Neural correlates of visual object recognition learning under temporal statistical regularities

Thesis

for the degree

doctor rerum naturalium (Dr. rer. nat.)

approved by the Faculty of Natural Sciences of
Otto von Guericke University Magdeburg

by M.Sc. Ehsan Kakaei

born on 16.01.1991 in Tabriz

Examiners: Prof. Dr. Jochen Braun
Prof. Dr. Tobias H. Donner

submitted on 19.02.2025
defended on 16.12.2025

Abstract

Neural correlates of visual object recognition learning under temporal statistical regularities

M.Sc. Ehsan Kakaei

We live in an environment that provides us with sensory information rich in spatio-temporal regularities, at different scales ranging from regularities within an entity (e.g. co-occurring visual features of an object) to relations between multiple elements (e.g. sequences of objects that belong to the same setting, such as trees and animals belonging to nature, or PC and printer belonging to an office). We implicitly learn these regularities and utilise them to establish stable, yet flexible, neural representations of single elements in the environment and their relative positions in space and time. This type of implicit learning is essential for cognitive development and facilitates learning and performance in different cognitive domains, such as visual search, motor learning, and object recognition. Additionally, altering these spatio-temporal regularities also changes neuronal and neural responses in the brain.

In our studies, we aimed to examine the effect of temporal regularities on object recognition learning and its underlying neural substrates. Specifically, we hypothesised that implicit learning of the temporal association between visual objects that are initially unknown to observers can facilitate their learning. For this reason, we sought to closely monitor the changes associated with learning in both cognitive performance and neural representations of individual objects. Moreover, we wondered how neural representations of learning under temporal regularities between multiple objects would compare to representations of learning at the level of individual objects.

On the behavioural level, we show that temporal regularities benefited participants in learning to recognise novel 3D objects and altered the order by which the objects were learnt. On the neural level, we identified brain regions that provided rich information on temporal regularities at both single object and on multi-object levels.

These two levels of regularities largely coexisted in ventral occipitotemporal regions. Moreover, we could closely monitor development of representations in single object level over runs and conclude that the brain of mature humans utilises the existing neural substrates to form representations of novel objects, but these representations are not stable and are subject to changes as one gains expertise on distinguishing previously seen objects from unseen objects of the same kind.

Zusammenfassung

Neurale Korrelate des Lernens der visuellen Objekterkennung unter zeitlichen statistischen Regularitäten.

M.Sc. Ehsan Kakaei

Wir leben in einer Umgebung, die uns mit sensorischen Informationen versorgt, die reich an spatio-temporalen Regularitäten sind, die sich über verschiedene Skalen erstrecken, von Regularitäten innerhalb eines Objekts (z. B. koexistierende visuelle Merkmale eines Objekts) bis hin zu Beziehungen zwischen mehreren Elementen (z. B. Sequenzen von Objekten, die zu demselben Kontext gehören, wie Bäume und Tiere in der Natur oder PC und Drucker in einem Büro). Wir lernen diese Regularitäten implizit und nutzen sie, um stabile, aber flexible neuronale Repräsentationen einzelner Elemente in der Umgebung und deren relative Positionen in Raum und Zeit zu etablieren. Diese Art des impliziten Lernens ist entscheidend für die kognitive Entwicklung und erleichtert das Lernen und die Leistung in verschiedenen kognitiven Bereichen, wie z. B. visuelle Suche, motorisches Lernen und Objekterkennung. Darüber hinaus verändert die Modifikation dieser spatio-temporalen Regularitäten auch die neuronalen und neuralen Reaktionen im Gehirn.

In unseren Studien zielten wir darauf ab, den Einfluss zeitlicher Regularitäten auf das Lernen der Objekterkennung und die zugrunde liegenden neuronalen Substrate zu untersuchen. Insbesondere stellten wir die Hypothese auf, dass das implizite Lernen der zeitlichen Assoziation zwischen visuellen Objekten, die den Beobachtern zunächst unbekannt sind, deren Lernen erleichtern kann. Aus diesem Grund wollten wir die mit dem Lernen verbundenen Veränderungen sowohl in der kognitiven Leistung als auch in den neuronalen Repräsentationen einzelner Objekte genau überwachen. Darüber hinaus fragten wir uns, wie sich die neuronalen Repräsentationen des Lernens unter zeitlichen Regularitäten zwischen mehreren Objekten im Vergleich zu den Repräsentationen des Lernens auf der Ebene einzelner Objekte verhalten würden.

Auf der Verhaltensebene zeigen wir, dass zeitliche Regularitäten den Teilnehmern

beim Lernen, neuartige 3D-Objekte zu erkennen, zugutekamen und die Reihenfolge, in der die Objekte gelernt wurden, veränderten. Auf der neuronalen Ebene identifizierten wir Gehirnregionen, die reichhaltige Informationen über zeitliche Regularitäten sowohl auf der Ebene einzelner Objekte als auch auf der Ebene mehrerer Objekte lieferten. Diese beiden Ebenen der Regularitäten existierten weitgehend koexistent in den ventralen okzipito-temporalen Regionen. Darüber hinaus konnten wir die Entwicklung der Repräsentationen auf der Ebene einzelner Objekte über die Durchgänge hinweg genau überwachen und schlussfolgern, dass das Gehirn reifer Menschen die bestehenden neuronalen Substrate nutzt, um Repräsentationen neuartiger Objekte zu bilden, diese Repräsentationen jedoch nicht stabil sind und Veränderungen unterliegen, während man Expertise im Unterscheiden zuvor gesehener Objekte von ungesehenen Objekten derselben Art erlangt.

Contents

Chapter 1

Introduction

While I stare at my keyboard, thinking of a way to break it to you what I am really going to talk about, I don't just see a keyboard. I see a set of tightly placed keys, each carrying a tiny drawing on them, arranged in a certain order. It took me years to learn what each key does, except for the magical "Scroll Lock"! First, I had to learn to recognise and distinguish each of the letters, signs, and numbers, then by more experience I learnt to locate each key relative to the other keys, which really helps to write a thesis faster. Actually, that is what this thesis is all about. It is about how we learn to recognise visual objects, differentiate them from the other objects, and how their relative order of them helps in this procedure of learning.

In this thesis, I will first introduce you to the concept of spatio-temporal regularities and discuss some influential works that examined effects of such regularities on cognitive functions. Then, I will provide a detailed introduction on neural substrates of object recognition in the visual domain and how neural spatio-temporal regularities affect these neural responses. Afterwards, I will briefly discuss our approach to the analysis of fMRI data using multi-voxel pattern analysis. Finally, I will discuss

the motivation and main findings of our studies before we read each study. The first study is a cognitive study that we conducted on the effect of temporal regularities on object recognition learning (Kakaei et al., 2021). In two later chapters, we will move on to two of our functional imaging studies that focus on neural correlates of object recognition learning (Kakaei and Braun, 2024a), and on how temporal regularities alter these neural substrates (Kakaei and Braun, 2024b).

1.1 Spatio-temporal regularities

One can define the spatio-temporal regularities as characteristics of a phenomenon that are predictable over space and/or time. This predictability can be as simple as the abundance of a single characteristic, or as complex as probabilistic relations between multiple characteristics governed by a certain abstract rule. Both humans and other animals can implicitly learn such regularities and use them to form expectations about their environment. This form of automatic learning of the regularities is termed as *statistical learning* which is thought to be essential for development and cognition. For example, in one of the earliest studies on statistical learning, Saffran and colleagues reported that adults (Saffran et al., 1996b) and 8-month-old infants (Saffran et al., 1996a) are sensitive to statistical regularities in sequences of sounds. The authors used transitional probability (predictability) between sounds (characteristics) to create pseudo-words that contained highly predictive sounds within words, but not between words (**Fig. 1.1.A**). Even in the absence of other acoustic information — such as tone, stress, or pauses — subjects could distinguish pseudo-words from non-words.

