Ann Math Artif Intell (2009) 57:233–248
DOI 10.1007/s10472-010-9196-8

An application of formal concept analysis to semantic neural decoding

Dominik Maria Endres · Peter Földiák · Uta Priss

Published online: 9 July 2010

Abstract This paper proposes a novel application of Formal Concept Analysis (FCA) to neural decoding: the semantic relationships between the neural representations of large sets of stimuli are explored using concept lattices. In particular, the effects of neural code sparsity are modelled using the lattices. An exact Bayesian approach is employed to construct the formal context needed by FCA. This method is explained using an example of neurophysiological data from the high-level visual cortical area STSa. Prominent features of the resulting concept lattices are discussed, including indications for hierarchical face representation and a product-of-experts code in real neurons. The robustness of these features is illustrated by studying the effects of scaling the attributes.

Keywords Formal concept analysis · FCA · Neural code · Sparse coding · High-level vision · STS · Bayesian classification · Semantic · Neural decoding

Mathematics Subject Classifications (2010) 06 · 92 · 62

D. M. Endres (✉)
Section for Theoretical Sensomotorics, Department of Cognitive Neurology,
Hertie Institute for Clinical Brain Research and Center for Integrative Neuroscience,
University Clinic Tübingen, Tübingen, Germany
e-mail: dominik.endres@klinikum.uni-tuebingen.de

P. Földiák
School of Psychology, University of St Andrews, Scotland, UK
e-mail: Peter.Foldiak@st-andrews.ac.uk

U. Priss
School of Computing, Edinburgh Napier University, Edinburgh, UK
e-mail: u.priss@napier.ac.uk

1 Introduction

Mammalian brains consist of billions of neurons, each capable of independent electrical activity. From an information-theoretic perspective, the patterns of activation or response of these neurons can be understood as the codewords comprising the neural code. The neural code describes which pattern of activity corresponds to what information item. We are interested in the (high-level) visual system, where such items may indicate the presence of a stimulus object or the value of some stimulus attribute, assuming that each time this item is represented the neural activity pattern will be the same or at least similar. *Neural decoding* is the attempt to reconstruct the stimulus from the observed pattern of activation in a given population of neurons [1–4]. Popular decoding quality measures, such as Fisher's linear discriminant [5] or mutual information [6] capture how accurately a stimulus can be determined from a neural activity pattern (e.g., [4]). While useful, these measures provide little information about the structure of the neural code, which is what we are concerned with here. Furthermore, we would also like to elucidate how this structure relates to the represented information items, i.e., we are interested in the semantic aspects of the neural code.

To explore the relationship between the representations of related items, Földiák [7] demonstrated that a sparse neural code can be interpreted as a graph (a kind of "semantic net"). Each codeword can then be represented as a set of active units (a subset of all units). The codewords can now be partially ordered under set inclusion: codeword $A \leq$ codeword B iff the set of active neurons of A is a subset of the active neurons of B. This ordering relation is capable of capturing semantic relationships between the represented information items. There is a duality between the information items and the sets representing them: a more general class corresponds to a smaller subset of active neurons, and more specific items are represented by larger sets [7]. Additionally, storing codewords as sets is especially efficient for sparse codes, i.e., codes with a low activity ratio (i.e., few active units in each codeword). Here, we apply Formal Concept Analysis (FCA) [8, 9] to these data because this duality can be interpreted as a Galois connection. The resulting concept lattices are an interesting representation of the relationships implicit in the code.

We would also like to be able to represent how the relationship between sets of active neurons translates into the corresponding relationship between the encoded stimuli. In our application, the stimuli are the formal objects, and the neurons are the formal attributes. The FCA approach exploits the duality of extensional and intensional descriptions and allows visual exploration of the data in lattice diagrams. FCA has shown to be useful for data exploration and knowledge discovery in numerous applications in a variety of fields [10, 11].

For the benefit of the FCA community, we provide more details on sparse coding in Section 2. As FCA is not (yet) a standard analysis technique in neuroscience, we also provide a short introduction to FCA in Section 3. A full account can be found in [9]. We demonstrate how the sparseness (or denseness) of the neural code affects the structure of the concept lattice in Section 4. Section 5 describes the generative classifier model which we use to build the formal context from the responses of

neurons in the high-level visual cortex of monkeys. Finally, we discuss the concept lattices so obtained in Section 6.

2 Sparse coding

One feature of neural codes which has attracted a considerable amount of interest is its *sparseness*. As detailed in [12], sparse coding is to be distinguished from local and dense distributed coding. At one extreme of low average activity ratio are local codes, in which each item is represented by a separate neuron or a small set of neurons. This way there is no overlap between the representations of any two items in the sense that no neuron takes part in the representation of more than one item. An analogy might be the coding of characters on a computer keyboard (without the Shift and Control keys), where each key encodes a single character. It should be noted that locality of coding does not necessarily imply that only one neuron encodes an item, it only says that the neurons are highly selective, corresponding to single significant items of the environment (e.g. a "grandmother cell"—a hypothetical neuron that has the exact and only purpose to be activated when someone sees, hears or thinks about their grandmother).

The other extreme (approximate average activity ratio of 0.5) corresponds to dense, or *holographic* coding. Here, an information item is represented by the combination of activities of all neurons. For N binary neurons this implies a representational capacity of 2^N. Given the billions of neurons in a human brain, 2^N is beyond astronomical. As the number of neurons in the brain (or even just in a single cortical area, such as the primary visual cortex) is substantially higher than the number of receptor cells (e.g. in the retina), the representational capacity of a dense code in the brain is much greater than what we can experience in a lifetime. Therefore the greatest part of this capacity is redundant.

