_{jk}}\log\frac{\Gamma(\tilde{\omega}^r_{jk}+n_{jk})}{\Gamma(\tilde{\omega}^r_{jk})} \quad \text{(A.16)}$$

$$=\tilde{\omega}^r_{jk}\Psi(\tilde{\omega}^r_{jk}+n_{jk})-\tilde{\omega}^r_{jk}\Psi(\tilde{\omega}^r_{jk}), \quad \text{(A.17)}$$

where $\Psi(\omega)=\frac{\partial}{\partial\omega}\log\Gamma(\omega)$. This expression is helpful when updating the other components of the distribution. Accordingly, we obtain expectation:

$$E[r_{jk}]=E_z[E_r[r_{jk}|z]]=\tilde{\omega}^r_{jk}E_z[\Psi(\tilde{\omega}^r_{jk}+n_{jk})-\Psi(\tilde{\omega}^r_{jk})]. \quad \text{(A.18)}$$

Under $q(\boldsymbol{z})$, each variable n_{jk} is the sum of N_j independent Bernoulli random variables $\delta(z_{ji}, k)$ for $1 \le i \le N_j$ with the probability of success $q(z_{ji} = k)$. Therefore, as suggested in (Teh et al., 2008), we can use the Central Limit Theorem and approximate this term using a Gaussian distribution for $n_{jk} > 0$.

Due to the independence of these Bernoulli variables, we have

$$\Pr(n_{jk}>0)=1-\prod_{i=1}^{N_j}(1-q(z_{ji}=k)), \quad \text{(A.19)}$$

$$E[n_{jk}]=E[n_{jk}|n_{jk}>0]\ \Pr(n_{jk}>0), \quad \text{(A.20)}$$

$$E[n^2_{jk}]=E[n^2_{jk}|n_{jk}>0]\ \Pr(n_{jk}>0), \quad \text{(A.21)}$$

which we can use to easily compute $E^+[n_{jk}] = E[n_{jk}|n_{jk} > 0]$ and $V^+[n_{jk}] = V[n_{jk}|n_{jk} > 0]$. We then calculate $E[r_{jk}]$ using Equation (A.18) by noting that

$$E_z[\Psi(\tilde{\omega}^r_{jk}+n_{jk})-\Psi(\tilde{\omega}^r_{jk})]\approx\Pr(n_{jk}>0)\left[\Psi(\tilde{\omega}^r_{jk}+E^+[n_{jk}])-\Psi(\tilde{\omega}^r_{jk})+\frac{V^+[n_{jk}]}{2}\Psi''(\tilde{\omega}^r_{jk}+E^+[n_{jk}])\right]. \quad \text{(A.22)}$$

Lastly, based on the auxiliary variable $\boldsymbol{r}$, we find that the optimal posterior distribution of the system weight stick-breaking parameters is given by $\upsilon_k\sim\mathrm{Beta}(\tilde{\omega}^{\upsilon,1}_k,\tilde{\omega}^{\upsilon,2}_k)$, with parameters:

$$\tilde{\omega}^{\upsilon,1}_k=1+\sum_j E[r_{jk}] \quad \text{(A.23)}$$

$$\tilde{\omega}^{\upsilon,2}_k=\gamma+\sum_{j,k'>k}E[r_{jk'}] \quad \text{(A.24)}$$

A.2.2 System memberships—The optimal posterior over the auxiliary variables defined in Equation (A.13) implies:

$$E[\log\, q^{*}(\boldsymbol{r}|\boldsymbol{z}) - \log\, p(\boldsymbol{z},\boldsymbol{r}\mid\pi,\alpha)] = \sum_{j}\,(\log\,\Gamma(\alpha + N_j) - \log\,\Gamma(\alpha)) + \sum_{jk} E_z\,\left[\log\,\Gamma(\tilde{\omega}^{r}_{jk}) - \log\,\Gamma(\tilde{\omega}^{r}_{jk} + n_{jk})\right]. \quad \text{(A.25)}$$

The Gibbs free energy as a function of the posterior distribution of a single membership variable *q*(*z$_{ji}$*) becomes

$$\begin{aligned}\mathscr{F}[q(z_{ji})] = &\sum_{k} q(z_{ji} = k)\,\log\,q(z_{ji} = k) - \sum_{k} E_z\,\left[\log\,\Gamma(\tilde{\omega}^{r}_{jk} + n_{jk})\right] \\ &- \sum_{k} q(z_{ji} = k)\sum_{s}\,[q(x_{jis} = 1)E[\log\,\phi_{ks}] + q(x_{jis} = 0)E[\log(1 - \phi_{ks})]] + \text{const.}\end{aligned} \quad \text{(A.26)}$$

We can simplify the second term on the right hand side of Equation (A.26) as:

$$E_z\,\left[\log\,\Gamma(\tilde{\omega}_{r_{ji}} + n_{jk})\right] = E_z\,\left[\delta(z_{ji},k)\,\log(\tilde{\omega}^{r}_{jk} + n^{\neg ji}_{jk}) + \log\,\Gamma(\tilde{\omega}^{r}_{jk} + n^{\neg ji}_{jk})\right], \quad \text{(A.27)}$$

$$= q(z_{ji} = k)E_{z^{\neg ji}}[\log(\tilde{\omega}^{r}_{jk} + n^{\neg ji}_{jk})] + E_{z^{\neg ji}}[\log\,\Gamma(\tilde{\omega}^{r}_{jk} + n^{\neg ji}_{jk})], \quad \text{(A.28)}$$

where $n^{\neg ji}_{jk}$ and $\boldsymbol{z}^{\neg ji}$ indicate the exclusion of voxel *i* in subject *j* and only the first term is a function of *q*(*z$_{ji}$*). Minimizing Equation (A.26) yields the following update for membership variables:

$$q(z_{ji} = k) \propto \exp\,\left\{E_{z^{\neg ji}}[\log(\tilde{\omega}^{r}_{jk} + n^{\neg ji}_{jk})] + \sum_{s}\,(q(x_{jis} = 1)E[\log\phi_{k,l}] + q(x_{jis} = 0)E[\log(1 - \phi_{k,s})])\right\},$$

In order to compute the first term on the right hand side, as with the Equation (A.22), we use a Gaussian approximation for the distribution of *n$_{jk}$*:

$$E_{z^{\neg ji}}[\log(\tilde{\omega}^{r}_{jk} + n^{\neg ji}_{jk})] \approx \log(\tilde{\omega}^{r}_{jk} + E[n^{\neg ji}_{jk}]) - \frac{V[n^{\neg ji}_{jk}]}{2(\tilde{\omega}^{r}_{jk} + E[n^{\neg ji}_{jk}])^{2}}. \quad \text{(A.29)}$$

A.2.3 Voxel activation variables—We form the Gibbs free energy as a function only of the posterior distribution of voxel activation variables ***x***. For notational convenience, we define $\psi_{jis} = \Sigma_k\, E[\log(\varphi_{ks})]q(z_{ji} = k)$ and $\bar{\psi}_{jis} = \Sigma_{k,l}\, E[\log(1 - \varphi_{ks})]q(z_{ji} = k)$ and obtain

$$\begin{aligned}\mathscr{F}[q(x)] = \sum_{x} q(x)\,\Bigg\{&\log\,q(x) - \sum_{jis}\,\Bigg[(1 - x_{jis})\bar{\psi}_{jis} + x_{jis}\,\Bigg(\psi_{jis} + E[\lambda_{ji}]\,\Bigg[E[a_{ji}]E[\boldsymbol{h}_j]^{t}\Xi^{t}_{js}(\boldsymbol{y}_{ji} - \sum_{d} E[e_{jid}]\boldsymbol{f}_{id}) \\ &- \frac{1}{2}E[a^{2}_{ji}]\,\Bigg(E[\boldsymbol{h}^{t}_{j}\Xi^{t}_{js}\Xi_{js}\boldsymbol{h}_j] + 2\sum_{s'\neq s} E[x_{jis'}]E[\boldsymbol{h}^{t}_{j}\Xi^{t}_{js}\Xi_{js'}\boldsymbol{h}_j]\Bigg)\Bigg)\Bigg]\Bigg\} + \text{const.}\end{aligned} \quad \text{(A.30)}$$

Minimization of this function with respect to *q*(***x***) = Π$_{j,i,s}$ *q*(*x$_{jis}$*) yields the update rule:

$$\begin{aligned} q(x_{jis}{=}1) \propto \exp\ \Bigg\{ \psi_{jis}+E[\lambda_{ji}]\ \Bigg[E[a_{ji}]E[\boldsymbol{h}_j]^t\Xi^t_{js}(\boldsymbol{y}_{ji}-\sum_d E[e_{jid}]\boldsymbol{f}_{id}) \\ -\frac{1}{2}E[a^2_{ji}]\ \mathrm{Tr}\ \left(E[\boldsymbol{h}_j\boldsymbol{h}^t_j]\ \left(\Xi^t_{js}\Xi_{js}+2\sum_{s'\neq s}E[x_{jis'}]\Xi^t_{js}\Xi_{js'} \right)\right)\Bigg]\Bigg]\Bigg\} \end{aligned} \quad (A.31)$$

$$q(x_{jis}{=}0) \propto \exp\ \{\bar{\psi}_{jis}\}, \quad (A.32)$$

where Tr(·) is the trace operator.

A.2.4 fMRI model variables—

We collect the free energy terms corresponding to the nuisance variables ***e***:

$$\begin{aligned} \mathscr{F}[q(e)]=\int_e q(e)\ \Bigg(\log\ q(e)+\frac{1}{2}e^2_{jid}\omega^{e,2}_{jd}-\frac{1}{2}e_{jid}\omega^{e,1}_{jd}+\sum_{j,i,d}\frac{E[\lambda_{ji}]}{2}\ \Big[e^2_{jid}\|\boldsymbol{f}_{id}\|^2 \\ -e_{jih}\boldsymbol{f}^t_{jd}\left(\boldsymbol{y}_{ji}-\sum_{d'\neq d}E[e_{jid'}]\boldsymbol{f}_{jd'}-E[a_{ji}]\sum_s E[x_{jis}]\Xi_{js}E[\boldsymbol{h}_j]\right)\Big]\Big]\Bigg) + \text{const.} \end{aligned} \quad (A.33)$$

Recall that we assume a factored form for *q*(***e***) = Π$_{j,i,d}$ *q*(*e$_{jid}$*). Minimizing with respect to this distribution yields $q(e_{jid})\ \propto\ \exp\left\{-\frac{1}{2}\tilde{\omega}^{e,2}_{jid}e^2_{jid}+\frac{1}{2}\tilde{\omega}^{e,1}_{jid}e_{jid}\right\}$, with the parameters $\tilde{\omega}^{e,1}_{jid}$and $\tilde{\omega}^{e,2}_{jid}$ given in the Table A.1.