Many studies followed the logic of Saffran et al. (1996a) and examined the effects of such statistical regularities in various domains such as visual search (e.g. Chun and Jiang, 1998) and motor learning (e.g. Hunt and Aslin, 2001). Two of such

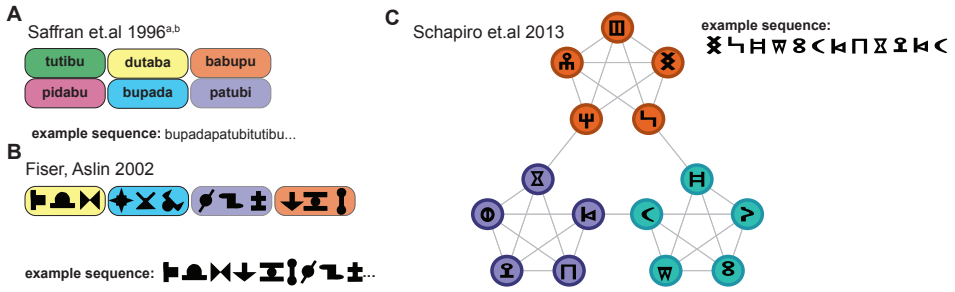

Figure 1.1: Examples from three major studies that examined the effect of spatio-temporal regularities on learning. **A)** Saffran et al. (1996a) showed that infants can learn to recognise pseudo-words composed of predictable sounds, from non-words. **B)** Fiser and Aslin (2002) expanded this paradigm to the visual modality. **C)** Schapiro et al. (2013) showed that the effect of predictability on learning is not limited to fixed order probabilities and can be generalised into more complicated probability distributions such as temporally grouped stimuli.

studies particularly focused on implicit learning of spatial (Fiser and Aslin, 2001) and temporal (Fiser and Aslin, 2002) aspects of the statistical regularities in visual scenes (**Fig. 1.1.B**) and showed that humans are sensitive to statistical regularities in multiple orders. Fiser and Aslin showed that subjects learn statistical characteristics of visual scenes that go beyond simple frequency statistics ($P(X)$ probability of observing stimulus X), and they can learn to form associations between co-occurring visual symbols in both forms of joint ($P(X, Y)$ probability of observing both stimuli X and Y) and conditional probabilities ($P(X|Y)$ probability of observing stimulus X when stimulus Y is observed).

Fast-forwarding to 2013, Schapiro et al. show that higher-order statistical regularities can be learned even when the transitional probabilities are uniform for all stimuli. In that study, the authors embedded higher-order temporal regularities in sequences of visual stimuli by assigning each stimulus to a community structure [1].

[1] a set of nodes that have a higher number of connections between each other than with nodes from the other sets. In the context of temporal statistical regularities, a community structure corresponds to a set of objects that follow each other with higher probability than the objects outside the community

After extensive training, subjects were able to parse the transition from one community structure to another.

This type of statistical regularity arising from manipulating the geometrical property of a transitional probability graph has gained a lot of attention and more studies have provided evidences on sensitivity to such statistical regularities in human subjects. For example, it has been shown that such regularities can facilitate motor learning (Karuza et al., 2017; Kahn et al., 2018). In our studies, we also examined cognitive benefits of higher-order statistical regularities arising from such geometrical manipulations and extended those findings to visual object recognition.

1.2 Neural correlates of object recognition

Before we jump into discussing the effects of spatio-temporal regularities in object recognition, I find it useful to first discuss the neural correlates of visual object recognition. Later, in the **section 1.3** we will discuss how the activities of these neural substrates are modulated by spatio-temporal regularities.

Visual information in the environment arrives at the brain from the eyes via the optic nerve, ipsilaterally. In the optic chiasm, part of this information crosses to the contralateral brain hemispheres and visual information of the same hemifield from both eyes stays on the contralateral hemisphere. Later, both ipsilateral and contralateral projections reach the Lateral Geniculate Nucleus (LGN) distinctively on separate layers. Even in such an early stage, various types of neurons are differentially responsive to some of the basic features of the visual stimulus such as spatial and temporal frequencies and colour (Derrington and Lennie, 1984; Schiller and Logothetis, 1990; Van Essen and Gallant, 1994).

Via projections between LGN and primary visual cortex (V1), visual information

arrives at Calcarine. This is where the first cortical process of object representation takes place. Various cells in this area show selectivity to more features such as orientation and direction of motion, in addition to the previous features extracted in LGN — colour and spatial and temporal frequencies (Hubel and Wiesel, 1962; Schiller et al., 1976; Van Essen and Gallant, 1994). Classically it is suggested that there is a hierarchical structure from this stage onwards and as we move higher in the hierarchy, via feed-forward connections, simpler features in the lower levels combine to form more complex features, receptive fields[2] of neurons become larger, and functional segregation of neurons in terms of their feature-selectivity diminishes (Felleman et al., 1997; Grill-Spector and Weiner, 2014). Selectivity of the neurons in V1 to features like orientation, spatial and temporal frequency turns into selectivity to angles and contours in V2, and higher in the hierarchy in V4 they become sensitive to shape and form (Van Essen and Gallant, 1994; Kravitz et al., 2013).

Unlike in V1 and V2 where neurons are more responsive to the basic features, the cells in V4 are selective to more complex object features (Kobatake and Tanaka, 1994) and by integrating information over this neural population one can encode natural images (Rust and DiCarlo, 2010). Moreover, tuning of the neurons in this region to object features is invariant to the stimulus position, but can be modulated by the attentional network through a top-down process. This points to the sensitivity of V4 to diagnostic object features that are relevant to behaviour (Mirabella et al. 2007; for a detailed review on the functionality of V4 see Roe et al. 2012).

In later visual processing areas located in the lateral occipital (LO) and ventral temporal (VT) cortices, the neural population is highly specialised and selective to both animate (e.g. face and body parts) and inanimate (e.g. tools and houses) objects, as well as scenes (Kanwisher et al., 1997; Epstein and Kanwisher, 1998; Grill-Spector et al., 2001; Kanwisher et al., 2002). Moreover, the activity of the neural populations is invariant to various modifications such as brightness, direction,

[2]a visual area to which a neuron responds when a visual stimulus is presented or is removed.

colour, and view (Booth and Rolls, 1998; Grill-Spector et al., 1999).

The responses of the neurons in VT are more specialised than the responses in LO, since neural activities in LO show sensitivity to objects in general while the neural tunings in VT are category specific (Grill-Spector et al., 2001; Grill-Spector and Weiner, 2014). This category-selectivity in VT is such that objects of different categories evoke distinct activity patterns over overlapping areas (Haxby et al., 2001). Moreover, these activity patterns are organised such that they exhibit a close relation with the categorical hierarchy. On the top of the hierarchy where we classify all objects in a 'super-class', we have the general object-selectivity of the neural population. This general-selectivity exhibits itself in the form of responses to complete objects (not scrambled images or parts) regardless of their category, over a vast cortical area in inferiotemporal (IT) region. As we move down the objects hierarchy tree, first we can distinguish neural representations of the animate objects from inanimate ones. By diving deeper into the hierarchy, we can further discriminate neural activities arising from various categories (e.g.tools, faces, body parts, buildings,etc.) and sub-categories (e.g. human vs. animal faces/body), usually shown in the form of more similar activities among the exemplars of each category or sub-category than between categories (Kiani et al., 2007; Kriegeskorte et al., 2008b).

In summary, we read that through anatomical projection between the retina and LGN visual information first reaches the subcortical areas where some basic features such as spatial and temporal contrast already start to form. Afterwards, this information is transferred to the primary visual cortex in the occipital lobe (V1) where more primary features, such as colours and spatio-temporal frequencies are formed. Neighbouring regions (V2 and V3) process and integrate this information to form representations of more complex features such as contours and motion. Further up the hierarchy in V4, sensitivity to diagnostic features of shapes increases and the neural activity becomes more behaviourally relevant. Finally, through recurrent feedback and feedforward connections between the occipital and temporal

areas, complex visual features get integrated and form representations of objects in various categories and subcategories (for a detailed review see Van Essen and Gallant 1994; Grill-Spector and Weiner 2014; Kravitz et al. 2013).

1.3 Object recognition and statistical learning

As introduced earlier in **section 1.1**, spatio-temporal regularities are defined as predictable characteristics of the sensory stimuli. In the visual domain, such characteristics can be in the form of features (such as colour, orientation, brightness, etc.), different views of a stimulus, or multiple stimuli with a specific spatio-temporal contiguity. In this section, we will read about some of the previous works that examined the effect of such regularities on neural activities underlying object recognition. First, we will discuss neural correlates of between-features/-views spatio-temporal regularities in the form of *view-invariant* representations. Then, we will read about the neural correlates of spatio-temporal regularities between multiple stimuli in the forms of "simple" and higher-order multi-object regularities.