Sparse codes (small average activity ratio) appear to be a favourable compromise between dense and local codes [13, 14]. The small representational capacity of local codes can be remedied with a modest fraction of active units per pattern because representational capacity grows exponentially with average activity ratio (for small average activity ratios). Thus, distinct items are much less likely to interfere when represented simultaneously. Furthermore, it is much more likely that a single layer network can learn to generate a target output if the input has a sparse representation. This is due to the higher proportion of mappings being implementable by a linear discriminant function. Learning in single layer networks is therefore simpler, faster and substantially more plausible in terms of biological implementation. By controlling sparseness, the amount of redundancy necessary for fault tolerance can be chosen. With the right choice of code, a relatively small amount of redundancy can lead to highly fault-tolerant decoding. For instance, the failure of a small number of units may not make the representation undecodable. Moreover, a sparse distributed code can represent values at higher accuracy than a local code. Such distributed coding is also referred to as coarse coding.

Sparse codes seem to be employed in the mammalian visual system [15] and are well suited to representing the visual environment we live in [16, 17]. It is also possible to define sparseness for graded or even continuous-valued responses (see e.g. [4, 12, 18]).

3 Formal concept analysis

Central to FCA[9] is the notion of the formal context $K := (G, M, I)$, which is comprised of a set of formal objects G, a set of formal attributes M and a binary relation $I \subseteq G \times M$ between members of G and M. In our application, the members of G are visual stimuli, whereas the members of M are the neurons. If neuron $m \in M$ responds when stimulus $g \in G$ is presented, then we write $(g, m) \in I$ or gIm. It is customary to represent the context as a cross table, where the row(column) headings are the object(attribute) names. For each pair $(g, m) \in I$, the corresponding cell in the cross table has an "×". Table 1, left, shows a simple example context.

The prime operator for subsets $A \subseteq G$ is defined as $A' = \{m \in M | \forall g \in A : gIm\}$, i.e., A' is the set of all attributes shared by the objects in A. Likewise, for $B \subseteq M$, B' is defined as $B' = \{g \in G | \forall m \in B : gIm\}$, i.e., B' is the set of all objects having all attributes in B.

Definition 1 ([9]) A **formal concept** of the context K is a pair (A, B) with $A \subseteq G$, $B \subseteq M$ such that $A' = B$ and $B' = A$. A is called the **extent** and B is the **intent** of the concept (A, B). $\mathfrak{B}(K)$ denotes the set of all concepts of the context K.

In other words, given the relation I, (A, B) is a concept if A determines B and vice versa. A and B are sometimes called *closed* subsets of G and M with respect to I. Table 1, right, lists all concepts of the context in Table 1, left. One can visualise the defining property of a concept as follows: if (A, B) is a concept, reorder the rows and columns of the cross table such that all objects in A are in adjacent rows, and all attributes in B are in adjacent columns. The cells corresponding to all $g \in A$ and $m \in B$ then form a rectangular block of "×"s with no empty spaces in between. In the example above, this can be seen (without reordering rows and columns) for

Table 1 *Left*: a simple example context, represented as a cross-table

	n1	n2	n3
monkeyFace	×	×	
monkeyHand		×	
humanFace	×		
spider			×

concept	extent (stimuli)	intent (neurons)
0	*ALL*	*NONE*
1	spider	n3
2	humanFace monkeyFace	n1
3	monkeyFace monkeyHand	n2
4	monkeyFace	n1 n2
5	*NONE*	*ALL*

The objects (rows) are 4 visual stimuli, the attributes (columns) are 3 (hypothetical) neurons n1,n2,n3. An "×" in a cell indicates that a stimulus elicited a response from the corresponding neuron. *Right*: the concepts of this context. Colours correspond to Fig. 1

concepts 1,3,4. For a graphical representation of the relationships between concepts, one defines an order on $\mathbb{B}(K)$:

Definition 2 [9] If (A_1, B_1) and (A_2, B_2) are concepts of a context, (A_1, B_1) is a **subconcept** of (A_2, B_2) if $A_1 \subseteq A_2$ (which is equivalent to $B_1 \supseteq B_2$). In this case, (A_2, B_2) is a **superconcept** of (A_1, B_1) and we write $(A_1, B_1) \leq (A_2, B_2)$. The relation $\leq$ is called the **order** of the concepts.

It can be shown [8, 9] that $\mathbb{B}(K)$ and the concept order form a complete lattice. The concept lattice of the context in Table 1, with full and reduced labelling, is shown in Fig. 1. Full labelling means that a concept node is depicted with its full extent and intent. A reduced labelled concept lattice shows an object only in the smallest (w.r.t. the concept order of Definition 2) concept of whose extent the object is a member. This concept is called the *object concept*, or the concept that *introduces* the object. Likewise, an attribute is shown only in the largest concept of whose intent the attribute is a member, the *attribute concept*, which *introduces* the attribute. The closedness of extents and intents has an important consequence for neuroscientific applications. Adding attributes to M (e.g. responses of additional neurons) will very probably grow $\mathbb{B}(K)$. However, the original concepts will be embedded as a substructure in the larger lattice, with their ordering relationships preserved.

The lattice diagrams make the ordering relationship between the concepts graphically explicit: concept 3 contains all "monkey-related" stimuli, concept 2 encompasses all "faces". They have a common lower neighbour, concept 4, which is the "monkeyFace" concept. The "spider" concept (concept 1) is incomparable to any other concept except the top and the bottom of the lattice. Note that these relationships arise as a consequence of the (here hypothetical) response behaviour of the neurons. We will show (Section 6) that the response patterns of real neurons can lead to similarly interpretable structures.