For the activation heights ***a***, we find

$$\begin{aligned} \mathscr{F}[q(\boldsymbol{a})]=\int_a q(\boldsymbol{a})\ \Bigg(\log\ q(\boldsymbol{a})+\frac{1}{2}a^2_{ji}\omega^{a,2}_j-\frac{1}{2}a_{ji}\omega^{a,1}_j \\ +\sum_{j,i}\frac{E[\lambda_{ji}]}{2}\ \Bigg[a^2_{ji}\sum_{s,s'}E[x_{jis}x_{jis'}]E[\boldsymbol{h}^t_j\Xi^t_{js}\Xi_{js'}\boldsymbol{h}_j]-a_{ji}\sum_s E[x_{jis}]E[\boldsymbol{h}_j]^t\Xi^t_{js}(\boldsymbol{y}_{ji}-\sum_d E[e_{jid}]\boldsymbol{f}_{id})\Bigg]\Bigg) + \text{const.} \end{aligned} \quad (A.34)$$

Assuming a factored form, minimization yields $q(a_{ji})\ \propto\ \exp\left\{-\frac{1}{2}a^2_{ji}\tilde{\omega}^{a,2}_{ji}+\frac{1}{2}a_{ji}\tilde{\omega}^{a,1}_{ji}\right\}$, ***a*** ≥ 0, with parameters $\tilde{\omega}^{a,1}_{ji}$and $\tilde{\omega}^{a,2}_{ji}$ given in Table A.1.

The terms relating to the noise precisions λ are computed as:

$$\begin{aligned}\mathscr{F}[q(\boldsymbol{\lambda})]=\int_{\boldsymbol{\lambda}}q(\lambda)\ &\left\{\log\ q(\boldsymbol{\lambda})-\sum_{j,i}\ \left(\log(\lambda_{ji})\ (\omega_j^{\lambda,1}-1)+\lambda_{ji}\omega_j^{\lambda,2}\right.\right.\\ &-\frac{T_j}{2}\log(\lambda_{ji})+\frac{\lambda_{ji}}{2}\ \left[\|\boldsymbol{y}_{ji}\|^2+E[a_{ji}^2]\sum_{s,s'}E(x_{jis}x_{jis'}]E[\boldsymbol{h}_j^t\Xi_{js}^t\Xi_{js'}\boldsymbol{h}_j]\right.\\ &+\sum_d\ \left(E[e_{jid}^2]\|\boldsymbol{f}_{id}\|^2+\sum_{d'\neq d}E[e_{jid}]E[e_{jid'}]\boldsymbol{f}_{jd}^t\boldsymbol{f}_{jd'}\right)-\boldsymbol{y}_{ji}^t\ \left(\sum_d E[e_{jid}]\boldsymbol{f}_{id}+E[a_{ji}]\sum_s E[x_{jis}]\Xi_{js}E[\boldsymbol{h}_j]\right)\\ &\left.\left.\left.+E[a_{ji}]\sum_{s,d}E[e_{jid}]E[x_{jis}]\boldsymbol{f}_{jd}^t\Xi_{js}E[\boldsymbol{h}_j]\right]\right)\right\}+\text{const.}\end{aligned}\tag{A.35}$$

Minimization with respect to $q(\lambda_{ji})$ yields $q(\lambda_{ji})\ \propto\ \exp\left\{\log(\lambda_{ji})(\tilde{\omega}_{ji}^{\lambda,1}-1)-\lambda_{ji}\tilde{\omega}_{ji}^{\lambda,2}\right\}$, where the parameters $\tilde{\omega}_{ji}^{\lambda,1}$and $\tilde{\omega}_{ji}^{\lambda,2}$ are given in Table A.1. Finally, we can write the term involving the HRF as:

$$\begin{aligned}\mathscr{F}[q(\boldsymbol{h})]=\int_h q(\boldsymbol{h})\ &\left(\log\ q(\boldsymbol{h})\right.\\ &+\sum_j\ \left[\frac{1}{2}\boldsymbol{h}_j^t\boldsymbol{\Lambda}\boldsymbol{h}_j\right.\\ &+\sum_i E[\lambda_{ji}]\sum_{s,s'}E[x_{jis}x_{jis'}]\boldsymbol{h}_j^t\Xi_{js'}^t\Xi_{js}\boldsymbol{h}_j\\ &-\boldsymbol{h}_j^t\boldsymbol{\Lambda}\bar{\boldsymbol{h}}\\ &-\sum_{i,s}E[\lambda_{ji}]E[a_{ji}]E[x_{jis}]\boldsymbol{h}_j^t\Xi_{js}^t\ \left(\boldsymbol{y}_{ji}\right.\\ &\quad\left.\left.\left.-\sum_d E[e_{jid}]\boldsymbol{f}_{id}\right)\right]\right.+\text{const.}\end{aligned}\tag{A.36}$$

Assuming an approximate factored posterior distribution $q(\boldsymbol{h}) = \Pi_j\ q(\boldsymbol{h}_j)$ and minimizing the above cost function shows that the posterior for each HRF is of the form $q(\boldsymbol{h}_j)\ \propto\ \exp\left\{-\frac{1}{2}\boldsymbol{h}_j^t\Xi_j\boldsymbol{h}_j+\frac{1}{2}\boldsymbol{h}_j^t\tilde{\omega}_j^h\right\}$ with parameters $\tilde{\omega}_j^h$ and Ξ presented in Table A.1.

A.2.5 System Activation Probabilities—For the system activation profiles, we find

$$\begin{aligned}\mathscr{F}[q[(\phi_{ks})]=\int_{\upsilon}q(\phi_{ks})\ &\left(\log\ q(\phi_{ks})-\sum_k\ \left[\left\{\omega^{\phi,1}+\sum_{j,i,s}q(z_{ji}=k)q(x_{jis}=1)\right\}\ \log\ \phi_{ks}+\right.\right.\\ &\left.\left.\left\{\omega^{\phi,2}+\sum_{j,i,s}q(z_{ji}=k)\right\}\ \log(1-\phi_{ks})\right]\right)+\text{const.}\end{aligned}\tag{A.37}$$

The minimum is achieved for $\phi_{ks}\sim\text{Beta}(\tilde{\omega}_{ks}^{\phi,1},\tilde{\omega}_{ks}^{\phi,2})$, with the following parameters:

$$\tilde{\omega}_{ks}^{\phi,1}=\omega^{\phi,1}+\sum_{j,i,s} q(z_{ji}=k)q(x_{jis}=1) \quad \text{(A.38)}$$

$$\tilde{\omega}_{ks}^{\phi,2}=\omega^{\phi,2}+\sum_{j,i,s} q(z_{ji}=k)q(x_{jis}=0) \quad \text{(A.39)}$$

Appendix

Table A.1

Update rules for computing the posterior *q* over the unobserved variables.

$\tilde{\omega}_{jk}^{r}=\exp\left(E[\log\ \alpha]+E[\log\ \upsilon_k]+\sum_{k'<k} E[\log(1-\upsilon_{k'})]\right)$
$E[r_{jk}]=\tilde{\omega}_{jk}^{r}E_z[\Psi(\tilde{\omega}_{jk}^{r}+n_{jk})-\Psi(\tilde{\omega}_{jk}^{r})]$
$\upsilon_k\sim\mathrm{Beta}(\tilde{\omega}_k^{\upsilon,1},\tilde{\omega}_k^{\upsilon,2})$
$\tilde{\omega}_k^{\upsilon,1}=1+\sum_j E[r_{jk}]$
$\tilde{\omega}_k^{\upsilon,2}=E[\gamma]+\sum_{j,k'>k} E[r_{jk'}]$
$\phi_{k,s}\sim\mathrm{Beta}(\tilde{\omega}_{k,s}^{\phi,1},\tilde{\omega}_{k,s}^{\phi,2})$
$\tilde{\omega}_{k,s}^{\phi,1}=\omega^{\phi,1}+\sum_{j,i} q(z_{ji}=k)q(x_{jis}=1)$
$\tilde{\omega}_{k,s}^{\phi,2}=\omega^{\phi,2}+\sum_{j,i} q(z_{ji}=k)q(x_{jis}=0)$
$a_{ji}\sim\mathrm{Normal}\ \left(\tilde{\omega}_{ji}^{a,1}(\tilde{\omega}_{ji}^{a,2})^{-1},(\tilde{\omega}_{ji}^{a,2})^{-1}\right)$
$\tilde{\omega}_{ji}^{a,2}=(\sigma_j^a)^{-1}+E[\lambda_{ji}]\sum_{s,s'} E[x_{jis}x_{jis'}]\mathrm{Tr}\ \left(E[\boldsymbol{h}_j\boldsymbol{h}_j^t]\Xi_{js}^t\Xi_{js'}\right)$
$\tilde{\omega}_{ji}^{a,1}=\mu_j^a\left(\sigma_j^a\right)^{-1}+E[\lambda_{ji}]\sum_s E[x_{jis}]E[\boldsymbol{h}_j]^t\Xi_{js}^t(\boldsymbol{y}_{ji}-\sum_d E[e_{jid}]\boldsymbol{f}_{jd})$
$\lambda_{ji}\sim\mathrm{Gamma}(\tilde{\omega}_{ji}^{\lambda,1},\tilde{\omega}_{ji}^{\lambda,2})$
$\tilde{\omega}_{ji}^{\lambda,1}=\kappa_j+\frac{T_j}{2}$

$$
\begin{aligned}
\tilde{\omega}_{ji}^{\lambda,2}=&\theta_j+\|\boldsymbol{y}_{ji}\|^2+\sum\nolimits_d \left(E[e_{jid}^2]\|\boldsymbol{f}_{jd}\|^2\right.\\
&\left.+\sum\nolimits_{d'\neq d}E[e_{jid}]E[e_{jid'}]\boldsymbol{f}_{jd}^t\boldsymbol{f}_{jd'}\right)\\
&+E[a_{ji}^2]\sum\nolimits_{s,s'}E[x_{jis}x_{jis'}]\mathrm{Tr}\ \left(E[\boldsymbol{h}_j\boldsymbol{h}_j^t]\Xi_{js}^t\Xi_{js'}\right)\\
&+E[a_{ji}]\sum\nolimits_{s,d}E[e_{jid}]E[x_{jis}]\boldsymbol{f}_{jd}^t\Xi_{js}E[\boldsymbol{h}_j]\\
&-\boldsymbol{y}_{ji}^t(\sum\nolimits_d E[e_{jid}]\boldsymbol{f}_{jd}\\
&+E[a_{ji}]\sum\nolimits_s E[x_{jis}]\Xi_{js}E[\boldsymbol{h}_j])
\end{aligned}
$$

$$e_{jid}\sim\mathrm{Normal}\ \left(\tilde{\omega}_{jid}^{e,1}(\tilde{\omega}_{jid}^{e,2})^{-1},(\tilde{\omega}_{jid}^{e,2})^{-1}\right)$$

$$\tilde{\omega}_{jid}^{e,2}=(\sigma_{jd}^e)^{-1}+E[\lambda_{ji}]\|\boldsymbol{f}_{jd}\|^2$$

$$\tilde{\omega}_{jid}^{e,1}=\mu_{jd}^e(\sigma_{jd}^e)^{-1}+E[\lambda_{ji}]\boldsymbol{f}_{jd}^t\ \left(\boldsymbol{y}_{ji}-\sum\nolimits_{d'\neq d}E[e_{jid'}]\boldsymbol{f}_{jd'}-E[a_{ji}]\sum\nolimits_s E[x_{jis}]\Xi_{js}E[\boldsymbol{h}_j]\right)$$

$$h_j\sim\mathrm{Normal}(\Xi_j^{-1}\tilde{\omega}_j^h,\Xi_j^{-1})$$

$$\Xi_j=\boldsymbol{\Lambda}+\sum\nolimits_i E[\lambda_{ji}]\sum\nolimits_{s,s'}E[x_{jis}x_{jis'}]\Xi_{\mathrm{js}'}^t\Xi_{js}$$

$$\tilde{\omega}_j^h=\boldsymbol{\Lambda}\bar{\boldsymbol{h}}+\sum\nolimits_{i,s}E[\lambda_{ji}]E[a_{ji}]E[x_{jis}]\Xi_{js}^t\ (\boldsymbol{y}_{ji}-\sum\nolimits_d E[e_{jid}]\boldsymbol{f}_{jd})$$

References

Aguirre G, Zarahn E, D'Esposito M. The Variability of Human, BOLD Hemodynamic Responses. NeuroImage. 1998; 8(4):360–369. [PubMed: 9811554]

Ances B, Leontiev O, Perthen J, Liang C, Lansing A, Buxton R. Regional differences in the coupling of cerebral blood flow and oxygen metabolism changes in response to activation: implications for BOLD-fMRI. NeuroImage. 2008; 39(4):1510–1521. [PubMed: 18164629]

Antoniak C. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. The Annals of Statistics. 1974; 2(6):1152–1174.