1.3.1 View-invariant representation

In unalike environments where visual settings such as brightness, visual angle, hue, and depth are different, humans and other animals retain their ability to recognise a familiar object (e.g. a friend's face). Interestingly, single neuron recordings (e.g. in IT; Booth and Rolls 1998), as well as BOLD activities (e.g. in LOC; Grill-Spector et al. 1999), exhibit that neural representations in various object-selective cortical regions stay invariant to such changes in visual settings. Previous studies have extensively examined such robustness in object recognition and have theorised that we make an abstract mental representation of the objects which is tolerant to changes

in the visual setting (*view-invariant* representation), the exact mechanism of which is under debate (for a detailed review see Palmeri and Tarr 2008).

One of the proposed mechanisms for establishing such a view-invariant representation is through plasticity between feature-/view-encoding neurons whose receptive fields are spatio-temporally correlated (McMahon and Leopold, 2012). In other words, when certain features/views regularly appear in close temporal proximity, the sets of neurons that encode those features/views tend to be associated with each other due to neural plasticity. Such an association would help to form a sturdy representation of an object, invulnerable to subtle changes in the visual setting of that object, and to generalise the representation to other objects of similar appearance and categories (Grill-Spector and Weiner, 2014).

There is a large body of evidence supporting the importance of spatio-temporal contiguity in establishing a view-invariant representation. On the behavioural level, Wallis and colleagues (Wallis and Bülthoff, 2001; Wallis et al., 2009), for example, demonstrated that humans use spatio-temporal contiguity between different views of faces to form facial identities. They showed that by altering regularities between consecutive views of a face, people tend to mistakenly identify one face as the other. This effect is not limited to faces and generalises to objects as well (Cox et al., 2005).

On the neural level, using similar alterations in visual stimuli, the influence of spatio-temporal regularities on view-invariant neural representations is studied in both humans and non-human primates. For instance, Li and DiCarlo showed that object-selective neurons in IT decrease their selectivity to an object when temporal contiguity of the visual experience breaks the identity of that object. They demonstrated that by changing the identity of an object from a 'good' object (i.e. an object to which a neuron responds selectively) to a 'poor' one, during a saccade, the selectivity of neurons to that 'good' object gradually declines as monkeys become more experienced (Li and DiCarlo, 2008, 2010, 2012). In humans, such susceptibility

of view-invariant neural representations to temporal contiguity has been observed using fMRI adaptation (also known as repetition suppression) — which emerges in the form of reduction in BOLD activity when identical or highly similar stimuli are presented repeatedly (Grill-Spector et al., 2006). Van Meel and Op de Beeck (2018) examined fMRI BOLD activity levels associated with different views of faces in the face-selective areas (OFA and FFA) and observed adaptation when the identity of the face remained unchanged between consecutive views (higher spatio-temporal regularity), in contrast to the absence of adaptation when the identity changed between (lower spatio-temporal regularity).

In summary, we read that the neural representations of visual objects in object-selective regions are formed such that they are robust to various changes in the visual stimuli, which is essential to object-recognition in novel environments. Spatio-temporal regularities are essential in the formation of such '*view-invariant*' representations, since manipulating these regularities wanes object-specific selectivity on the neuronal level. Presumably, neural plasticity between neural populations encoding features or views that are spatio-temporally correlated plays a role in maintaining and flexibly changing these representations.

1.3.2 Multi-object association learning

Establishing a *view-invariant* representation of an object is a tremendous achievement for the visual system. But what if the visual system could utilise the mechanism that it uses for associating different views or features, to additionally form a representation of the likely context of an object. Eventually, there is much evidence supporting that cortical representations of visually distinct objects in an environment are not completely independent and there is a degree of interdependency between them. In fact, it has been observed that when objects repeatedly appear in a sequence, their neural representations tend to be more correlated with each other (Miyashita,

1988).

Single neuron recordings from IT have shown that when macaque monkeys are trained extensively to memorise an association between paired stimuli, neurons respond differently to paired than unpaired stimuli. Some neurons selectively responded to both of the paired stimuli suggesting that these neurons are encoding the pair. A second type of neurons did not respond during the presentation of the cue stimulus, only to respond after the cue and prior to recall of the second stimulus acting like a predictively encoding neuron (Sakai and Miyashita, 1991). Some neurons selectively respond to the predicting stimulus but have reduced response for the trailing stimulus, exhibiting a form of predictive suppression (Meyer and Olson, 2011). These correlated responses of the neurons to paired stimuli, during the cuing or recall period, only develop after the association is learned, even when this association is implicit and task irrelevant (Erickson and Desimone, 1999). Neural responses to these first-order statistical regularities (i.e. one stimulus is directly predictive of the other stimulus) are not limited to paired-stimulus associations and has been observed in multi-stimulus associations as well (Meyer et al., 2014).

Imaging studies in humans, have shown two different sets of regions involved in implicit learning of multi-object associations. The first set of regions is domain specific, with responses that depend on the modality of the stimuli, and the second set is domain general and is responsive to spatio-temporal regularities in different modalities (for a review see Batterink et al. 2019; Fiser and Lengyel 2022). In the visual domain, many of the domain specific regions that are responsive to multi-object association learning are also involved in object recognition. For example, it is shown that increasing the predictability of upcoming stimuli — such as a sequence made of predefined sets of triplets, similar to the paradigm of Fiser and Aslin (2002) (see **section 1.1** and **Figure 1.1.B**) — can increase the level of BOLD activity in certain visual areas of the brain such as LOC and VOTC (Turk-Browne et al., 2009). These results are in line with the results of non-human primates studies that

we discussed above and point to a possible coexisting representation of objects and their spatio-temporal relations in the visual areas. Such representations would help to establish a flexible abstraction of the environment by associating objects that regularly appear together to neurally similar responses.

Sensitivity to multi-object association learning goes beyond brain regions responsive to visual stimuli and is also observed in other brain regions that are shared between different domains (Fiser and Lengyel, 2022). Using different learning paradigms and methods of analysis, a significant amount of studies have pointed out that medial temporal lobe, especially hippocampus, is involved in the extraction of spatio-temporal regularities between multiple objects.(e.g. Turk-Browne et al. 2010; Gheysen et al. 2011; Schapiro et al. 2012; Hsieh et al. 2014; Hindy et al. 2016). In a paired-association learning task Turk-Browne et al. (2010) showed that the BOLD activity of cueing objects (i.e. stimuli that predict the upcoming stimulus) was enhanced after subjects implicitly learned the associations. In a similar design, by applying representational similarity analysis (RSA), it was observed that when objects are paired (Schapiro et al., 2012), or form fixed sequences (Hsieh et al., 2014) their multi-voxel representations in hippocampus become more similar after learning the association.

Such an increased representational similarity between objects was also observed when the spatio-temporal regularities were on higher order (i.e. preceding object did not necessarily predict the upcoming object). As discussed before, even in the absence of one-to-one associations (as in paired association learning), objects can be temporally grouped such that the members of a group can predict upcoming member, but not objects that belong to the other groups. Using such a temporally clustered design, Schapiro et al. (2013, 2016) observed that representations of the objects that belonged to the same cluster (*community-structure*) were more similar than the representations of objects in different clusters, in various brain regions including HC, the inferio-frontal gyrus,the superio-temporal gyrus, and the anterio-temporal lobe.

In summary, we learned that spatio-temporal regularities between multiple objects alter the neural activities in various brain regions which can be divided into domain-specific and domain-general networks. The domain-specific network is subject to altered activity when the regularities are on lower statistical order while more complex and abstract regularities are represented in higher-level brain areas which are domain general (Batterink et al., 2019; Fiser and Lengyel, 2022).

1.4 Multi-voxel pattern analysis

Since we wanted to examine the changes of the brain activity pattern during the course of learning, we sought to first identify areas of the brain in which individual objects are represented in the sense that information about object identity is encoded. Instead of applying the classical univariate fMRI analysis methods (e.g. generalized linear model [**GLM**]) which address the statistical significance of the event-related voxel activities, we employed a multivoxel pattern analysis (MVPA) method. MVPA typically uses the information provided by a set of voxels, even if they are unresponsive to the desired conditions, as features (predictors) to classify and distinguish desired categories of stimuli. To decide which set of voxels should be included in the analysis, conventional MVPA utilizes either: 1) voxels limited in an anatomical region, 2) voxels in close spatial proximity (e.g. within a spherical searchlight),or 3) voxels responsive to the categories computed by univariate analysis (for a complete review on MVPA see Norman et al. 2006). However, these options did not serve our purposes — performing a whole brain study while benefiting from abundance of information provided by multivoxel analysis — since they either: 1) would have limited us to a single region, 2) would have been computationally expensive to cover the whole brain, or 3) would have discarded information available in the 'unresponsive' voxels.