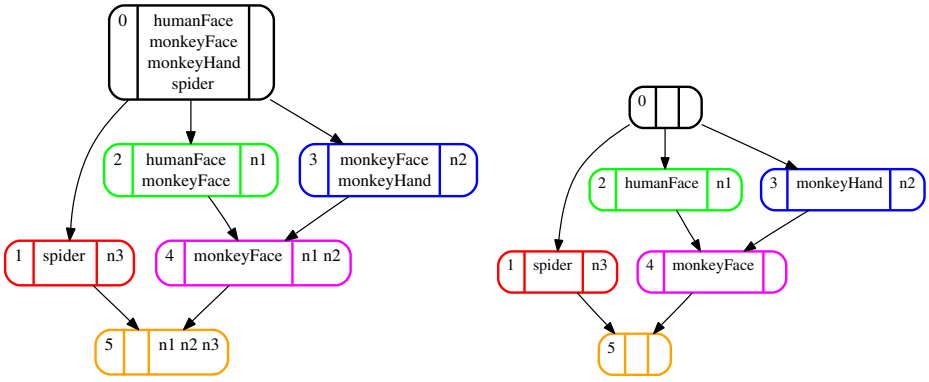


Fig. 1 Concept lattice computed from the context in Table 1. Each node is a concept, arrows represent superconcept relation, i.e., an arrow from X to Y reads: X is an upper neighbour of Y. Colours correspond to Table 1, right. The number in the leftmost compartment is the concept number. Middle compartment contains the extent, rightmost compartment the intent. *Left*: fully labelled concepts, i.e., all members of extents and intents are listed in each concept node. *Right*: reduced labelling. An object/attribute is only listed in the extent/intent of the smallest/largest concept that contains it. Reduced labelling is very useful for drawing large concept lattices

From a decoding perspective, a fully labelled concept shows those stimuli that have activated at least the set of neurons in the intent. In contrast, the stimuli associated with a concept in reduced labelling will activate the set of neurons in the intent, but no others. The fully labelled concepts show stimuli encoded by activity of the active neurons of the concept without knowledge of the firing state of the other neurons. Reduced labels, on the other hand show those stimuli that elicited a response *only* from the neurons in the intent.

4 Concept lattices of local, sparse and dense codes

In the case of a binary neural code, the sparseness of a codeword is inversely related to the fraction of active neurons. The inverse of the average fraction of active neurons across all codewords is the sparseness of the code [12, 14]. To study what structural effects different levels of sparseness would have on a neural code, we generated random codes, i.e., each of 10 stimuli was associated with randomly drawn responses of 10 neurons, subject to the constraints that the code be perfectly decodable and that the sparseness of each codeword was equal to the sparseness of the code. Figure 2 shows the contexts (represented as cross-tables) and the concept lattices

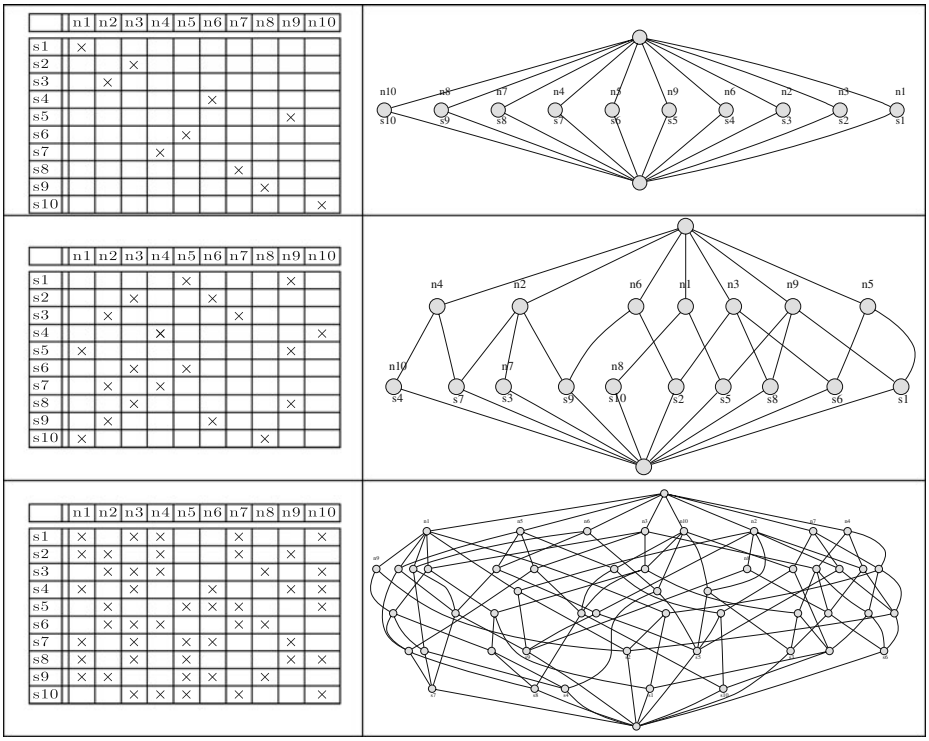


Fig. 2 Contexts (represented as cross-tables) and concept lattices for a local, sparse and dense random neural code with reduced labelling. Each context was built out of the responses of 10 (hypothetical) neurons (n1, ..., n10) to 10 stimuli (s1, ..., s10). Each node represents a concept

of a local code (activity ratio 0.1), a sparse code (activity ratio 0.2) and a dense code (activity ratio 0.5). In a local code, the response patters to different stimuli have no overlapping activations, hence the lattice representing this code is an anti-chain with top and bottom element added. Each concept in the anti-chain introduces (at least) one stimulus and (at least) one neuron. In contrast, a dense code results in a larger number of concepts which introduce neither a stimulus nor a neuron. The lattice of the dense code also contains substantially longer chains between the top and bottom nodes than in the case of sparse and local codes.

The most obvious differences between the lattices is the total number of concepts. A dense code, even for a small number of stimuli, will give rise to a large number of concepts, because the neuron sets representing the stimuli are very probably going to have non-empty intersections. These intersections are potentially the intents of concepts which are larger than those concepts that introduce the stimuli. Hence, the latter are found towards the bottom of the lattice. This implies that they have large intents, which is a consequence of the density of the code. Determining these intents thus requires the observation of a large number of neurons, which is unappealing from a decoding perspective. The local code does not have this drawback, but is hampered by a small encoding capacity (maximal number of concepts with non-empty extents): the concept lattice in Fig. 2 is the largest one which can be constructed for a local code comprised of 10 binary neurons. Which of the above structures is most appropriate depends on the conceptual structure of the environment to be encoded.