Balslev D, Nielsen F, Frutiger S, Sidtis J, Christiansen T, Svarer C, Strother S, Rottenberg D, Hansen L, Paulson O, Law I. Cluster analysis of activity-time series in motor learning. Human Brain Mapping. 2002; 15(3):135–145. [PubMed: 11835604]

Baumgartner R, Ryner L, Richter W, Summers R, Jarmasz M, Somorjai R. Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis. Magnetic Resonance Imaging. 2000; 18(1):89–94. [PubMed: 10642106]

Baumgartner R, Scarth G, Teichtmeister C, Somorjai R, Moser E. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part I: reproducibility. Journal of Magnetic Resonance Imaging. 1997; 7(6):1094–1108. [PubMed: 9400854]

Baumgartner R, Windischberger C, Moser E. Quantification in functional magnetic resonance imaging: fuzzy clustering vs. correlation analysis. Magnetic Resonance Imaging. 1998; 16(2):115–125. [PubMed: 9508268]

Baune A, Sommer F, Erb M, Wildgruber D, Kardatzki B, Palm G, Grodd W. Dynamical cluster analysis of cortical fMRI activation. NeuroImage. 1999; 9(5):477–489. [PubMed: 10329287]

Beckmann C, Smith S. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Transactions on Medical Imaging. 2004; 23(2):137–152. [PubMed: 14964560]

Beckmann C, Smith S. Tensorial extensions of independent component analysis for multisubject FMRI analysis. NeuroImage. 2005; 25(1):294–311. [PubMed: 15734364]

Biswal B, Ulmer J. Blind source separation of multiple signal sources of fMRI data sets using independent component analysis. Journal of Computer Assisted Tomography. 1999; 23(2):265–271. [PubMed: 10096335]

Bullmore E, Long C, Suckling J, Fadili J, Calvert G, Zelaya F, Carpenter T, Brammer M. Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domains. John Wiley & Sons. 2001; 12(2):61–78.

Burock M, Dale A. Estimation and detection of event-related fMRI signals with temporally correlated noise: A statistically efficient and unbiased approach. Human Brain Mapping. 2000; 11(4):249–260. [PubMed: 11144754]

Chuang K, Chiu M, Lin C, Chen J. Model-free functional MRI analysis using Kohonen clustering neural network and fuzzy C-means. IEEE Transactions on Medical Imaging. 1999; 18(12):1117–1128. [PubMed: 10695525]

Cox R, Jesmanowicz A. Real-time 3D image registration for functional MRI. Magnetic Resonance in Medicine. 1999; 42(6):1014–1018. [PubMed: 10571921]

Dale A. Optimal experimental design for event-related fMRI. Human Brain Mapping. 1999; 8(2–3):109–114. [PubMed: 10524601]

Epstein R, Parker W, Feiler A. Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. Journal of Neuroscience. 2007; 27(23):6141. [PubMed: 17553986]

Fadili M, Ruan S, Bloyet D, Mazoyer B. A multistep unsupervised fuzzy clustering analysis of fMRI time series. Human Brain Mapping. 2000; 10(4):160–178. [PubMed: 10949054]

Filzmoser P, Baumgartner R, Moser E. A hierarchical clustering method for analyzing functional MR images. Magnetic resonance imaging. 1999; 17(6):817–826. [PubMed: 10402588]

Friston, K.; Ashburner, J.; Kiebel, S.; Nichols, T.; Penny, W., editors. Statistical Parametric Mapping: the Analysis of Functional Brain Images. Elsevier: Academic Press; 2007.

Golay X, Kollias S, Stoll G, Meier D, Valavanis A, Boesiger P. A new correlation-based fuzzy logic clustering algorithm for fMRI. Magnetic Resonance in Medicine. 1998; 40(2):249–260. [PubMed: 9702707]

Golland, P.; Golland, Y.; Malach, R. Detection of spatial activation patterns as unsupervised segmentation of fMRI data. Proceedings of MICCAI: International Conference on Medical Image Computing and Computer Assisted Intervention; Springer; 2007. p. 110-118.volume 4791 of *LNCS*

Golland Y, Golland P, Bentin S, Malach R. Data-driven clustering reveals a fundamental subdivision of the human cortex into two global systems. Neuropsychologia. 2008; 46(2):540–553. [PubMed: 18037453]

Goutte C, Hansen L, Liptrot M, Rostrup E. Feature-space clustering for fMRI meta-analysis. Human Brain Mapping. 2001; 13(3):165–183. [PubMed: 11376501]

Goutte C, Toft P, Rostrup E, Nielsen F, Hansen L. On clustering fMRI time series. NeuroImage. 1999; 9(3):298–310. [PubMed: 10075900]

Greve D, Fischl B. Accurate and robust brain image alignment using boundary-based registration. NeuroImage. 2009; 48(1):63–72. [PubMed: 19573611]

Grill-Spector K, Malach R. The human visual cortex. Annual Review of Neuroscience. 2004; 27:649–677.

Handwerker D, Ollinger J, D'Esposito M. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage. 2004; 21(4):1639–1651. [PubMed: 15050587]

Jarmasz M, Somorjai R. EROICA: exploring regions of interest with cluster analysis in large functional magnetic resonance imaging data sets. Concepts in Magnetic Resonance Part A. 2003; 16A(1):50–62.

Jbabdi S, Woolrich M, Behrens T. Multiple-subjects connectivity-based parcellation using hierarchical dirichlet process mixture models. NeuroImage. 2009; 44(2):373–384. [PubMed: 18845262]

Kanwisher, N. The ventral visual object pathway in humans: evidence from fMRI. In: Chalupa, L.; Wener, J., editors. The Visual Neurosciences. MIT Press; 2003. p. 1179-1189.

Kanwisher N. Functional specificity in the human brain: A window into the functional architecture of the mind. Proceedings of the National Academy of Science. 2010; 107(25):11163.

Kanwisher N, Yovel G. The fusiform face area: a cortical region specialized for the perception of faces. Philosophical transactions of the Royal Society. Series B, Biological Sciences. 2006; 361(1476):2109–2128.

Kim S, Smyth P. Hierarchical Dirichlet processes with random effects. Advances in Neural Information Processing Systems. 2007; 19:697–704.

Kuhn H. The Hungarian Method for the assignment problem. Naval Research Logistics Quarterly. 1955; 2:83–97.

Lange T, Roth V, Braun M, Buhmann J. Stability-based validation of clustering solutions. Neural Computation. 2004; 16(6):1299–1323. [PubMed: 15130251]

Lashkari, D.; Sridharan, R.; Vul, E.; Hsieh, P.; Kanwisher, N.; Golland, P. Nonparametric hierarchical Bayesian model for functional brain parcellation. Proceedings of MMBIA: IEEE Computer Society Work-shop on Mathematical Methods in Biomedical Image Analysis; IEEE; 2010a. p. 15-22.

Lashkari D, Vul E, Kanwisher N, Golland P. Discovering structure in the space of fMRI selectivity profiles. NeuroImage. 2010b; 50(3):1085–1098. [PubMed: 20053382]

Makni S, Ciuciu P, Idier J, Poline J. Joint detection-estimation of brain activity in functional MRI: a multichannel deconvolution solution. IEEE Transactions on Signal Processing. 2005; 53(9):3488–3502.

Makni S, Idier J, Vincent T, Thirion B, Dehaene-Lambertz G, Ciuciu P. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. NeuroImage. 2008; 41(3):941–969. [PubMed: 18439839]

Marrelec, G.; Benali, H.; Ciuciu, P.; Poline, J-B. Bayesian estimation of the hemodynamic response function in functional MRI. Bayesian Inference and Maximum Entropy Methods Workshop; IOP Institute of Physics; 2002. p. 229-247.volume 617 of *AIP*

McKeown M, Makeig S, Brown G, Jung T, Kindermann S, Bell A, Sejnowski T. Analysis of fMRI data by blind separation into independent spatial components. Human Brain Mapping. 1998; 6(3): 160–188. [PubMed: 9673671]

McKeown M, Sejnowski T. Independent component analysis of fMRI data: examining the assumptions. Human Brain Mapping. 1998; 6(5–6):368–372. [PubMed: 9788074]

Miezin F, Maccotta L, Ollinger J, Petersen S, Buckner R. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. Neuroimage. 2000; 11(6):735–759. [PubMed: 10860799]

Minka T. Automatic choice of dimensionality for PCA. Advances in Neural Information Processing Systems. 2001:598–604.

Moser E, Baumgartner R, Barth M, Windischberger C. Explorative signal processing in functional MR imaging. International Journal of Imaging Systems and Technology. 1999; 10(2):166–176.

Moser E, Diemling M, Baumgartner R. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part II: quantification. Journal of Magnetic Resonance Imaging. 1997; 7(6)

Pitman J. Poisson–Dirichlet and GEM invariant distributions for split-and-merge transformations of an interval partition. Combinatorics, Probability and Computing. 2002; 11(5):501–514.

Schacter D, Buckner R, Koutstaal W, Dale A, Rosen B. Late Onset of Anterior Prefrontal Activity during True and False Recognition: An Event-Related fMRI Study. NeuroImage. 1997; 6(4):259–269. [PubMed: 9417969]

Schwarz G. Estimating the dimension of a model. The Annals of Statistics. 1978; 6(2):461–464.

Smith S, Beckmann C, Ramnani N, Woolrich M, Bannister P, Jenkinson M, Matthews P, McGonigle D. Variability in fMRI: A reexamination of inter-session differences. Human Brain Mapping. 2005; 24(3):248–257. [PubMed: 15654698]

Spiridon M, Fischl B, Kanwisher N. Location and spatial profile of category-specific regions in human extrastriate cortex. Human Brain Mapping. 2006; 27(1):77–89. [PubMed: 15966002]

Teh Y, Jordan M, Beal M, Blei D. Hierarchical dirichlet processes. Journal of the American Statistical Association. 2006; 101(476):1566–1581.

Teh Y, Kurihara K, Welling M. Collapsed variational inference for HDP. Advances in Neural Information Processing Systems. 2008; 20:1481–1488.

Teh Y, Newman D, Welling M. A collapsed variational bayesian inference algorithm for latent dirichlet allocation. Advances in Neural Information Processing Systems. 2007; 19:1353–1360.