In order to cope with these limitations, we applied our analysis on the voxels within each of the 758 functional parcels defined by MD758 parcellation (Dornas and Braun, 2018). This parcellation provides highly consistent activity correlations between voxels of a single parcel, low functional correlations between voxels of different parcels, relevance with anatomical structures, as well as high cluster quality. Since MD758 is based on consistent functional correlations between voxels of anatomical regions — defined by the AAL90 atlas (Tzourio-Mazoyer et al., 2002) — we adopted our analysis to this parcellation in order to benefit from both the spatial correspondence and functional correlations between voxels. In other words, we hypothesised that the consistent functional correlations between voxels' activity within a parcel might also provide consistent information on the task in hand.

In the following sections we will take a close look at the steps we took during our MVPA. First we will introduce the linear discriminant analysis we used to distinguish activity patterns of different stimuli. Then, we will discuss how we elicited activity patterns within each parcel, and which measures we used to characterise our observations.

1.4.1 Direct linear discriminant analysis

To identify the brain areas in which cortical representations are informative about object identity, we sought to find a representational space (that is a multi-dimensional space with each dimension corresponding to one measurement unit, such as electrode, voxel, time,etc.) where objects are distinguishable by their identities. One can find such a space by providing the multi-unit (here multivoxel) activity patterns and corresponding IDs of each measurement to a supervised machine learning algorithm. For this purpose we chose direct linear discriminant analysis (DLDA) — which is a modified version of Fisher's linear discriminant analysis (LDA)— since it has a

straight forward approach and has been successfully applied to various tasks in different fields of studies such as computer vision, and image recognition (Yu and Yang, 2001; Ye et al., 2006). The logic behind both classical LDA and DLDA is to find a space in which distances between different classes are maximized while distances between items within each class are minimized. In the context of object recognition, this corresponds to finding a space in which multi-unit activities of observations for an object are highly similar, while those of different objects are highly dissimilar.

It can be shown that finding such a space corresponds to finding a projection matrix $\mathbf{G}$ such that the ratio $\mathbf{J} = \hat{S}_B/\hat{S}_W$ of scatter between classes $\hat{\mathbf{S}}_\mathbf{B}$ over scatter within classes $\hat{\mathbf{S}}_\mathbf{W}$ in the projected data is maximized. This is equivalent to:

$$J = \frac{G^T S_B G}{G^T S_W G}; \qquad \frac{\partial J}{\partial G} = 0 \tag{1.1}$$

where S_B and S_W are the scatter matrices of the data before projection.

The DLDA additionally reduces the dimensionality of data which is essential when the dimensionality of the data is very high but sample size is much lower. In short, this is done by first diagonalizing the S_B matrix, discarding non-informative dimensions (for a classification with κ classes, only the first $\kappa - 1$ eigenvectors are informative), and then diagonalizing the S_W matrix (for full details see Yu and Yang 2001).

To acquire data needed for obtaining the most discriminable space using DLDA, we first extracted the multivoxel activity pattern of each object. This was done by recording the BOLD signal of each trial, within each parcel, over a 9 seconds window (from 2 to 11 seconds after trial onset). For each parcel with N_{vox} voxels, such recording corresponds to a single data point in a $N_{dim} = 9 \times N_{vox}$ dimensional space.Since we wanted to establish a space capable of decoding the identity of the recurrent objects, we trained the classifier using only the activity patterns of the $\kappa = 15$ recurrent objects, resulting in a 14-dimensional representational space.

Once the representational space was established by DLDA, we wanted to validate that the identity of the objects was indeed decodable. When data is high-dimensional and number of classes are much lower than the dimension of the data, a classifier can easily over-fit the model. To overcome this problem, it is a common practice to train the classifier on a set of data and cross-validate it on another set (also known as testing) to conclude if the representation of the classes can be generalised. Here, we randomly selected 10% of the activity patterns of each recurrent object (approximately 20 trials) for the cross-validation and the remaining 90% (approximately 190 trials) was used for training the classifier.

We used several measures to evaluate how decodable identity of the objects is (see **Appendix A.1** for exact calculation of each measure). First, we quantified accuracy of the classifier by assigning the ID of the closest object centroid (Euclidean distance) as the inferred object, then defining the **accuracy** $\alpha = h/N$ ratio of hits h (number of correctly classified test data) to total number of test data N. Second, we measured separability between object pairs i and j by projecting test data of these classes on the line connecting their centroids acquired in the training phase, and calculating the **discriminability** measure $\delta_{\mathbf{ij}} = |\mu_i - \mu_j| / \sqrt{0.5(\sigma_i^2 + \sigma_j^2)}$ where μ and σ are respectively the mean and standard deviation values. Third, we measured overall multi-class discriminability by calculating ***F*-ratio** measure defined as the ratio of between-class to within-class variance $\mathbf{F} = SS_B(N - \kappa)/SS_W(\kappa - 1)$ (Anderson, 2001). Since we wanted to account for the individual differences between subjects, moreover, stimulus-set was different between subjects and in all conditions (sequence types), we performed DLDA for each subject in different sequence types separately.

1.5 Motivation and main findings

As discussed before, we use the statistical contingencies between the elements to enhance our understanding and our interaction with the environment. We discussed different types of statistical contingencies, in the form of spatio-temporal regularities, and their importance in development and cognition. One of the main motivations of my studies was to examine whether such benefits extend to visual object recognition. As presented in the (Kakaei et al., 2021, **Chapter 2**), we studied this by presenting subjects with two sets of novel 3D objects that they had to learn to recognise over two weeks. While a set of objects formed temporal clusters (similar to those of Schapiro et al. 2013), hence enforcing a high-order temporal association between the objects, objects in the other set lacked such a strong association between each other. We in fact confirm that high-order spatio-temporal regularities between visual stimuli enhance the performance of subjects in recognising novel 3D objects. Additionally, thanks to our trial-by-trial approach in analysing performance of the subjects, we could estimate the time when subjects start to recognise an object and we observed that the onset of familiarity was not distributed randomly and was related to the temporal "context" in which an object is presented. Specifically, we observed that mutually predictive objects tend to be learned together.

In addition to examining the benefits of spatio-temporal regularities on recognition performance, we aimed to study the effect of these regularities on the activity of the underlying neural substrate in object recognition. Our fMRI data analysis method provided us with detailed neural representations of individual objects on the trial level which we used in two adjoined studies.

In the study (Kakaei and Braun, 2024a, **Chapter 3**) we aimed to identify the brain areas that are responsive to individual objects during an object recognition learning task, and study the changes in their response patterns as subjects became

familiar with a group of inanimate 3D objects and grew expertise in recognising them. For this purpose, we trained and cross-validated an algorithm, which employs a supervised learning procedure, to decode identity of the objects based on their corresponding multi-voxel activity pattern. We extended previous research on object recognition by showing that identity of inanimate 3D objects are decodable in both low-level and high-level visual areas. This was unprecedented, since previous works usually decode objects on the level of categories (e.g. animate vs inanimate, faces vs body parts, houses vs. cars,etc.) and focus on regions of interests (ROI) that are located in inferio-temporal and lateral-occipital regions. In our study, in addition to these regions, some other regions located in both low-level and high-level areas also provide enough information to decode objects, even on the level of exemplars. (Our individual objects were all generated under the same category. Hence, decoding individual objects corresponds to decoding exemplars in a category). We could show that the multi-voxel activity in various brain regions in both dorsal and ventral visual pathways, as well as some regions in fronto-parietal network, provide us with sufficient decodable information on identity of the individual objects.

By measuring the decodability of the neural information in finer blocks, and by tracing the representations of objects in individual trials, we achieved detailing the learning-related changes in neural representations throughout the experiment. We showed that as subjects learned to recognise objects and distinguish them from unfamiliar ones, neural representations of the learned and not-learned objects diverged from each other. This was particularly more evident in the parieto-frontal and occipito-temporal networks.

In the other adjoined study (Kakaei and Braun, 2024b, **Chapter 4**) we aimed to investigate the effect of temporal regularities on the activities of the neural substrate underlying object recognition. Specifically, we examined neural correlates of temporal regularities in two time scales: First, we studied temporal regularities at the single object level, which corresponds to learning the temporal associations between

different views/features of an object (order of 3 seconds). This was done by examining BOLD signals from the brain regions that formed view-invariant representations of objects, such that the identity of individual objects could be decoded irrespective of the initial viewpoint or orientations. For this part, we employed the representations of individual objects that we acquired in the adjoined publication (Kakaei and Braun, 2024a). On the second time-scale, we focused on temporal regularities at the level of multiple objects, which corresponds to multi-object temporal association learning (order of 30 seconds). For this purpose, we randomly assigned objects to temporal clusters, as we did in Kakaei et al. (2021), and compared representational similarity of the objects within clusters to representational similarity of the objects between clusters.