5 Building a formal context from responses of high-level visual neurons

To explore whether FCA is a suitable tool for interpreting real neural codes, we constructed formal contexts from the responses of high-level visual cortical cells in area STSa (part of the temporal lobe) of monkeys. Characterising the responses of these cells is a difficult task. They exhibit complex nonlinearities and invariances which make it impossible to apply linear techniques, such as reverse correlation [19], that were shown to be useful in understanding the responses of neurons in early visual areas [20, 21]. The concept lattices obtained by FCA might enable us to display and browse these invariances: if the response of a subset of cells indicates the presence of an invariant feature in a stimulus, then all stimuli having this feature should form the extent of a concept whose intent is given by the responding cells.

5.1 Physiological data

The data were obtained through [22], where the experimental details can be found. Briefly, spike trains were obtained from single neurons within the upper and lower banks of the superior temporal sulcus (STSa) of an awake and behaving monkey (*Macaca mulatta*) via standard extracellular recording techniques [23]. During the recordings, the monkey had to perform a fixation task. This area contains cells which are responsive to faces and other complex shapes. Extracellular voltage fluctuations were measured, and the stereotypical action potentials (i.e., 'spikes') of the neuron were detected and their temporal sequence was recorded resulting in a 'spike train'. These spike trains were turned into distinct samples, each of which contained the spikes from −300 ms before to 600 ms after the stimulus onset with a temporal

resolution of 1 ms. The stimulus set consisted of 1704 images, containing colour and black and white views of human and monkey head and body, animals, fruits, natural outdoor scenes, abstract drawings and cartoons. Stimuli were presented for 55 ms each without inter-stimulus gaps in random sequences. While this rapid serial visual presentation (RSVP) paradigm complicates the task of extracting stimulus-related information from the spike trains, it has the advantage of allowing for the testing of a large number of stimuli. A given cell was tested on a subset of 600 or 1200 of these stimuli, each stimulus was presented between 1–15 times.

The data were previously analysed with respect to the stimulus selectivity of individual cells only. Previous neural population decoding studies were aimed at identifying stimulus labels (e.g. [2, 3]) only. This paper presents the first analysis of the semantic structure of neural data with FCA.

5.2 Bayesian thresholding

In order to apply FCA, we extracted binary attributes from the raw spike trains. We will experiment with many-valued attributes to describe the neural response, but first we employ a simple binary thresholding. This binary attribute should be as informative about the stimulus as possible, to allow for the construction of meaningful concepts. We do this by Bayesian thresholding, as detailed below. This procedure also avails us of a null hypothesis $H_0 =$ "the responses contain no information about the stimuli".

The usual way of obtaining binary responses from neurons is thresholding the spike count within a certain time window. This is a relatively straightforward task, if the stimuli are presented well separated in time and a large number of trials per stimulus are available. Then latencies and response offsets are often clearly discernible and thus choosing the time window is not too difficult. However, under RSVP conditions with few trials per stimulus, response separation becomes more tricky, as the responses to subsequent stimuli will tend to follow each other without an intermediate return to baseline activity. Moreover, neural responses tend to be rather noisy. We will therefore employ a simplified version of the generative Bayesian Bin classification algorithm (BBCa) [24], which was shown to perform well on RSVP data [25].

BBCa was designed for the purpose of inferring stimulus labels $g \in \{0; \ldots; S-1\}$ from a continuous-valued, scalar measure z of a neural response. The range of z is divided into a number of contiguous bins. Within each bin, the observation model for the g is multinomial with S possible stimulus labels (outcomes), i.e., g assumes value l with probability p_l such that $\sum_{l=0}^{S-1} p_l = 1$. Furthermore, because the p_l in each bin are unknown *a priori*, we employ a Dirichlet prior [26] to express this ignorance:

$$p(p_0, \ldots, p_{S-1}) = \frac{\Gamma\left(\sum_{l=0}^{S-1} \alpha_l\right)}{\prod_{l=0}^{S-1} \Gamma(\alpha_l)} \prod_l p_l^{\alpha_l - 1} \tag{1}$$

The Dirichlet prior is chosen for analytical convenience: it is conjugate to the multinomial observation model. Thus the posterior is Dirichlet too, and inference of the p_l can be done in closed form by adjusting the parameters α_l. For details, the reader is referred to [24, 26]. We show in [24] that one can iterate/integrate over all possible bin boundary configurations efficiently, making exact Bayesian

inference feasible. Moreover, the marginal likelihood (or model evidence) becomes thus available, which can be used to infer the posterior distribution over all spike counting windows. We make two simplifications to BBCa: 1) z is discrete, because we are counting spikes and 2) we use models with only 1 bin boundary Z_0 in the range r of z, i.e.,

$$P(g = l_i | z = z_i) = \begin{cases} p_{l_i} \text{ if } z_i \leq Z_0 \\ q_{l_i} \text{ otherwise} \end{cases} \tag{2}$$

$$\sum_{l=0}^{S-1} p_l = 1 \ , \ \sum_{l=0}^{S-1} q_l = 1 \tag{3}$$

$$p(p_0, \ldots, p_{S-1}) = \frac{\Gamma\left(\sum_{l=0}^{S-1} \alpha_l\right)}{\prod_{l=0}^{S-1} \Gamma(\alpha_l)} \prod_{l=0}^{S-1} p_l^{\alpha_l - 1} \tag{4}$$

$$p(q_0, \ldots, q_{S-1}) = \frac{\Gamma\left(\sum_{l=0}^{S-1} \beta_l\right)}{\prod_{l=0}^{S-1} \Gamma(\beta_l)} \prod_{l=0}^{S-1} q_l^{\beta_l - 1} \tag{5}$$