Thirion, B.; Faugeras, O. Feature detection in fMRI data: the information bottleneck approach. Proceedings of MICCAI: International Conference on Medical Image Computing and Computer Assisted Intervention; Springer; 2003. p. 83-91.volume 2879 of *LNCS*

Thirion B, Faugeras O. Feature characterization in fMRI data: the Information Bottleneck approach. Medical Image Analysis. 2004; 8(4):403–419. [PubMed: 15567705]

Thirion B, Pinel P, Mériaux S, Roche A, Dehaene S, Poline J. Analysis of a large fMRI cohort: statistical and methodological issues for group analyses. NeuroImage. 2007a; 35(1):105–120. [PubMed: 17239619]

Thirion, B.; Tucholka, A.; Keller, M.; Pinel, P.; Roche, A.; Mangin, J.; Poline, J. High level group analysis of FMRI data based on Dirichlet process mixture models. Proceedings of IPMI: International Conference on Information Processing in Medical Imaging; Springer; 2007b. p. 482-494.volume 4584 of *LNCS*

Tukey, J. Exploratory data analysis. Addison-Wesley; 1977.

Wang, X.; Grimson, W.; Westin, C. Tractography segmentation using a hierarchical dirichlet processes mixture model. Proceedings of IPMI: International Conference on Information Processing in Medical Imaging; Springer; 2009. p. 101-113.volume 5636 of *LNCS*

Wei X, Yoo S, Dickey C, Zou K, Guttmann C, Panych L. Functional MRI of auditory verbal working memory: long-term reproducibility analysis. NeuroImage. 2004; 21(3):1000–1008. [PubMed: 15006667]

Woolrich M, Ripley B, Brady M, Smith S. Temporal autocorrelation in univariate linear modeling of FMRI data. NeuroImage. 2001; 14(6):1370–1386. [PubMed: 11707093]

> We provide a nonparametric hierarchical Bayesian for fMRI data across subjects. > Inference based on the model learns patterns of functional specificity from data. > The model estimates the number, activation profile, and spatial extent of patterns. > Results in a visual fMRI study agree with prior findings and suggest novel patterns.

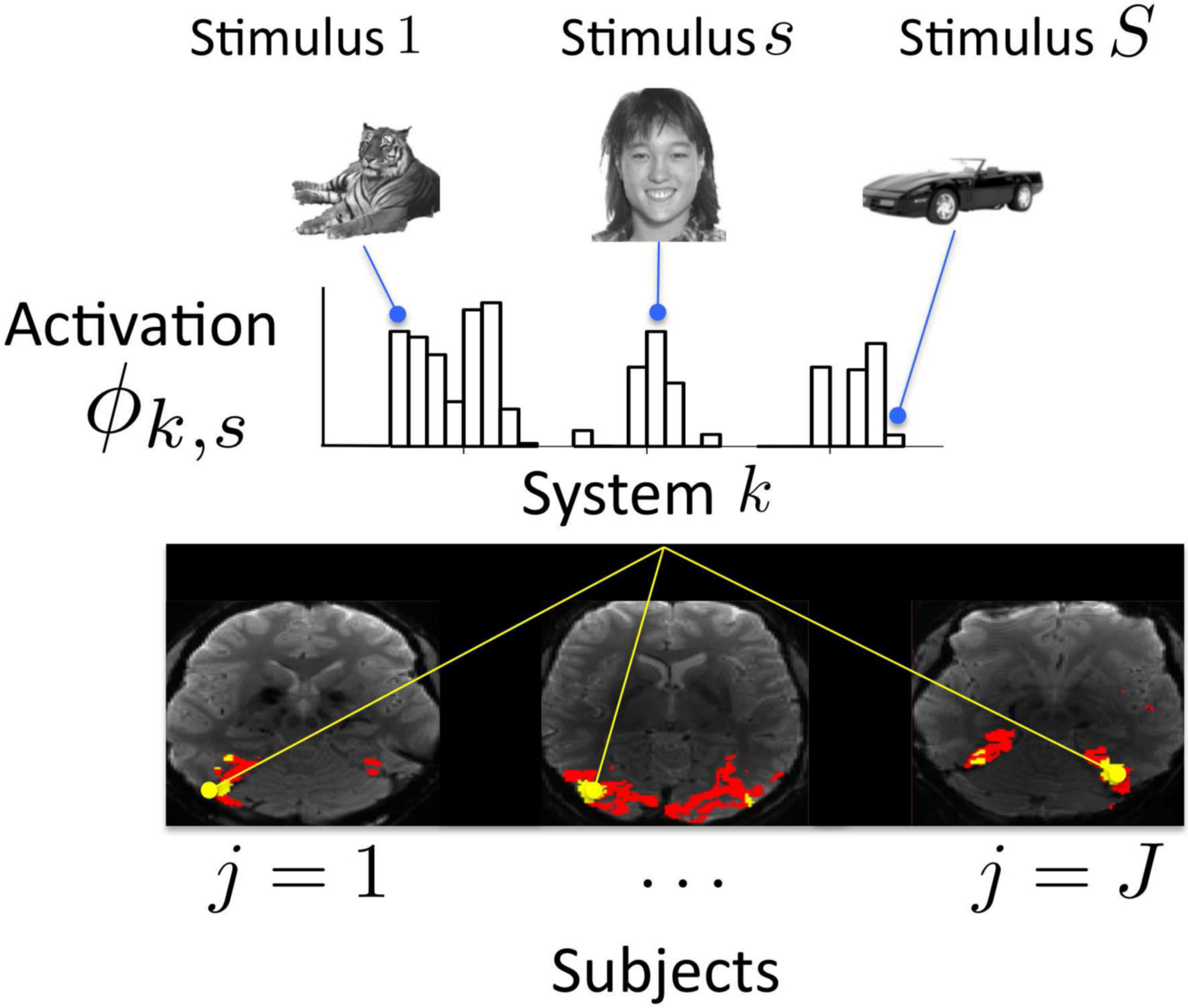

Fig. 1.

Schematic diagram illustrating the concept of a system. System *k* is characterized by vector $[\varphi_{k1}, \cdots, \varphi_{kS}]^t$ that specifies the level of activation induced in the system by each of the *S* stimuli. This system describes a pattern of response demonstrated by collections of voxels in all *J* subjects in the group.

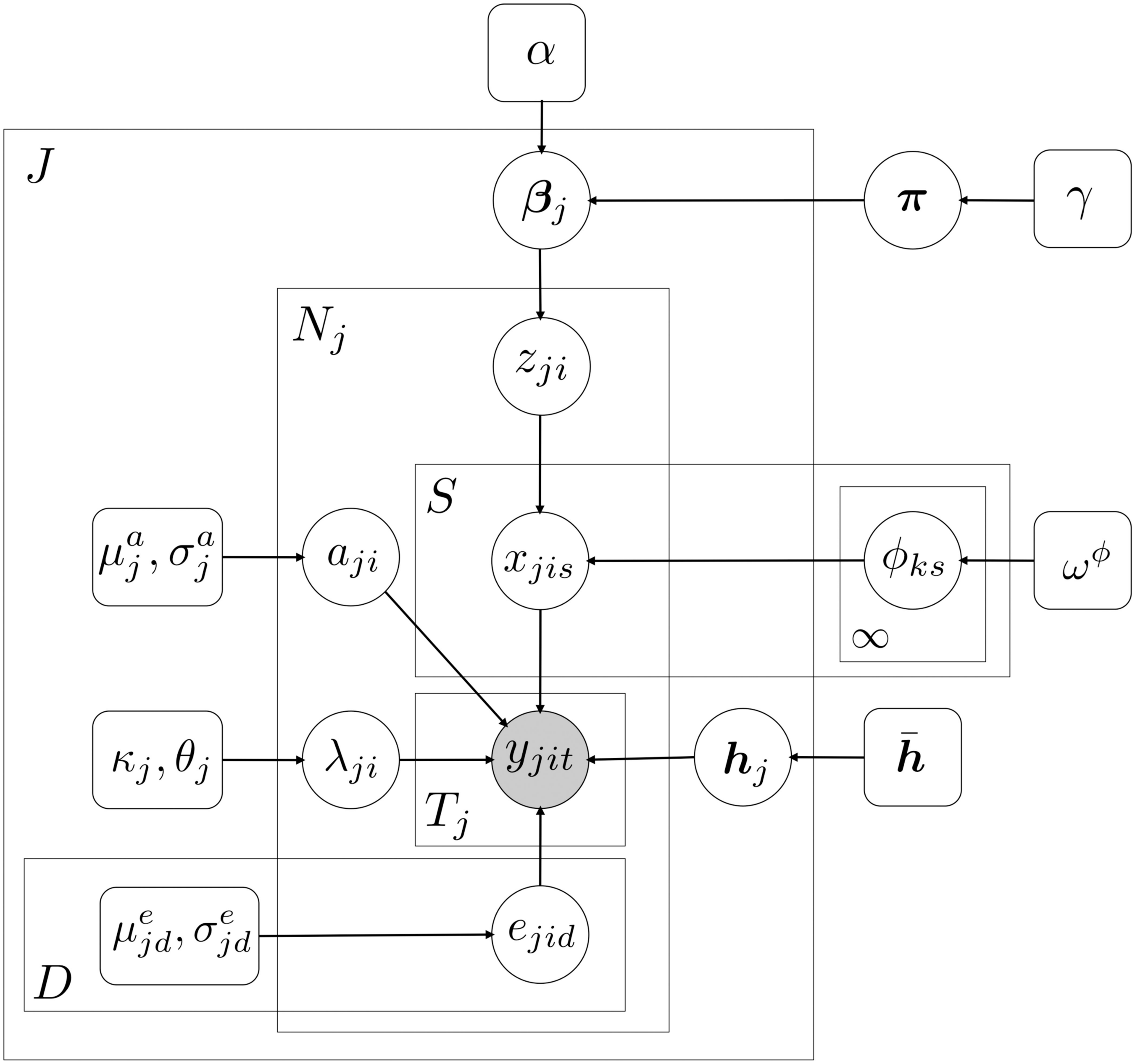

Fig. 2.

Full graphical model that expresses dependencies among latent and observed variables across subjects. Circles and squares indicate random variables and model parameters, respectively. Observed variables are shaded. For a description of different variables, see Table 1.