At the level of the lower temporal-scale (views/features associations), our results validated and extended the previous work on view-invariant representations of objects by showing that the identity of individual objects is decodable in both ventral and dorsal visual pathways, starting from primary visual areas and extending ventrally to inferio-temporal regions and dorsally to the inferio-frontal cortex.

At the level of the higher temporal-scale (multi-object associations), we identified two sets of brain networks. The first set consists of domain-specific regions (i.e. regions that are sensitive to spatio-temporal regularities if the stimuli belong to a specific domain such as visual or auditory) — specifically in the ventral occipito-temporal areas — in which the multi-voxel representations of objects that were temporally clustered were more similar than the representations of the objects in different clusters. The cortical distribution of these regions is such that there is a large overlap between regions that are sensitive to temporal contiguity at the single object level and at the multi-object level. In other words, our results suggest that sensitivity to temporal regularities at multiple levels coexists in the ventral visual stream, and confirm previous findings on the sensitivity of the object-selective regions to temporal contiguity between objects, as we discussed in **Section 1.3.2**.

The second set of regions consists of domain-general regions (i.e. regions that are sensitive to spatio-temporal regularities irrespective of the domain of the stimuli) — specifically the middle frontal, inferio-frontal, superio-frontal, and superior parieto-temporal networks. These regions are associated with cognitive processes that require a more abstract representation of the relations between various attributes of stimuli, such as engagement in a high-load working-memory task, establishing cognitive maps, and adopting complex decision strategies. In our study, the results of the representational similarity analysis in these regions showed that the representations of the objects in the same temporal clusters were more dissimilar than the representations of objects in different clusters. This resembles a "context-specific" map where the representations of objects within a specific context (here a temporal cluster) are distinguishable, but this map resets between different contexts.

In summary, we read that as adult humans come across a new set of visual objects, they presumably utilise the neural substrates that are developed for recognising other objects of the similar category, to form representations of the newly learned objects. Thanks to the spatio-temporal contiguity between various views or features of each object, such representations become robust to changes in visual settings. The formation of such representations seems to interact with inter-object representations. Specifically, when the visual scenery contains bonus information about the inter-object relations, humans implicitly learn to form associations between these objects which in turn helps them to recognise these objects faster. Interestingly, the cortical areas responsible for such associative learning seem to spatially overlap with areas that also encode representations of individual objects.

1.6 Conclusion

Since one of the main objectives of this work was to evaluate changes in neural representations associated with recognition learning, we decided to develop an experimental paradigm during which subjects would gain expertise in visual objects. Considering that fMRI has a low temporal resolution, we needed to slow down this learning process. In the paradigm that we developed, human performance gradually improved over several days, making it possible to assess cortical representations over several successive sessions of fMRI. To slow down learning sufficiently, we not only developed complex 3D shapes, but also presented these from different points of view and in different states of rotation on every trial. In order to make sure that subjects would gain visual expertise, we decided that all the stimuli should belong to the same subcategory (i.e. inanimate 3D objects).

Once we established our paradigm, we sought to study the benefits of temporal regularities in visual object recognition learning. Specifically, we utilised the design of Schapiro et al. (2013) to embed a "higher-order temporal regularity" in the sequences during which the objects were presented. We established behaviourally that humans implicitly learn higher-order temporal regularities while viewing sequences of complex shapes. Even though the presence of higher-order regularity was not disclosed and instructions did not refer to it, the time-course of learning was significantly altered by it. Changes included faster learning rate and a changed order of learning individual objects. If observers noticed anything about temporal context, it was the fact that objects tended to repeat at shorter intervals.

In order to study the neural representations of individual objects and their changes associated with learning,we developed a new paradigm for multi-voxel pattern analysis in fMRI, which makes it practical to query the encoding of complex information over the entire brain. The two key ingredients of the paradigm are a comparatively

fine parcellation of grey matter voxels, and a numerically tractable analysis of spatio-temporal pattern with order of thousands dimensions (O($\sim$ 100) voxels by O($\sim$ 10) TRs). We have made available both innovations to the scientific community.

We provide evidence on decodability of complex object shapes' identity in a distributed network of parcels (124 of 758 parcels) in ventral and dorsal visual areas, and in fronto-parietal regions. This includes 70 parcels in the occipital cortex, 18 in the fusiform or temporal cortex —basically along the entire ventral occipito-temporal pathway— and 29 in the parietal cortex, and 7 in the frontal cortex.

We were able to track changes in the geometry of cortical representations as visual expertise for recurring objects was being acquired and consolidated. Following Kriegeskorte et al. (2008a), we have established gradual changes in cortical representations of complex objects in terms of the representational distances (representational similarity analysis). While the representation of objects that recurred many times and that observers attempted to memorize was broadly stable, we observed an expansion (or diversification) of the response distributions, so that all objects together scattered more uniformly over the available representational space.

For objects that appeared only once and that observers did not attempt to memorize (non-recurring objects), the results were quite different. Here, an increasing dissociation between recurring and non-recurring objects was accompanied by a substantial contraction (or stereotypisation) in the distribution of non-recurring responses, so that the representation of the non-recurring objects shifted to the margin of the representational space.

To establish cortical representations of the higher-order temporal regularities, we have modified representational similarity analysis to compare representational similarity for objects within and between different temporal communities. We have shown that representational distances are highly confounded by temporal proximity

and require a robust test controlling for such "contaminations". Our approach effectively eliminated these confounding effects, as demonstrated by comparing results recorded with and without temporal context in the presentation sequences. As the diagnostic sensitivity of our paradigm was considerably lower for temporal context than for object identity, it was not too surprising that we identified only 27 of 758 parcels with "community sensitivity".

Along the ventral occipio-temporal pathway, we identified 11 parcels in which the representations of the objects that belonged to the same temporal community were more similar than the representations of the objects in different communities (*positively* community-sensitive). This substantial anatomical overlap between cortical representations of individual objects and temporal communities is perhaps our most important finding, as it suggests that the same cortical pathway comes to represent regularities in both lower and higher temporal scales.

In the higher association cortex, including parietal, frontal, superior temporal and insular cortex, we identified 12 parcels that were *negatively* community-sensitive, in the sense that representational distances were larger within than between communities. These parcels were not selective for object shape and can be interpreted as providing "context-specific" cognitive maps that reset whenever the sensory sequence enters a new context.

Chapter 2

Visual object recognition is facilitated by temporal community structure

Brief Communication

Visual object recognition is facilitated by temporal community structure

Ehsan Kakaei,[1,2,3] Stepan Aleshin,[2,3] and Jochen Braun[2,3]

[1]European Structural and Investment Funds Graduate School on Analysis, Imaging, and Modelling of Neuronal and Inflammatory Processes , Otto-von-Guericke University, 39120 Magdeburg, Germany; [2]Institute of Biology, Otto-von-Guericke University, 39120 Magdeburg, Germany; [3]Center for Behavioral Brain Sciences, Otto-von-Guericke University, 39120 Magdeburg, Germany

Humans and others primates are highly attuned to temporal consistencies and regularities in their sensory environment and learn to predict such statistical structure. Moreover, in several instances, the presence of temporal structure has been found to facilitate procedural learning and to improve task performance. Here we extend these findings to visual object recognition and to presentation sequences in which mutually predictive objects form distinct clusters or "communities." Our results show that temporal community structure accelerates recognition learning and affects the order in which objects are learned ("onset of familiarity").

[Supplemental material is available for this article.]

Our understanding of the world is grounded in sensory experience. Typically, this experience consists of contiguous streams of sensations that are richly structured in both time and space (Schapiro and Turk-Browne 2015). Such statistical structure may involve simple correlations of pairs of sensory events or, more generally, clusters of correlations between mutually predictive events forming a "temporal community" (Schapiro et al. 2013). Both humans and other primates (Miyashita 1988) can learn to predict such statistical regularities in space and time (Fiser and Aslin 2001, 2002). Moreover, statistical structure can be exploited explicitly or implicitly to enhance task performance. For example, predictable presentation order can facilitate motor learning (Kahn et al. 2018), language learning (Saffran et al. 1996), visual search (Chun and Jiang 1998; Jiang and Wagner 2004; Sisk et al. 2019), and conditional associative learning (Hamid et al. 2010).