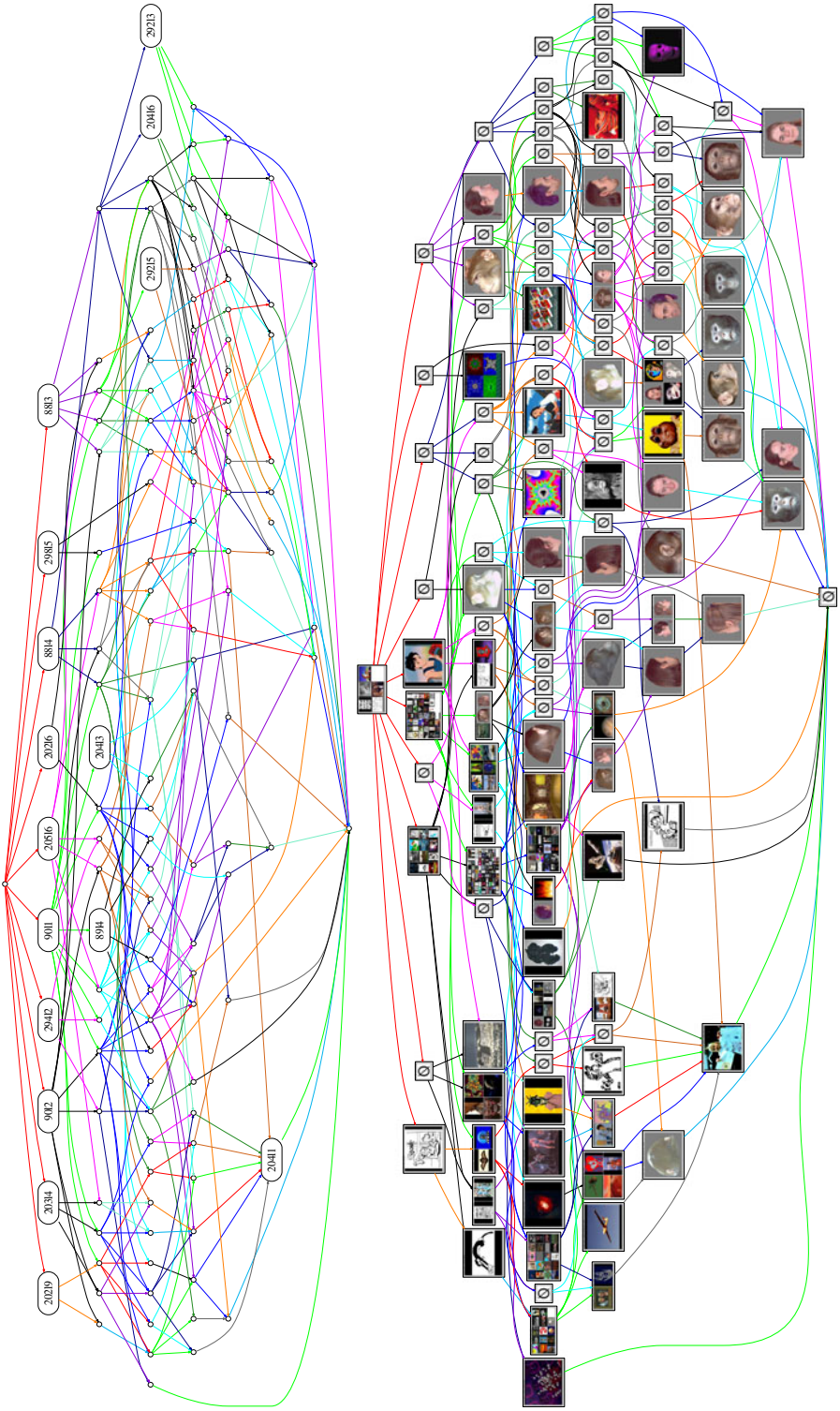
$$p(Z_0) = \frac{1}{|r|}. \tag{6}$$

We have no a priori preferences for any stimulus label, thus we choose $\forall l : \alpha_l = \beta_l = 1$.

The bin membership (in the higher bin) of a given neural response can then serve as the binary attribute required for FCA, since BBCa weighs bin configurations by their classification (i.e., stimulus label decoding) performance. We proceed in a straight Bayesian fashion: since the bin membership is the only variable we are interested in, all other parameters (counting window size and position, class membership probabilities, bin boundaries) are marginalised. This minimises the risk of spurious results due to "contrived" information (i.e., choices of parameters) made at some stage of the inference process. Afterwards, the probability P_u that the response belongs to the upper bin is thresholded at a probability of 0.5, i.e., if the probability is larger than 0.5, then there will be a cross in the context.

In addition to this simple binarisation, we also experimented with attribute scaling [9] to investigate the robustness of the prominent features of the resulting lattices. The attributes of our contexts are derived from probabilities, so 0.5 is the natural discretisation point if one wants to minimise the chance of misclassifying a neural response as 'above threshold' when it really is below, and vice versa. However, this discretisation inevitably injects noise into the attributes. One might wonder how much our results are affected by this noise. We addressed this question by scaling the bin membership probability P_u ordinally at $P_u > 0.4$, $P_u > 0.5$ and $P_u > 0.6$, thereby creating three attributes for the three response levels of each neuron. This scaling has e.g. the effect that stimuli which evoked a response with $P_u > 0.6$ from a given neuron would be introduced below a stimulus which evoked a response with $0.4 \leq P_u < 0.6$ from the same neuron.

Since BBCa yields exact model evidences, it can also be used for model comparison. Running the algorithm with no bin boundaries in the range of z effectively yields the probability of the data given the "null hypothesis" H_0: z does not contain


29213
20446
29215
8813
29085
8814
20263
20433
20656
9011
8944
29421
9082
20341
20334
20219

◀ **Fig. 3** *Left/top*: a lattice with reduced labelling on the attributes which are the cells, i.e., cells (e.g. '20219') are only shown in their attribute concepts. A small circle denotes a concept which introduces no cells. All edges originating at a concept have the same colour to facilitate visual determination of the ordering relations between concepts. The colours have no meaning beyond that. The majority of cells are introduced in lower neighbours of the top concept. *Right/bottom*: the same lattice as on the right with reduced labelling on the stimuli, i.e., stimulus images are only shown in their object concepts. The ∅ indicates that an extent is the intersection of the upper neighbours' extents, i.e., no new stimuli are introduced by this concept. This lattice shows an emphasis on "face" and "head" concepts, with cartoon faces introduced towards the top and monkey faces towards the bottom. For details, see text

any information about g. We can then compare it against the alternative hypothesis described above (i.e., the information which bin z is in tells us something about g) to determine whether the cell has responded at all.