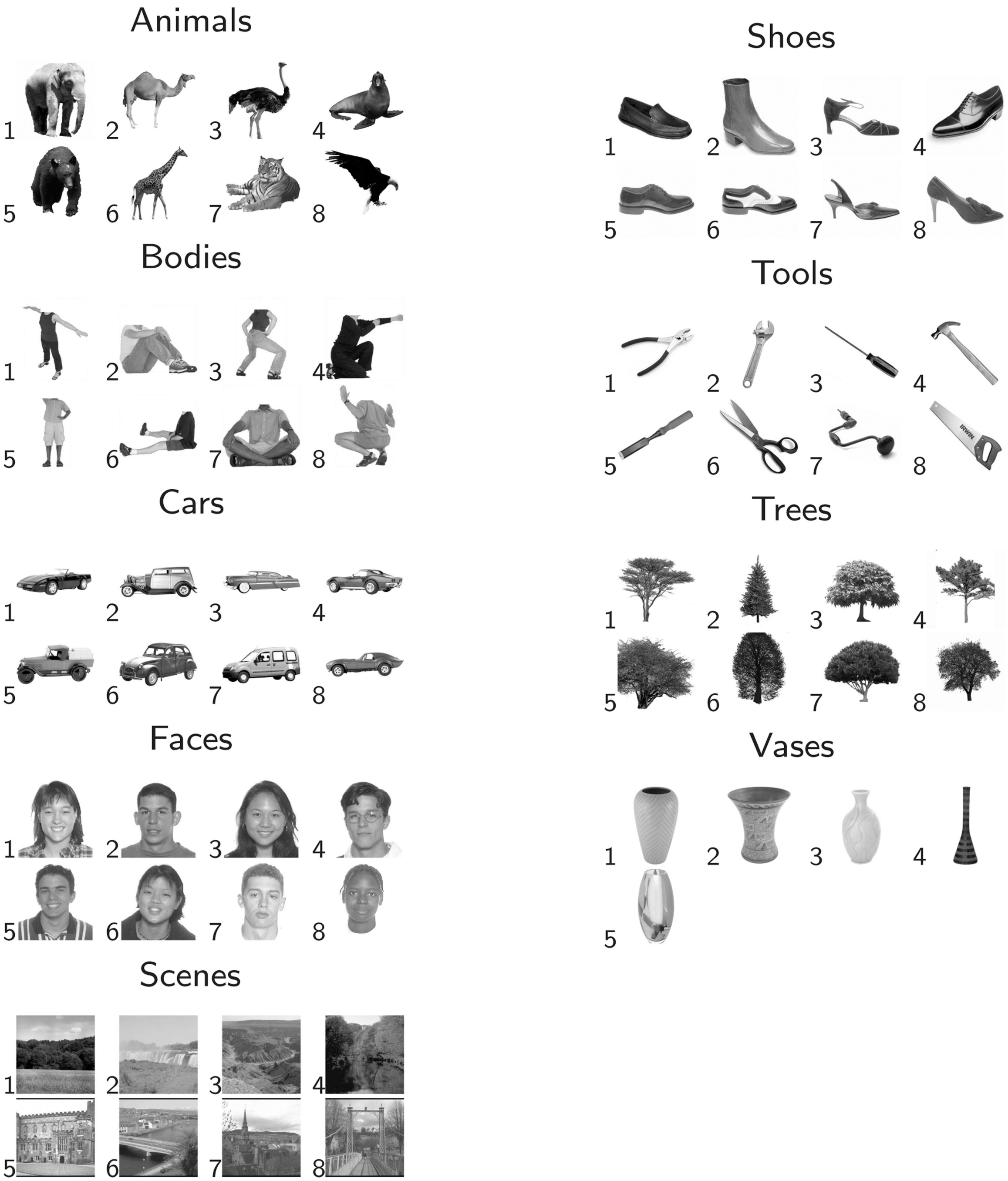

Fig. 3.

The 69 images used as stimuli in the experiment.

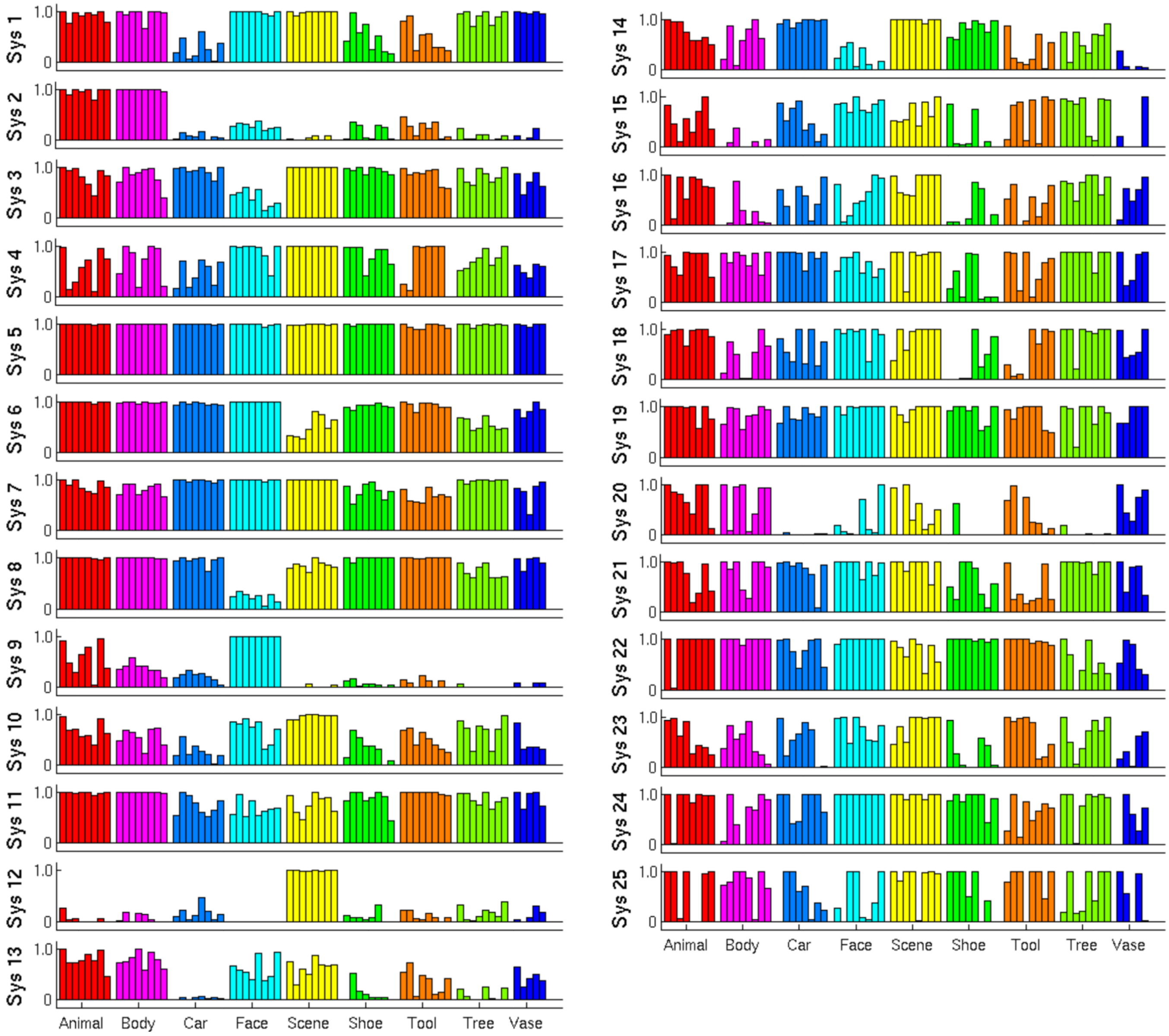

Fig. 4.

System profiles of posterior probabilities of activation for each system to different stimuli. The bar height correspond to the posterior probability of activation.

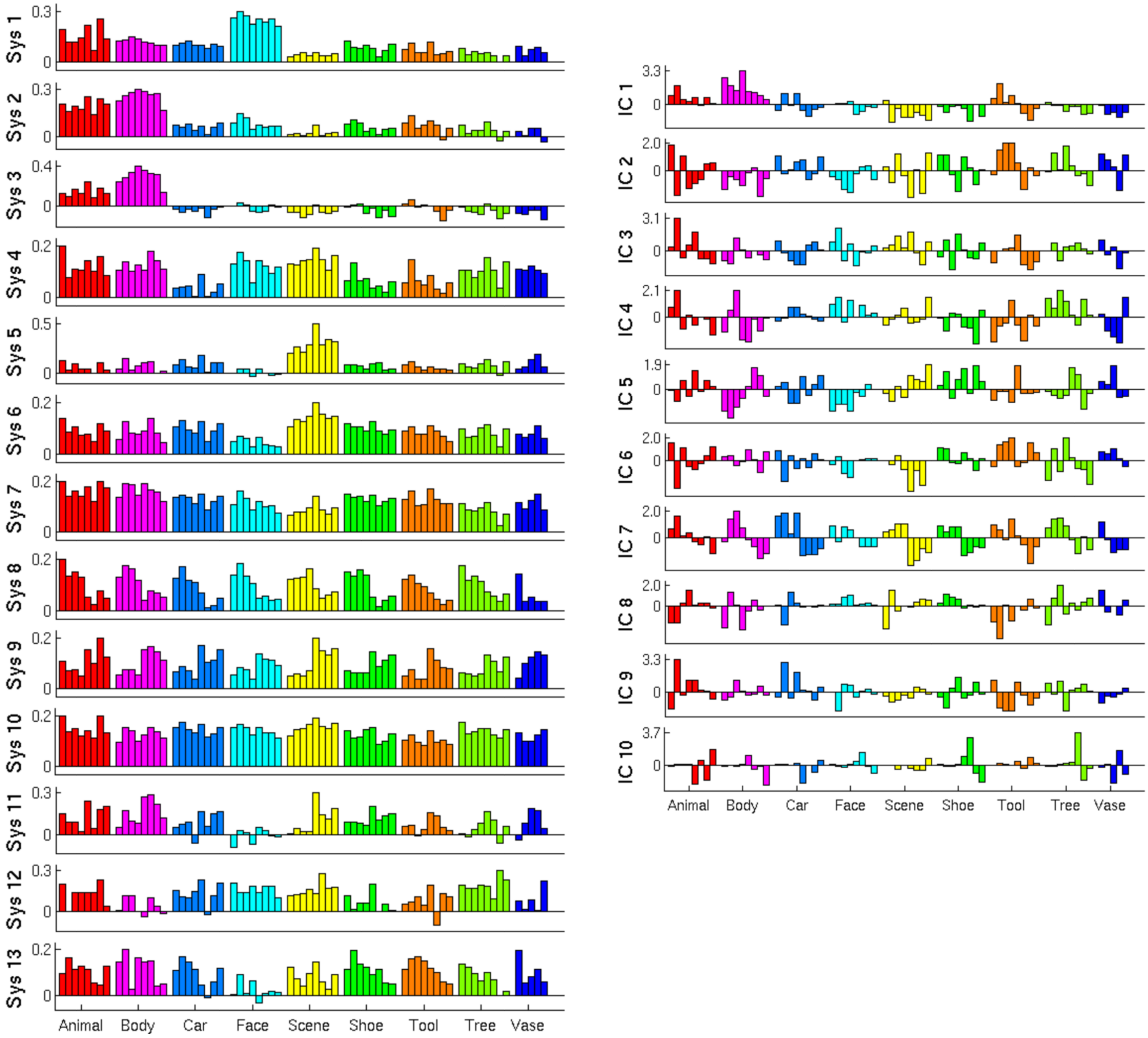

Fig. 5.

Left: system selectivity profiles estimated by the finite mixture of functional systems (Lashkari et al., 2010b). The bar height corresponds to the value of components of normalized selectivity profiles. Right: profiles of independent components found by the tensorial group ICA (Beckmann and Smith, 2005). The bar height corresponds to the value of the independent components. Both sets of profiles are defined in the space of the 69 stimuli.