In general, implicit (unsupervised) learning of temporal structure is thought to provide a biological basis for important cognitive functions, including the formation of episodic memories, learning of task-sets, model-based planning, and structural learning (e.g., Kemp and Tenenbaum 2008; Rigotti et al. 2010; Gershman 2017; Russek et al. 2017). To improve experimental access to these phenomena, we sought behavioral evidence for interactions between learning at different hierarchical levels, namely, learning of individual objects and learning of the temporal context in which such objects are experienced.

Sequences of visual presentations may exhibit different kinds of temporal structure arising from sequential dependencies. A simple kind of structure is sequential dependency between consecutively presented items (i.e., an increased probability of item X, given preceding item Y). A more complex kind of structure arises when sequential dependencies are clustered within subsets of items. This leads to longer-term dependencies (i.e., an increased probability of item X, given recent item Z) and extended sequences of items that are mutually predictive (Schapiro et al. 2013; Karuza et al. 2017; Kahn et al. 2018).

The mechanisms of visual object recognition have been studied extensively (Wallis and Bülthoff 1999) with considerable evidence supporting "feature-based mechanisms" that represent three-dimensional objects in terms of multiple two-dimensional features/views (plus interpolations) (Bülthoff and Edelman 1992). Presumably, temporal regularities arise naturally in handling three-dimensional objects and help associate distinct two-dimensional views and/or features (Wallis and Bülthoff 1999). For example, when nonhuman primates learn to categorize initially unfamiliar objects, they readily form neural representations for arbitrary two-dimensional features that are diagnostic for category (Sigala and Logothetis 2002; Sigala et al. 2002). Interestingly, such representations automatically encompass predictive sequential dependencies between successive trials, even when its diagnostic information is redundant (Miyashita 1988; Wallis 1998).

The effect of sequential dependencies between successive trials on visual object recognition was investigated by two previous studies, which found a reaction time advantage (Barakat et al. 2013) and a recognition memory advantage (Otsuka and Saiki 2016) for target objects that consistently follow particular objects, compared with target objects that follow varying objects. Here we extended these findings in two ways: First, we monitored the formation of recognition memory more closely and comprehensively (every presentation of every object), and second, we considered the effect of clustered dependencies creating "temporal communities" of objects (which are typically experienced for nine successive presentations).

We investigated performance of observers in a visual object recognition learning task under three conditions: (1) "strongly structured" sequences comprising distinct temporal communities (clusters of mutually predictive objects), (2) "weakly structured" sequences with uniform sequential dependence, and (3) "random" or "unstructured" sequences without sequential dependence. All sequences were generated as random walks on graphs of $n = 15$ distinct objects (Fig. 1A), in which nodes represented distinct objects and edges represented possible transitions (in both directions). As one sequence comprised 180 object presentations, each graph was

Corresponding author: ehsankakaei91@gmail.com
Article is online at http://www.learnmem.org/cgi/doi/10.1101/lm.053306.120.

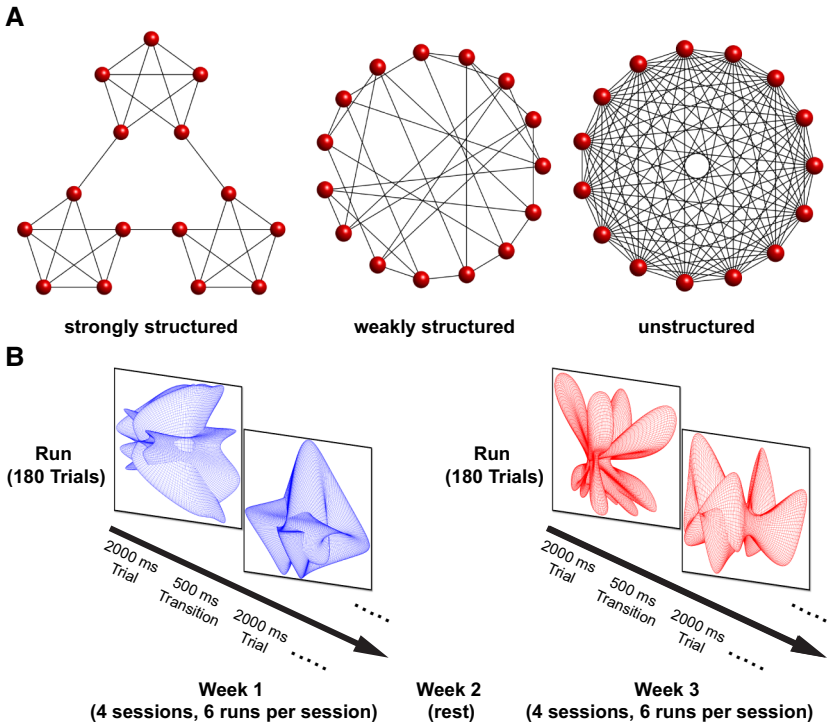

Figure 1. Presentation sequence and trial structure. (*A*) Presentation sequences were generated as (nearly) random walks on three types of graphs, with nodes representing a distinct object and edges representing possible transitions (in both directions). A sparsely connected, modular graph generated "strongly structured" sequences with distinct community structures (*left*), a sparsely connected, nonmodular graph generated "weakly structured" sequences (*middle*), and a full connected graph generated "unstructured" or "random" sequences (*right*). (*B*) Presentation sequences consisted of 180 complex, three-dimensional objects (shown rotating for 2 sec about a randomly oriented axis in the frontal plane). Of these, 170 ± 0.04 (mean ± SEM) objects were recurring, and 9.2 ± 0.04 objects were nonrecurring. Observers categorized each object as "familiar" or "unfamiliar." Over the four sessions of 1 wk, observers performed 24 runs and viewed 4320 presentations, with every recurring object appearing at least 250 times.

traversed multiple times (~11.3 times). Graphs were either modular and sparsely connected ("strongly structured" sequences), or nonmodular and sparsely connected ("weakly structured" sequences), or nonmodular and fully connected ("unstructured" or "random" sequences). In "strongly structured" sequences, approximately 9.2 ± 0.1 successive presentations (mean ± SEM) featured objects of the same temporal community.

One presentation sequence ("run") comprised exactly 180 objects and on average included 9.2 ± 0.04 (mean ± SEM) nonrecurring objects appearing exactly once during the entire experiment. Nonrecurring objects were spaced 14–19 presentations apart. The remaining 170 ± 0.04 objects were recurring and were selected by performing a pseudorandom walk on a graph (Fig. 1A), albeit with some restrictions: no direct repeats and returns were permitted (e.g., X–X or X–Y–X) and all $n = 15$ objects were repeated comparably often (11.4 ± 0.04 repetitions). The repetition latency for any given object ranged from three to >60 presentations. Very short latencies (of three to five presentations) were far more common in strongly structured sequences than in weakly structured or unstructured sequences (Supplemental Fig. S8).

To control the difficulty of shape recognition, ensure initial unfamiliarity of all objects, and minimize interference from semantic associations, we generated complex three-dimensional objects by convolving two closed Bezier curves in a plane. Complexity was controlled by number and the position of random seeds for the two curves. The pairwise dissimilarity of the resulting complex objects was statistically unrelated to their pairwise distance in the presentation sequence (see Supplemental Fig. S1). To ensure this, dissimilarity was quantified in terms of the vector distance between depth maps (of resolution 64 × 64 × 64) obtained from six viewing directions along the three principal component axes.

Objects were presented for 2 sec rotating with an angular velocity of 144 deg/sec about an axis in the frontal plane. Starting angle and axis orientation were randomized for each trial, forcing observers to become familiar with the full three-dimensional shape (rather than just certain features). Presentation periods were separated by 0.5-sec transition periods, during which the previous object disappeared toward a distant location on the right, while the next object approached from a distant location on the left. This was intended to encourage observers to imagine a spatially extended sequence of distinct objects (Supplemental_Movie_S1).

Twenty healthy observers (eight males and 12 females, aged 25 to 34 yr old) participated in three experiments. Two experiments compared "strongly structured" and "unstructured" sequences, and one experiment compared "strongly structured" and "weakly structured" sequences. All observers had normal or corrected to normal vision and were paid for their participation. Ethical guidelines of the Centre for Neuroscientific Innovation and Technology, Magdeburg, were followed.