5.3 Cell selection

The experimental data consisted of recordings from 26 cells. To minimise the risk that the computed neural responses were a result of random fluctuations, we excluded a cell if 1) H_0 was more probable than 10^{-6} or 2) the posterior standard deviations of the counting window parameters were larger than 20 ms, indicating large uncertainties about the response timing. Cells which did not respond above the threshold included all cells excluded by the above criteria (except one). Furthermore, since not all cells were tested on all stimuli, we also had to select tuples of subsets of cells and stimuli such that all cells in a tuple were tested on all stimuli. Incidentally, this selection can also be accomplished with FCA, by determining the concepts of a context with $gIm =$"stimulus g was tested on cell m" and selecting those with a large number of stimuli $\times$ number of cells. One of these cell and stimulus subset pairs (16 cells, 310 stimuli) was selected for further exemplary analysis, but the lattices computed from the other subset pairs displayed similar features.

6 Results

To analyse the neural code, the thresholded neural responses were used to build stimulus-by-cell-response contexts. We performed FCA on these with ColibriConcepts,[1] created stimulus image montages[2] and plotted the lattices.[3] In these graphs, the images represent the formal objects.

Figure 3, right/bottom, shows a lattice which has an emphasis on "face" and "head" concepts. The concepts in the right half of the lattice introduce predominantly different views of faces and heads, with 'back of the head' stimuli concentrated in the middle of the lattice, whereas front and side views are grouped together towards the right. The concepts introducing human and cartoon faces (i.e., with extents consisting of general "face" images) tend to be higher up in the lattice and their intents tend

[1] Available at http://code.google.com/p/colibri-concepts/

[2] Via ImageMagick, available at http://www.imagemagick.org

[3] With Graphviz, available at http://www.graphviz.org

to be small. In contrast, the lower concepts introduce mostly single monkey faces (and faces of the monkey's caregivers), with the bottom concepts having intents of ≥ 7 cells. We may interpret this as an indication that the neural code has a higher "resolution" for faces of conspecifics (and other "important" faces) than for faces in general, i.e., other monkeys are represented in greater detail in a monkey's brain than humans or cartoons. This feature can be observed in most lattices we generated.

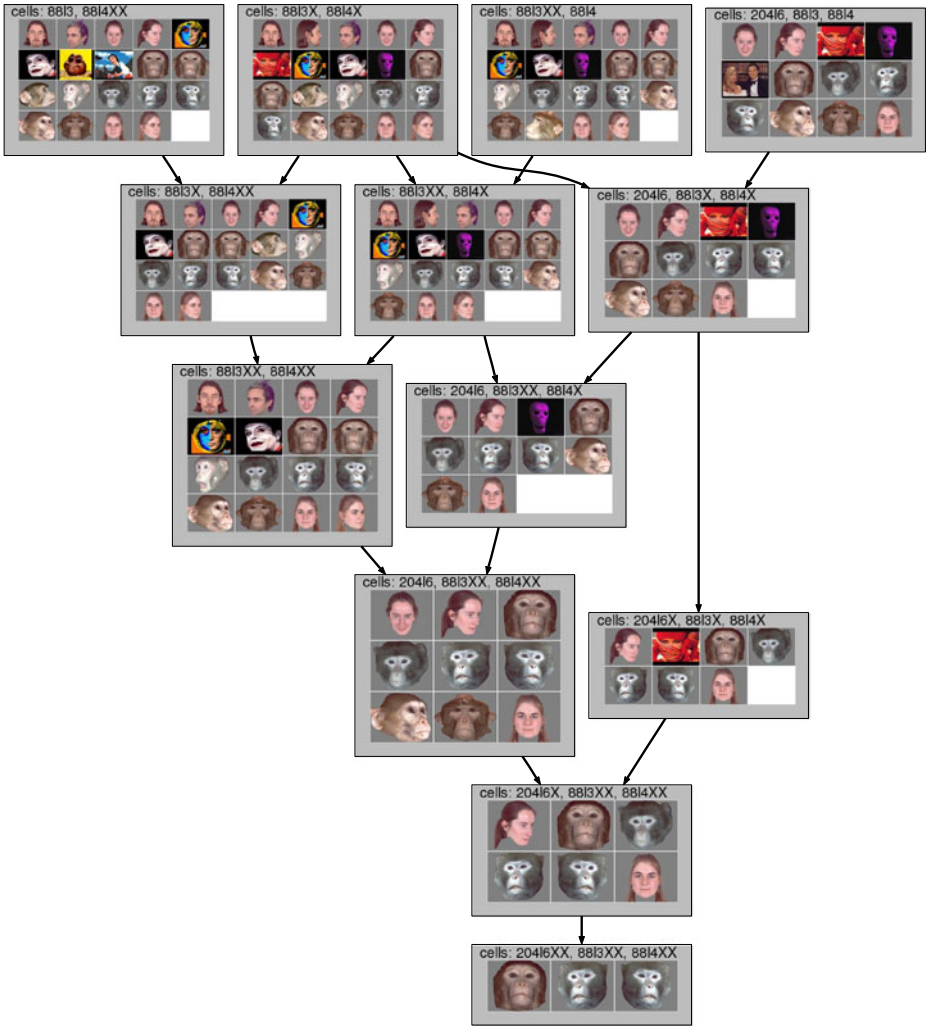


Fig. 4 A subgraph of a lattice with full labelling. Bin membership probabilities P_u were scaled ordinally (see Section 5.2). Ordinal scaling could potentially lead to a distortion of the order between e.g. cartoon and monkey faces (cf. Fig. 3), if the former evoked a stronger response than the latter. However, this does not seem to be the case: monkey and caregiver faces can still be found at the bottom of the graph, whereas cartoon face only appear at the top. The *top frame* of each concept shows the cells which comprise the intent. The steps of the ordinal scaling are indicated by appending 'X's to the cellname. For example, '204l6' means that cell '204l6' had a $P_u > 0.4$, '204l6X' indicates a $P_u > 0.5$ and '204l6XX' stands for $P_u > 0.6$

Thus, monkey STSa cells are not just responsive to faces in general, but to specific subclasses, such as monkey faces, in particular. Figure 3, left/top, shows the same lattice as in the right/bottom half of the figure, but with reduced labelling on the cells. Note that the majority of the cells are introduced in concepts directly below the top concept, but not all of these concepts introduce stimuli. This highlights the importance of looking at a population of cells to decode specific stimulus information (such as 'it's the face of my caregiver').