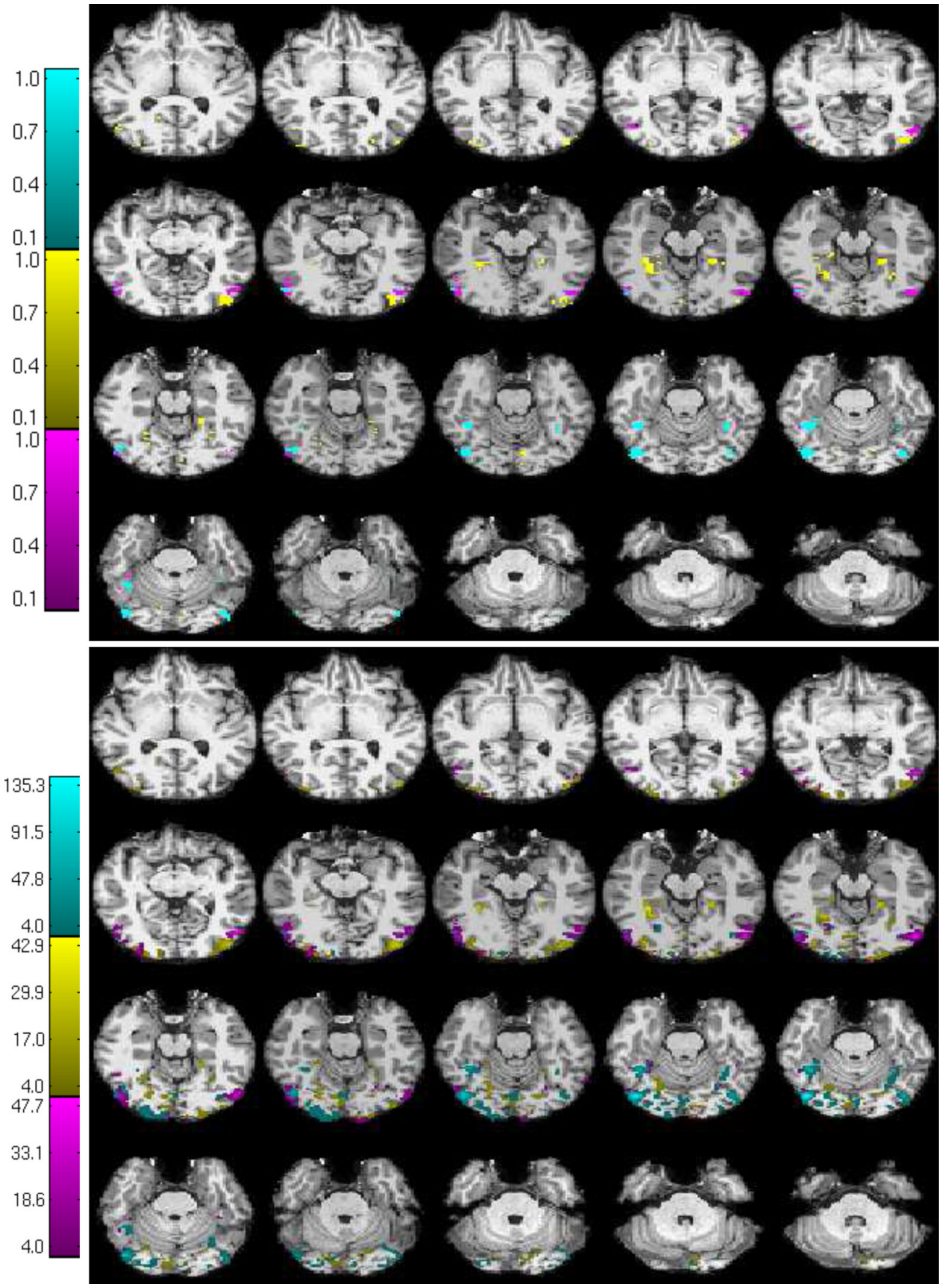

Fig. 6.

Top: membership probability maps corresponding to systems 2, 9, and 12, selective respectively for bodies (magenta), scenes (yellow), and faces (cyan) in one subject. Bottom: map representing significance values $-\log_{10} p$ for three contrasts bodies-objects (magenta), faces-objects (cyan), and scenes-objects (yellow) in the same subject.

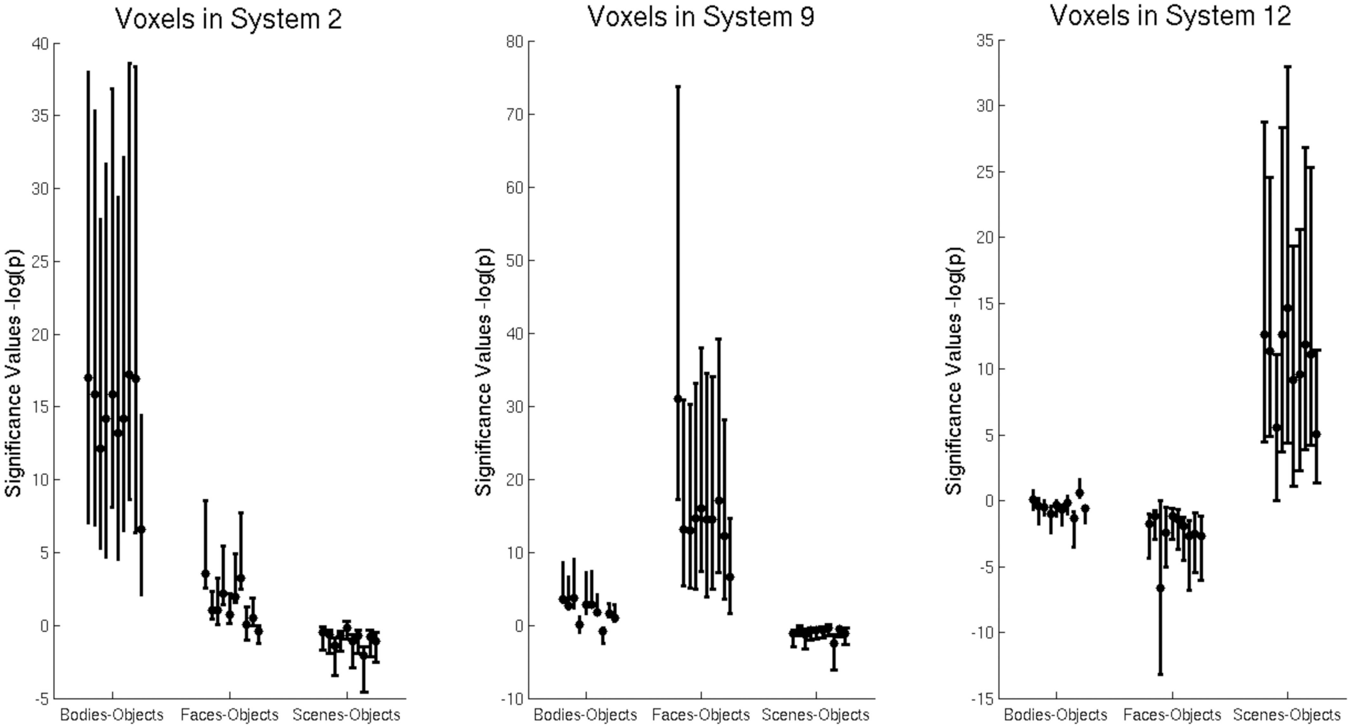

Fig. 7.

The distributions of significance values across voxels in systems 2, 9, and 12 for three different contrasts. For each system and each contrast, the plots report the distribution for each subject separately. The black circle indicates the mean significance value in the area; error bars correspond to 25th and 75th percentiles. Systems 2, 9, and 12 contain voxels with high significance values for bodies, faces, and scenes contrasts, respectively.

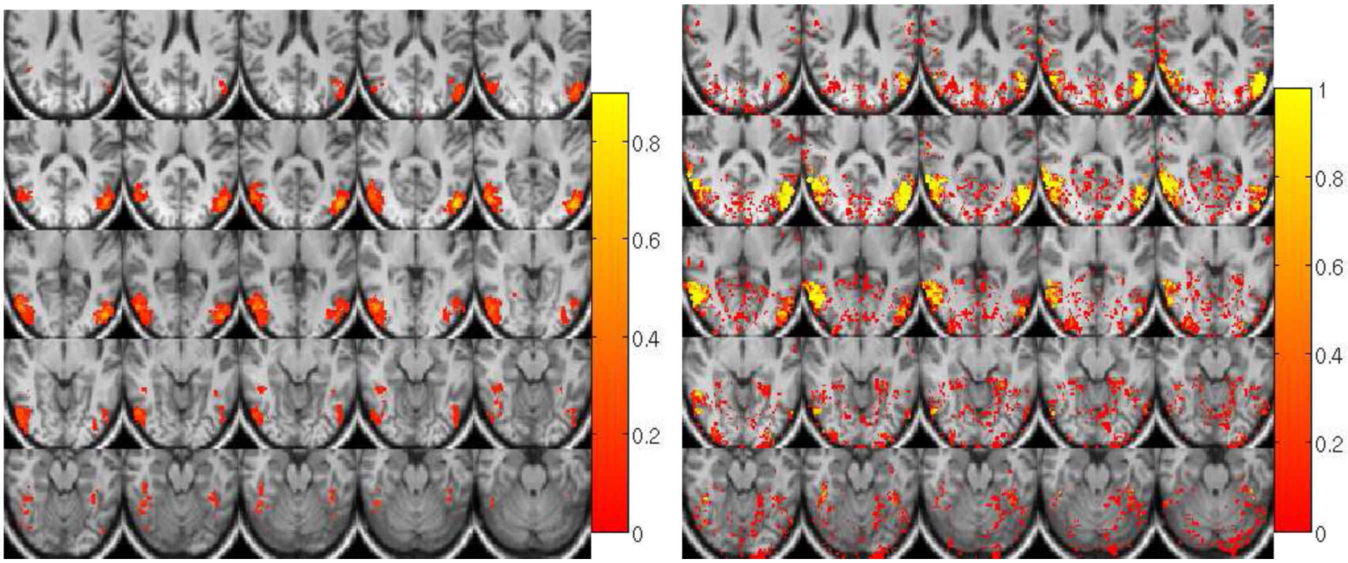

Fig. 8.

Left: the proportion of subjects with voxels in the body-selective system 2 at each location after nonlinear normalization to the MNI template. Right: the group probability map of the body-selective component 1 in the ICA results.

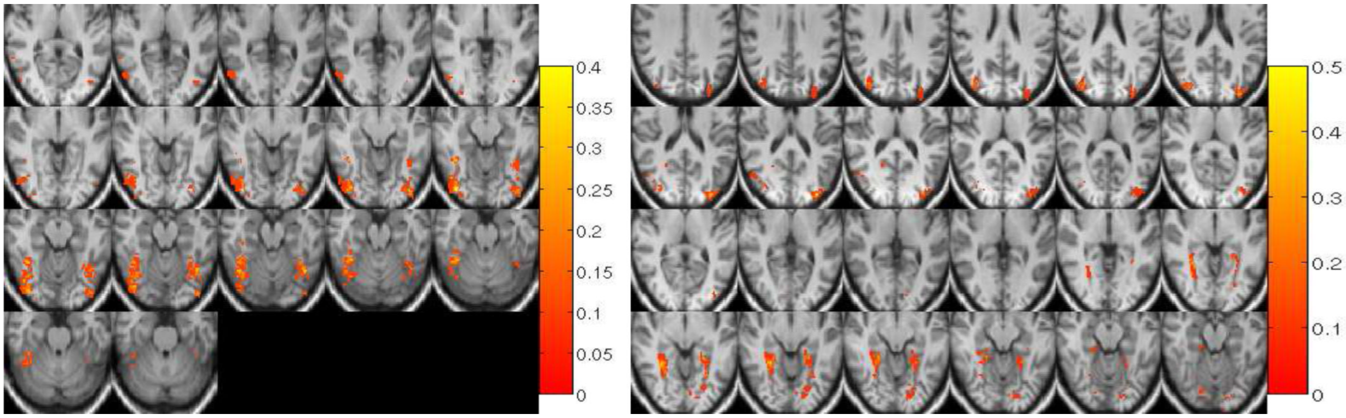

Fig. 9.

Group normalized maps for the face-selective system 9 (left), and the scene-selective system 12 (right).