In order to monitor the progress of recognition learning as closely as possible, observers were required to classify every object presented as either "familiar" (seen previously) or "unfamiliar" (never seen previously). For each observer, a fresh set of 30 pairwise dissimilar objects was generated. The set was divided arbitrarily into two subsets of 15 objects, one used for "structured" sequences and the other for "unstructured" sequences. In addition, we generated a larger number (~500) of nonrecurring objects, which appeared exactly once during the entire experiment. During each trial, the observer categorized the current object as "familiar," "unfamiliar," and "not sure," by pressing a key. No feedback was provided. Observers performed this task on four different days within 1 wk, with six sequences per day (24 sequences overall). Accordingly, observers viewed 4320 presentations during which every recurring object appearing at least 250 times. After pausing for a week, observers repeated the experiment with entirely new objects and with sequences generated from another graph (Fig. 1B). Observers were told that each condition used new objects that were never shown before. To further emphasize this point, object color changed between conditions. The order of conditions (structured or unstructured) was counter-balanced between observers. Observer instructions did not mention presentation order (sequence structure).

At the end of each week of testing, observers were required to additionally perform a validation task, to assess the extent to which objects had become familiar (Supplemental_Movie_S2; Supplemental_Material). In this task, observers viewed for 30 sec

an array of 12 simultaneously rotating objects, of which three were randomly selected from the 15 "recurring" objects and nine objects were entirely new (never seen before). Observers were asked to pick out the three most "familiar" objects and received binary feedback ("all correct" or "one or more incorrect"). All observers approached ceiling performance (proportion correct >0.95) in all conditions (all sequence structures), confirming that almost all recurring objects had become familiar.

To establish the progress of recognition learning, we analyzed 250 repetitions (over four sessions and 24 sequences) of every recurring object. To this end, we considered "sliding windows" with $N_w = 5$ successive presentations of a given object (for details see Supplemental Fig. S3). Note that some windows bridged successive presentation sequences and/or sessions. For each window and "recurring" object, we computed the proportion of "familiar" responses ("hit rate") (Fig. 2A). As "familiar" objects were common, some false positives were to be expected. To take this into account, we also established a "false alarm rate" for each session, as the fraction of "nonrecurring" objects not categorized as "unfamiliar" (Fig. 2B). Combining hit rate (of a window) with false alarm rate (of the concomitant session), we performed a simplified sensitivity analysis (Macmillan and Creelman 2004) to obtain a corrected classification performance ρ and decision bias b for each window and "familiar" object (see the Supplemental_Material). Alternative sensitivity analyses and performance measures (A', d'; Stanislaw and Todorov 1999) did not materially alter the results.

The resulting corrected performance ρ (mean and SEM, assuming binomial variability) is shown in Figure 2C. Performance increased nearly monotonically, but was consistently superior when objects were presented with "strongly structured" sequences with "temporal community structure" than when they were presented in unstructured sequences. This difference was significant after ~60 presentations. As expected, observers rapidly developed a liberal bias (favoring "familiar" responses), which weakened somewhat over subsequent sessions (Fig. 2D).

We also analyzed the time-development of average response times (RTs). Consistent with the performance results, RTs decreased faster for strongly structured sequences than for unstructured sequences (Supplemental_Material; Supplemental Fig. S2).

In addition to the gradual increase in the probability of recognizing recurring objects, we also sought to determine the point in time at which individual objects became familiar ("onset of familiarity"). We defined this point in two alternative ways: (1) as the first window in which corrected performance exceeded a threshold of $\rho \geq 0.875$ (high threshold approach) or (2) as the window in which entropy $H_\rho = -[\rho log_2(\rho) + (1-\rho) log_2(1-\rho)]$ of corrected performance reached its peak value (low threshold approach). Note that entropy peaks at the transition from exclusively "unfamiliar" to exclusively "familiar" responses.

After establishing the "onset of familiarity" for each object, we ranked all objects by order of onset and established the "onset separation" between object pairs in terms of onset rank (Δn) and presentation rank (Δk). The median separation of successive onsets (defined by threshold or entropy) was nine or 16 presentations, respectively. Interestingly, the median separation of successive onsets in same cluster was roughly thrice as long, with 24 and 50

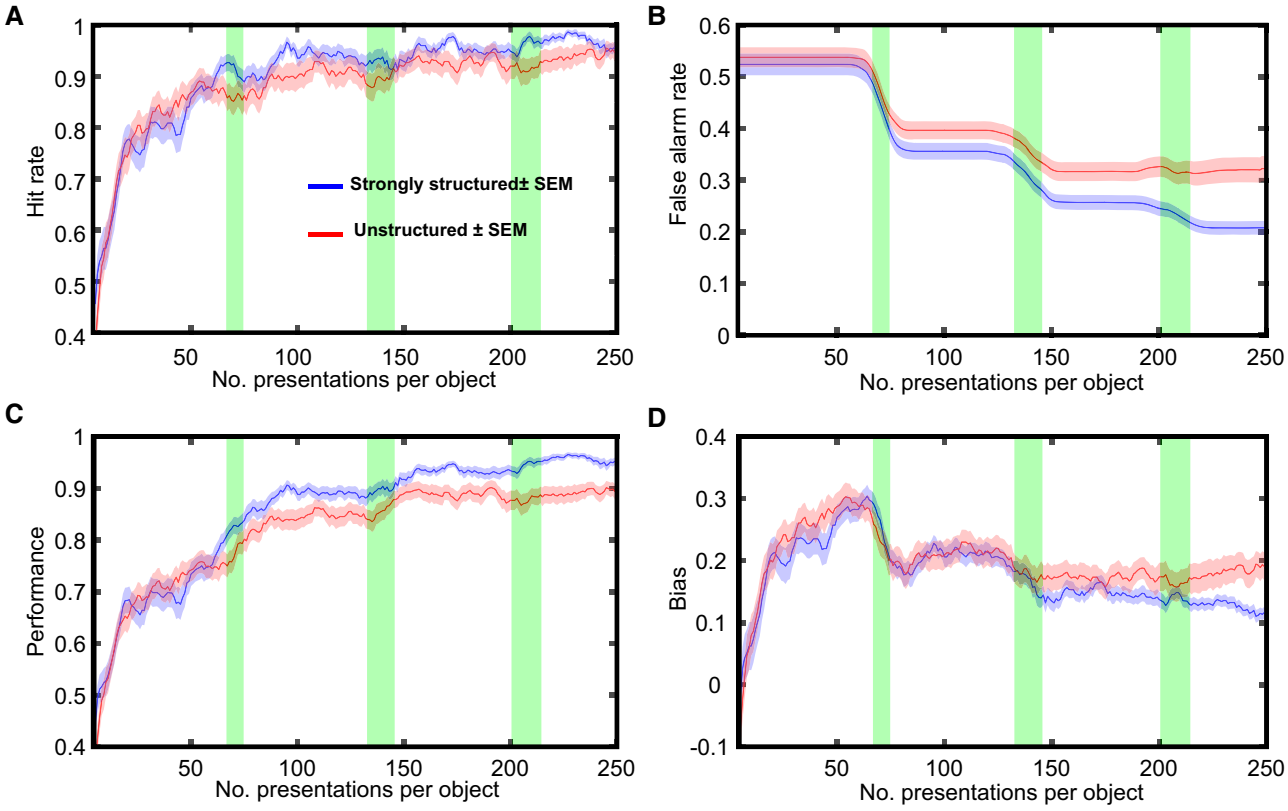

Figure 2. Time course of recognition learning. (*A*) Average hit rate (recurring categorized as familiar, per window) increases with the number of presentations of a given object. (*B*) Average false alarm rate (nonrecurring not categorized as unfamiliar, per session) decreases with the number of presentations. (*C*) Average corrected performance ρ increases nearly monotonically with presentation number. It was consistently larger for strongly structured sequences (with temporal community structure) than for unstructured sequences. (*D*) Average criterion bias b, as a function of presentation number. Green regions indicate the transition between sessions (20%–80% of objects in previous session).

presentations, respectively, implying that successive onsets occurred during separate visits to a given community.

In strongly structured sequences, one may distinguish objects pairs XY that are "adjacent" [follow each other with P(Y|X) = 0.25] or "nonadjacent" [never follow each other, P(Y|X) = 0]. In addition, one may distinguish object pairs within the same community (either adjacent or nonadjacent) and between different communities (also either adjacent or nonadjacent). Note that the objects linking different communities ("linking objects") contribute both "adjacent" pairs in different communities and "nonadjacent" pairs in the same community (Fig. 3B). We analyzed the "onset of familiarity" for different object pairs (as defined above), specifically, the probability that the two members of a pair exhibit successive onsets ($\Delta n = 1$) or nearly successive ($\Delta n = 2$) onsets. Interestingly, the probability of successive onsets was significantly higher than chance for objects in the same community (null hypothesis H_0: "onsets" are ordered randomly) (Fig. 3A). Moreover, we found the probability of successive "onsets" to be significantly elevated for "adjacent" objects in the same cluster, insignificantly elevated for "adjacent" objects in different clusters ("linking objects"), and significantly reduced for "nonadjacent" objects in different clusters ($P < 0.05$; corrected for false discovery rate of multiple comparisons) (Fig. 3B; Benjamini and Hochberg 1995).