To demonstrate that this face hierarchy is not an artifact of thresholding noisy neural responses, we selected a subset of 3 face-selective cells and ordinally scaled the bin membership probabilities (as described in Section 5.2). A subgraph of the resulting lattice with fully labelled extents is shown in Fig. 4. Ordinal scaling could potentially lead to a distortion of the ordering between e.g. cartoon and monkey faces, if the former evoked a stronger response than the latter. However, this does not seem to be the case: monkey and caregiver faces can still be found at the bottom of the graph, whereas cartoon faces only appear at the top.

Figure 5 shows a subgraph from another lattice with full labelling and ordinally scaled probabilities. Full labelling is of interest in these applications because viewing the full extent simultaneously gives an impression of "what this concept is about". The concepts in the left half of the graph are face concepts, whereas the extents of the concepts in the right half also contain a number of non-face stimuli. Most of the latter have something "roundish" about them. The bottom concepts, being subordinate to both the "round" and the "face" concepts, contain stimuli with both characteristics,

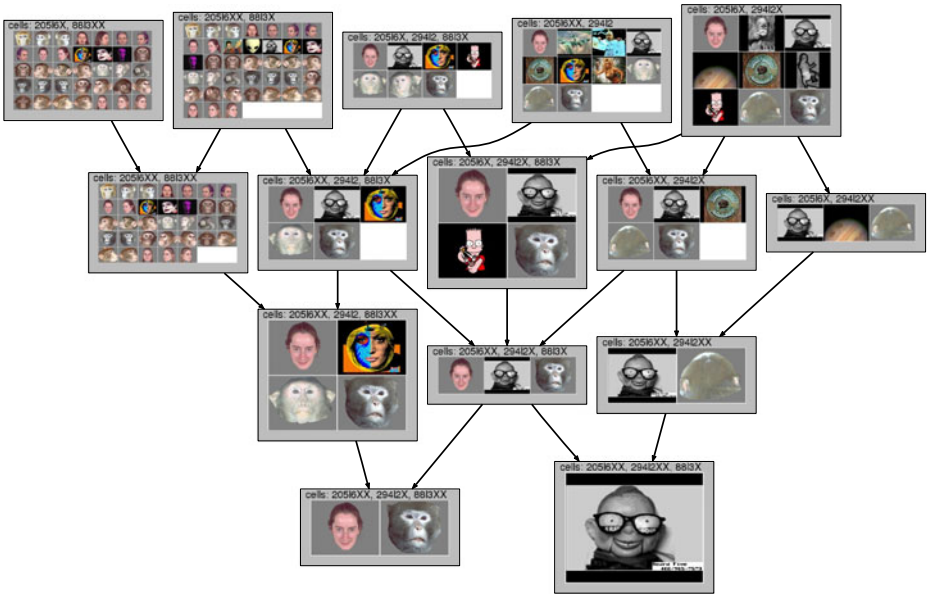


Fig. 5 A subgraph of a lattice with full labelling. The concepts on the right side are not exclusively "face" concepts, but most members of their extents have something "roundish" about them. The *top frame* of each concept shows the cells which comprise the intent. The steps of the ordinal scaling are indicated by appending 'X's to the cellname. For example, '20416' means that cell '20416' had a $P_u > 0.4$, '20416X' indicates a $P_u > 0.5$ and '20416XX' stands for $P_u > 0.6$

which points towards a product-of-experts (PoE) encoding [27]. In PoE, each 'expert' can be thought of as an attribute (or attribute combination) of the represented item. These experts are expected to correspond to meaningful aspects of the information items. Several examples of this kind can be found in the other graphs of the complete concept lattices, which cannot be included in this paper.

7 Conclusion

We demonstrated the potential usefulness of FCA for the exploration and interpretation of neural codes. This technique is feasible even for high-level visual codes, where linear decoding methods [20, 21] fail, and it provides qualitative information about the structure of the code which goes beyond stimulus label decoding [1–4]. The semantic structure of neural data has previously been analysed with tree-based clustering methods [28]. Imposing a tree structure on the data may be inappropriate for neural data that reflects a more general semantic structure, as supported by our results.

Individual concepts have an interpretation from the perspective of both a theoretical neuroscientist and also for a neuron trying to decode categories. The activation of the neurons of a concept's intent with reduced labelling on the stimuli show the stimulus category encoded by these neurons assuming that all other neurons are inactive. However, from a neuron's perspective it is more plausible to consider the full extent (stimuli). The category formed by the full extent can be decoded by observing the responses of only these neurons, and the activation of all other neurons can be ignored for this purpose. This interpretation is a useful answer to the long-standing issue of distributed versus local neural encoding [29]. Our results suggest that at least on our limited sample of neurons and stimuli, only a small number (e.g. 7) of neurons are needed to form quite specific concepts. One can only speculate at this point how adding substantially more stimuli and neurons scales this result but based on these results we would expect a relatively small number of neurons per concept. This is consistent with the extremely limited relative connectivity of real neurons, which on average can connect to a small fraction of other neurons (approx. $10^4/10^{11} = 10^{-7}$ in the human brain). FCA suggests that making these relatively small number of connections to the correct subset of other neurons can lead to the representation of useful categories, without considering the complete pattern of neural activity within an area.