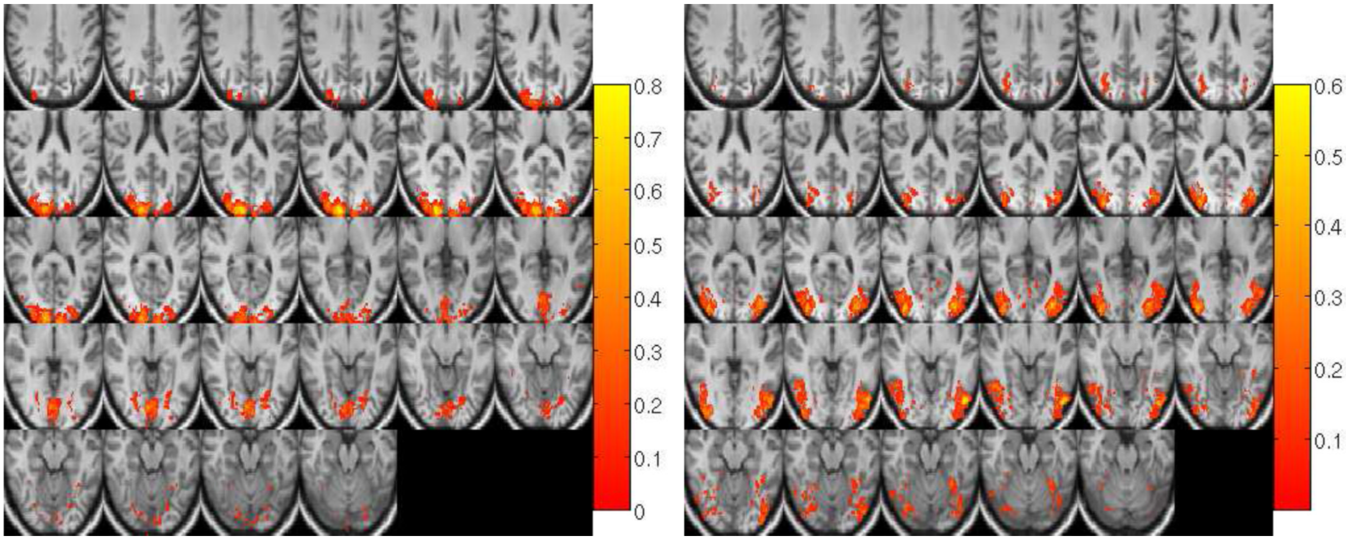

Fig. 10.
Group normalized maps for system 1 (left), and system 8 (right) across all 10 subjects.

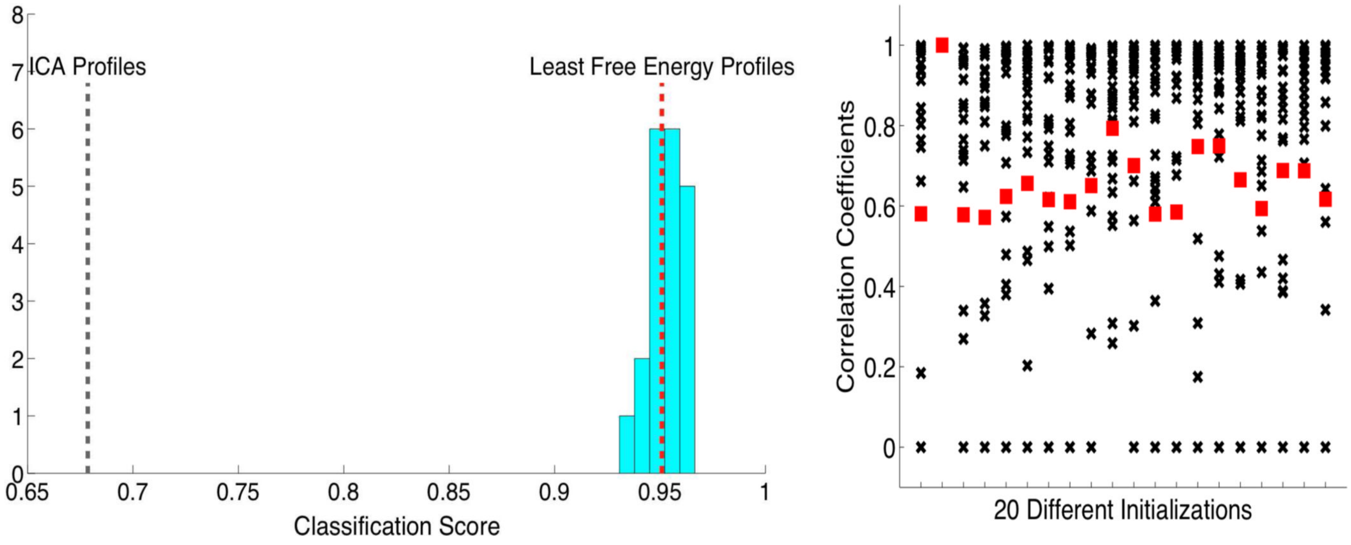

Fig. 11.

(Left) the histogram of classification scores for 20 different initializations of system memberships. The figure denotes the classification scores for the best result based on Gibbs free energy (Figure 4) and tensorial group ICA for comparison. (Right) Consistency scores of all different profiles found by 20 different initializations when matched to the best results in terms of Gibbs free energy. Red squares denote the average consistency score for each initialization.

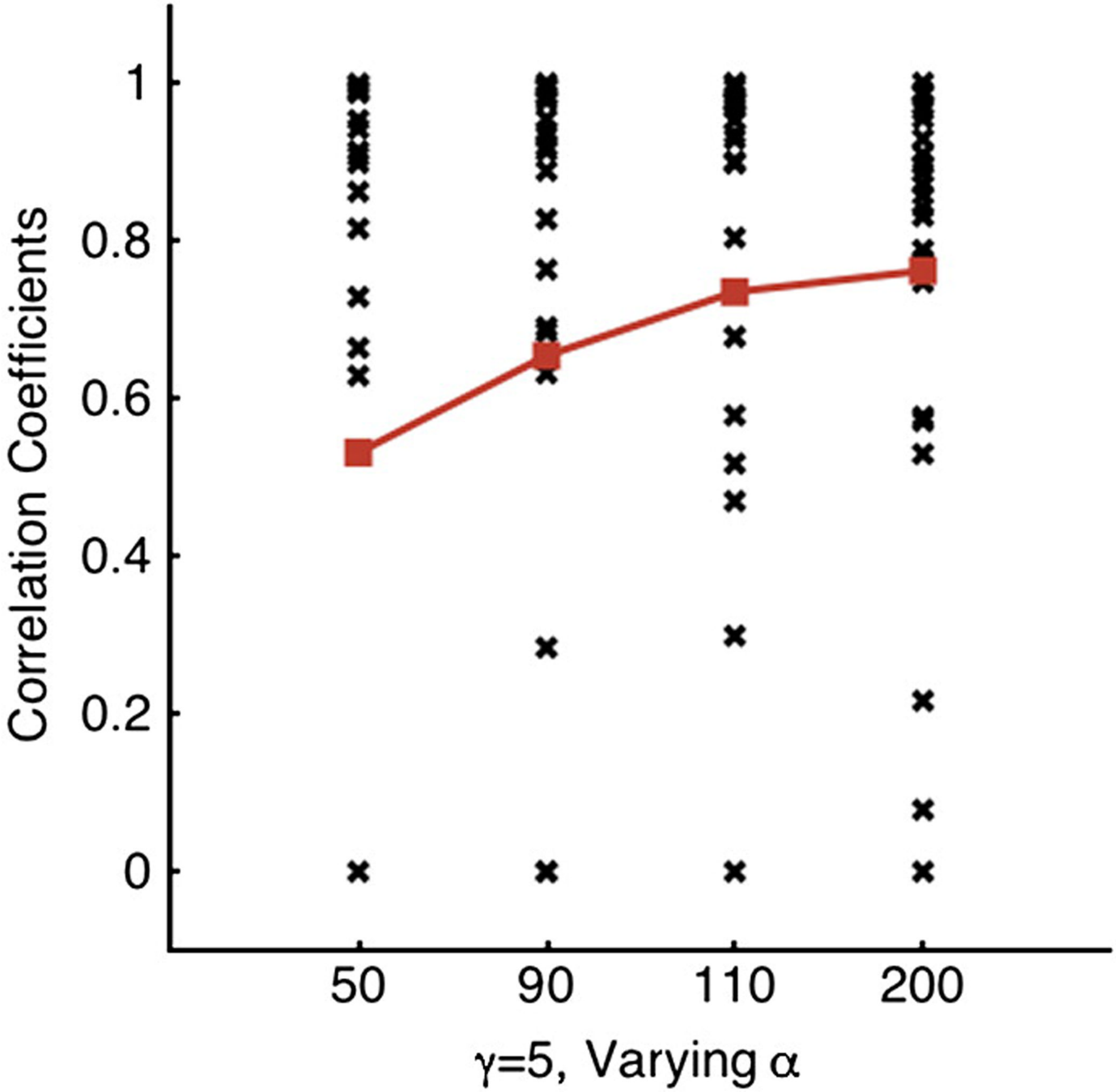

Fig. 12.

Consistency scores for the results found with different HDP scale parameters when matched to the results found with $\gamma = 5$ and $\alpha = 100$. Red squares denote the average correlation coefficients for all profiles found with any given parameter pair.

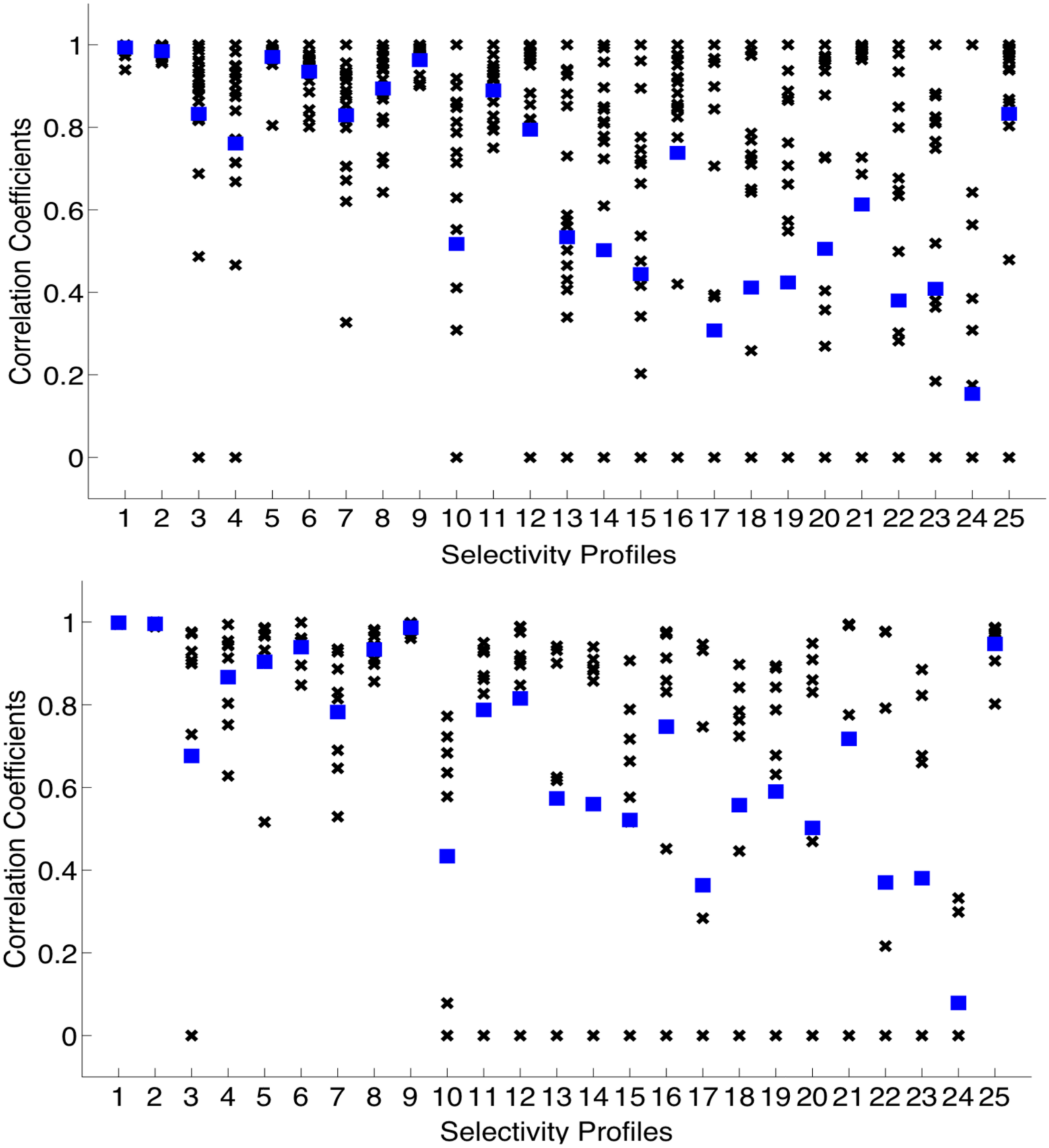

Fig. 13.

(Top) The histogram of classification scores matched to each of the 25 system profiles of Figure 4 for different initializations. (Bottom) The histogram of classification scores matched to each of the 25 system profiles of Figure 4 for different HDP scale parameters. Blue squares denote the average of all consistency scores for each system profile.

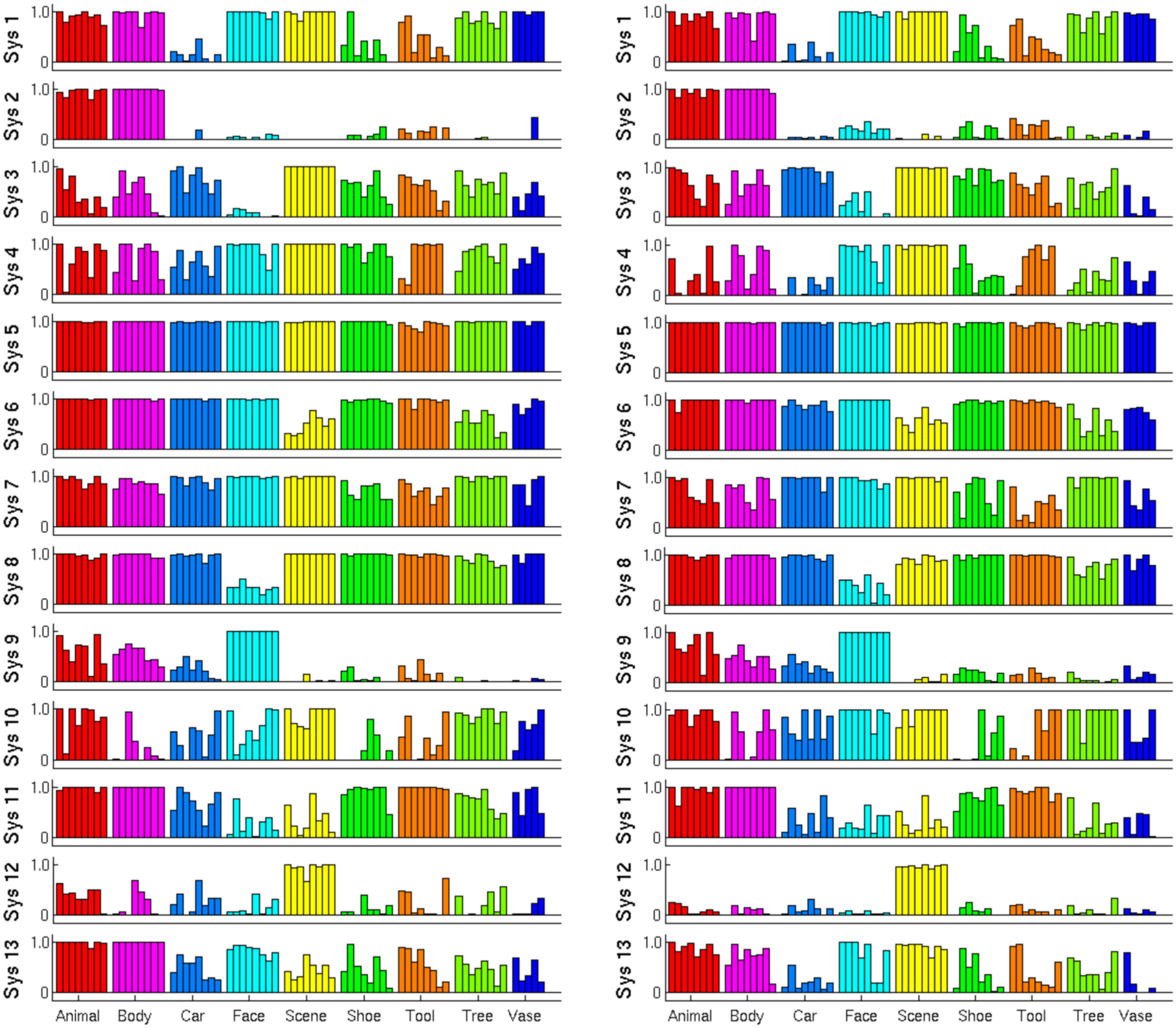

Fig. 14.

System profiles of activation probabilities found by applying the method to two independent sets of 5 subjects. The profiles were first matched across two groups using the scheme described in the text, and then matched with the system profiles found for the entire group (Figure 4).

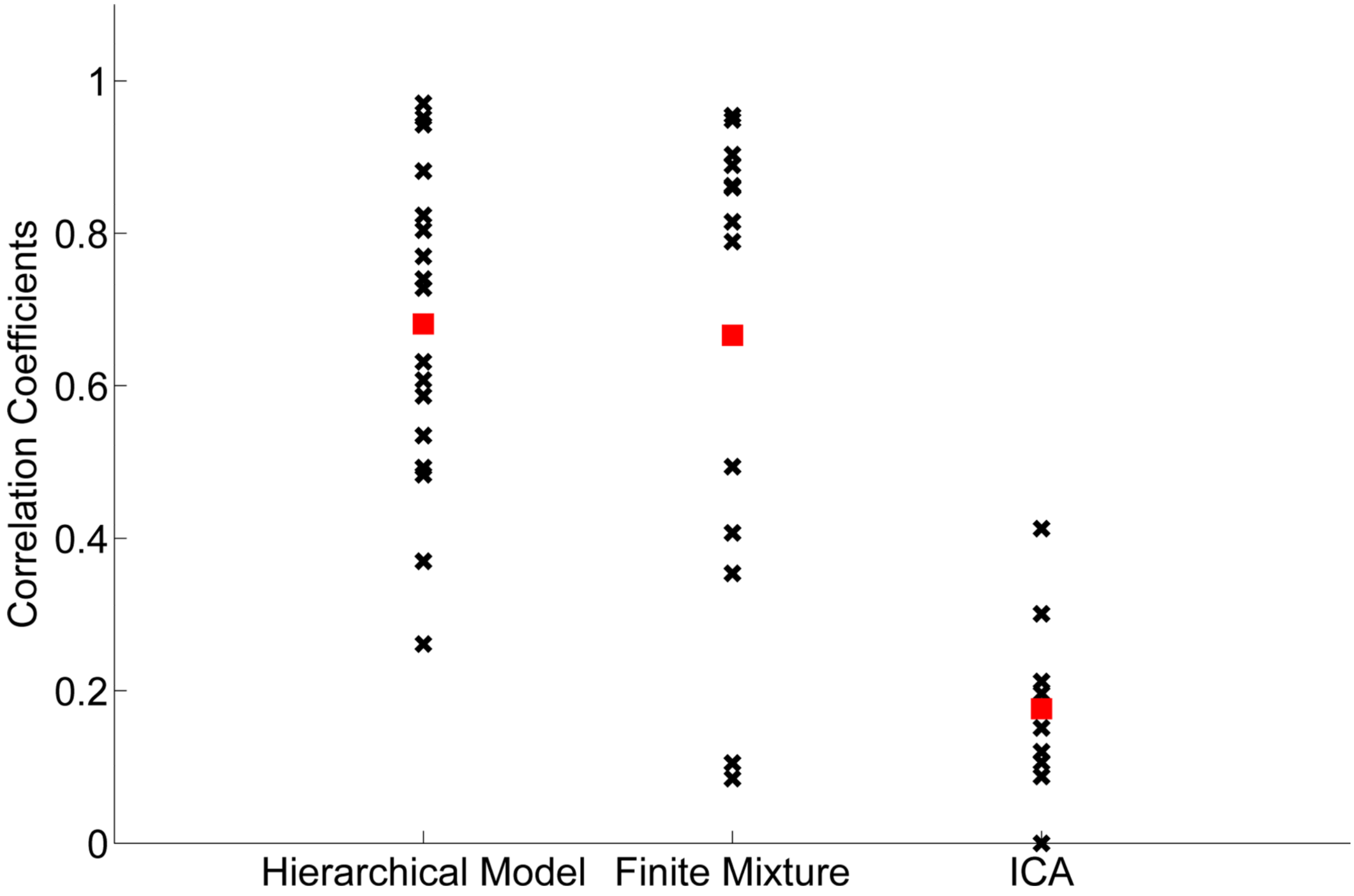

Fig. 15.

The correlation of profiles matched between the results found on the two separate sets of subjects for the three different techniques. Red squares denote the average correlation coefficients for each set of profiles.

Table 1

Variables and parameters in the model.

x_{jis}	binary activation of voxel *i* in subject *j* for stimulus *s*
z_{ji}	multinomial unit membership of voxel *i* in subject *j*
φ_{ks}	activation probability of system *k* for stimulus *s*
$\boldsymbol{\beta}_j$	system prior vector of weights in subject *j*
π_k	group-level prior weight for system *k*
α, γ	HDP scale parameters
y_{jit}	fMRI signal of voxel *i* in subject *j* at time *t*
e_{jid}	nuisance regressor *d* contribution to signal at voxel *i* in subject *j*
a_{ji}	amplitude of activation of voxel *i* in subject *j*
$\boldsymbol{h}_j$	a finite-time HRF vector in subject *j*
λ_{ji}	variance reciprocal of noise for voxel *i* in subject *j*
μ_j^a, σ_j^a	prior parameters for response amplitudes in subject *j*
$\mu_{jd}^e, \sigma_{jd}^e$	prior parameters for nuisance regressor *d* in subject *j*
$\omega^{\varphi,1}$, $\omega^{\varphi,2}$	prior parameters for actviation probabilities φ
κ_j, θ_j	prior parameters for noise variance in subject *j*

Table 2

Number of systems and the resulting classification scores for the results found when varying the HDP scale parameters around the pair (γ, α) = (5, 100).

γ = 5, varying α	50	90	100	110	200
Number of Systems	15	19	25	22	24
Classification Score	0.96 ± 0.15	0.94 ± 0.17	0.95 ± 0.16	0.97 ± 0.13	0.96 ± 0.15
α = 100, varying γ	3	4	5	6	10
Number of Systems	13	20	25	21	34
Classification Score	0.95 ± 0.16	0.96 ± 0.14	0.95 ± 0.16	0.95 ± 0.15	0.95 ± 16