As detailed above, the majority of the cells are introduced in concepts directly below the top concept. Thus, observing that a given cell is active does generally not imply the activity of any other cells (though there are some exceptions, e.g. cell '89l4', which is subordinate to '90l1' in Fig. 3, right). In other words, neurons represent information items which are logically largely independent. Together with the aforementioned 'clustering' of visually similar stimuli and the introduction of more specific stimuli towards the bottom of the lattice, this observation provides further evidence for the hypothesis that visual cortical neurons implement a product-of-experts style code, where each code element (neuron) indicates a constraint on what is being represented. Morevoer, the fact that many of the concepts we found are easily interpretable suggests that these constraints explicit [29].

Clearly, however, our application of FCA for this analysis is still in its infancy. It would be very interesting to repeat the analysis presented here on data obtained from simultaneous multi-cell recordings, to elucidate whether the conceptual structures derived by FCA are used for decoding by real brains. On a larger scale than single neurons, FCA could also be employed to study the relationships in fMRI data [30, 31]. The averaging inherent in fMRI imaging will erase some of the fine details of the lattice, but we hope that its basic structure will be preserved.

Acknowledgements D. Endres was supported by MRC fellowship G0501319. We thank D. Xiao and D. Perrett for making the data available to us.

References

1. Georgopoulos, A.P., Schwartz, A.B., Kettner, R.E.: Neuronal population coding of movement direction. Science **233**(4771), 1416–1419 (1986)
2. Földiák, P.: The 'ideal homunculus': statistical inference from neural population responses. In: Eeckmann, F., Bower, J. (eds.) Computation and Neural Systems, pp. 55–60. Kluwer, Norwell (1993)
3. Oram, M.W., Földiák, P., Perrett, D.I., Sengpiel, F.: The 'ideal homunculus': decoding neural population signals. Trends Neurosci. **21**, 259–265 (1998)
4. Quiroga, R.Q., Reddy, L., Koch, C., Fried, I.: Decoding visual inputs from multiple neurons in the human temporal lobe. J. Neurophysiol. **98**(4), 1997–2007 (2007)
5. Duda, O.R., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, New York (2001)
6. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, New York (1991)
7. Földiák, P.: Sparse neural representation for semantic indexing. In: XIII Conference of the European Society of Cognitive Psychology (ESCOP-2003). http://www.st-andrews.ac.uk/∼pf2/escopill2.pdf (2003)
8. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In: Rival, I. (ed.) Ordered Sets, pp. 445–470. Reidel, Dordrecht-Boston (1982)
9. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer (1999)
10. Ganter, B., Stumme, G., Wille, R. (eds.): Formal concept analysis, foundations and applications. In: Lecture Notes in Computer Science, vol. 3626. Springer (2005)
11. Priss, U.: Formal concept analysis in information science. Annu. Rev. Inf. Sci. Technol. **40**, 521–543 (2006)
12. Földiák, P., Endres, D.: Sparse coding. Scholarpedia **3**(1), 2984. http://www.scholarpedia.org/article/Sparse_coding (2008)
13. Földiák, P.: Forming sparse representations by local anti-Hebbian learning. Biol. Cybern. **64**, 165–170 (1990)
14. Földiák, P.: Sparse coding in the primate cortex. In: Arbib, M.A. (ed.) The Handbook of Brain Theory and Neural Networks, 2nd edn., pp. 1064–1068. MIT Press (2002)
15. Olshausen, B.A., Field, D.J., Pelah, A.: Sparse coding with an overcomplete basis set: a strategy employed by V1. Vis. Res. **37**(23), 3311–3325 (1997)
16. Olshausen, B.A.: Learning linear, sparse, factorial codes. Technical Report AIM 1580 (1996)
17. Simoncelli, E.P., Olshausen, B.A.: Natural image statistics and neural representation. Annu. Rev. Neurosci. **24**, 1193–1216 (2001)
18. Rolls, E.T., Treves, A.: The relative advantages of sparse versus distributed encoding for neuronal networks in the brain. Netw. **1**, 407–421 (1990)
19. Dayan, P., Abbott, L.F.: Theoretical Neuroscience. MIT Press, London, Cambridge (2001)
20. Jones, J.P., Palmer, L.A.: An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J. Neurophysiol. **58**(6), 1233–1258 (1987)
21. Ringach, D.L.: Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. J. Neurophysiol. **88**, 455–463 (2002)
22. Földiák, P., Xiao, D., Keysers, C., Edwards, R., Perrett, D.I.: Rapid serial visual presentation for the determination of neural selectivity in area STSa. Prog. Brain Res. **144**, 107–116 (2004)
23. Oram, M.W., Perrett, D.I.: Time course of neural responses discriminating different views of the face and head. J. Neurophysiol. **68**(1), 70–84 (1992)

24. Endres, D., Földiák, P.: Exact Bayesian bin classification: a fast alternative to bayesian classification and its application to neural response analysis. J. Comput. Neurosci. **24**(1), 21–35 (2008). doi:10.1007/s10827-007-0039-5
25. Endres, D.: Bayesian and information-theoretic tools for neuroscience. Ph.D. thesis, School of Psychology, University of St. Andrews, U.K. http://hdl.handle.net/10023/162 (2006)
26. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2007)
27. Hinton, G.E.: Products of experts. In: Ninth International Conference on Artificial Neural Networks ICANN 99, number 470 in ICANN (1999)
28. Kiani, R., Esteky, H., Mirpour, K., Tanaka, K.: Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J. Neurophysiol. **97**(6), 4296–4309 (2007)
29. Földiák, P.: Neural coding: non-local but explicit and conceptual. Curr. Biol. **19**(19), R904–R906 (2009)
30. Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L.: Identifying natural images from human brain activity. Nature **452**, 352–255 (2008). doi:10.1038/nature06713
31. Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M., Morito, Y., Tanabe, H., Sadato, N., Kamitani, Y.: Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron **60**, 915–929 (2008)