We conclude that temporal community structure had a significant effect on the order of recognition learning in the sense that familiarity of one object in a community facilitated familiarity of another object in the same community, provided the latter was "adjacent" [i.e., followed the former sometimes, P(Y|X) = 0.25]. Interestingly, no such "domino effect" was observed for the objects linking two different communities (i.e. adjacent objects in different communities).

The results presented in Figures 2 and 3 were replicated with an additional eight observers in a second experiment of almost identical design (Supplemental Figs. S4, S6).

To dissociate the effects of cluster-membership and adjacency, we also conducted a third experiment, in which six further observers viewed either "weakly structured" presentation sequences (during 1 wk) or "strongly structured" sequences (during another week). To generate "weakly structured" sequences without temporal communities, we generated sparsely connected graphs with exactly four links per node, but without any triangular link formations (Maslov and Sneppen 2002; Rubinov and Sporns 2010). Recognition learning was faster for "strongly structured" sequences than for "weakly structured" ones. The "domino" effect described above was again observed for "strongly structured" sequences (with both "onset" definitions), but to some extent also for "weakly structured" sequences (for one "onset" definition). Thus, the ordering of "onsets" of familiarity may be affected both by community membership and by adjacency in the presentation sequence (Supplemental Figs. S5, S7).

In this study, we investigated the effect of temporal community structure by comparing more or less structured presentation sequences. First, in "weakly structured" sequences, sparse connectivity of the generative graph ensured that each object predicted the next object with 25% probability (one of four possibilities). Second, in "strongly structured" sequences, the (equally sparse) generative graph was clustered into three communities of five objects, so that each object predicted the community membership of the next object with 90% probability (18 of 20 possibilities).

Previous studies of statistical learning did not aim to closely follow the learning of individual items (Siegelman et al. 2018). Here we sought to monitor the degree of familiarity of each individual object over successive presentations (Fig. 2). Whereas classification performance improved monotonically with presentation number for all sequences, a significant performance advantage developed quickly (over 60 to 70 presentations) for "strongly structured" sequences compared with either "unstructured" or "weakly structured" sequences (Supplemental Fig. S5). Note that recognition performance improved comparably over time, with or without having practiced stimulus-response mapping in a separate training session (experiments 2 and 3). Accordingly, we do not believe that motor learning contributed appreciably to these results.

Thanks to close monitoring, we could almost always determine the onset of familiarity for an individual object. Interestingly, the ordering of onsets did not appear to be fully random, in that objects of the same community ("temporal community") tended to become familiar after one another more often than expected by chance. Interestingly, this "domino effect" typically did not occur within one "extended visit" to a community but over subsequent visits to a given community. This "domino effect" was particularly pronounced for adjacent objects in the same

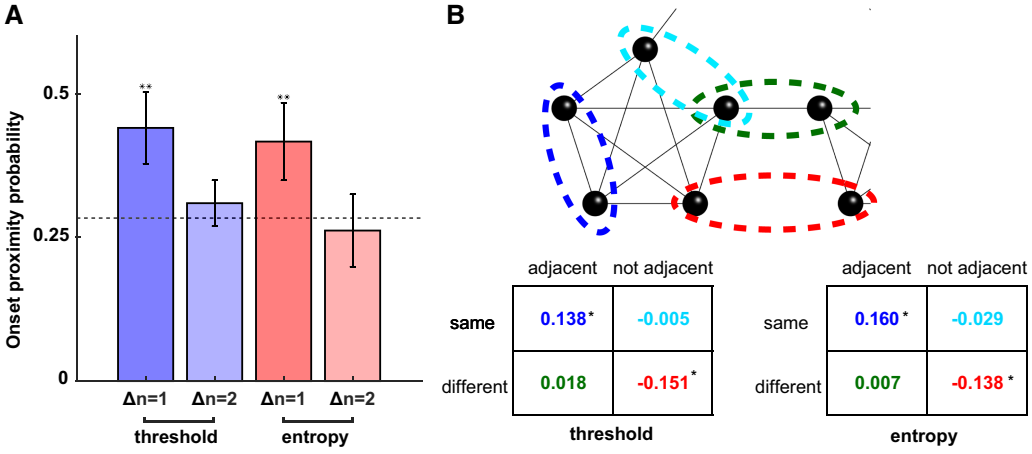

Figure 3. Analysis of the onset of familiarity with individual objects. (*A*) Successive onsets of familiarity ($\Delta n = 1$) are far more likely ([**] $P < 0.005$) for objects within the same cluster than would be expected by chance (dashed line). For nearly successive onsets ($\Delta n = 2$) this effect was not observed. (*B*) Comparison of frequency of successive onsets, compared with chance level, for objects pairs either in the same cluster (outlined blue and cyan) or in different clusters (green and red), which are either adjacent (blue and green) or nonadjacent on the graph (cyan and red). Frequency is significantly elevated ([*] $P < 0.05$ FDR corrected) for adjacent objects in the same cluster (blue) and suppressed for nonadjacent objects in different clusters (red).

community, but was not observed for adjacent objects in different communities. As a similar effect was observed for adjacent objects in "weakly structured" sequences without communities, there seems to be a contribution of frequent temporal proximity.

At the end of training, all objects had become familiar and could be retrieved explicitly from long-term memory, for both structured and unstructured sequences. The reason for the observed difference in learning rates remains unclear. One possibility is that structured sequences pose a reduced working-memory load, facilitating encoding and accelerating learning. When large sets of items are divided ("chunked") into subsets, both chunked and nonchunked items benefit and are learned more readily. Presumably, chunking reduces the dimensionality of the classification problem presented by each item (just like chunking the search array in an odd-man-out task reduces the dimensionality of target detection). This reduced dimensionality could then lower working-memory load and facilitate classification by comparison with long-term memory for both familiar (chunked) items and unfamiliar (nonchunked) items. Another important factor might be that temporal communities reduce repetition latencies (Supplemental Fig. S8). There is evidence that timely repetitions help consolidate memories, whereas delayed repetitions leave memories prone to disruption (Thalmann et al. 2019).

Previous studies of the effect of "temporal community structure" have shown that cluster borders are detectable (Schapiro et al. 2013) and that such borders elevate reaction time (Kahn et al. 2018; Karuza et al. 2019). As border items are thought to facilitate encoding/retrieval (Swallow et al. 2009), one might have expected accelerated recognition learning for "linking objects" that join two different clusters. However, in our paradigm, neither learning rate nor ordering of onsets of familiarity distinguished "linking objects" from other objects. In fact, our results suggest that any chunking benefits (Thalmann et al. 2019) apply more to objects within clusters than to objects that "link" clusters.

In summary, we showed that the presence of temporal communities of mutually predictive objects accelerates recognition learning for complex, three-dimensional objects and alters the order of recognition learning such that members of a group are often learned after one another (but separated by many intervening presentations).

Acknowledgments

The project was funded by the federal state Saxony-Anhalt and the European Structural and Investment Funds (ESF, 2014-2020), project number ZS/2016/08/80645.

References

Barakat BK, Seitz AR, Shams L. 2013. The effect of statistical learning on internal stimulus representations: predictable items are enhanced even when not predicted. *Cognition* **129:** 205–211. doi:10.1016/j.cognition.2013.07.003

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B (Methodol)* **57:** 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bülthoff HH, Edelman S. 1992. Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc Natl Acad Sci* **89:** 60–64. doi:10.1073/pnas.89.1.60

Chun MM, Jiang Y. 1998. Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn Psychol* **36:** 28–71. doi:10.1006/cogp.1998.0681

Fiser J, Aslin RN. 2001. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci* **12:** 499–504. doi:10.1111/1467-9280.00392

Fiser J, Aslin RN. 2002. Statistical learning of higher-order temporal structure from visual shape sequences. *J Exp Psychol Learn Mem Cogn* **28:** 458–467. doi:10.1037/0278-7393.28.3.458

Gershman SJ. 2017. Predicting the past, remembering the future. *Curr Opin Behav Sci* **17:** 7–13. doi:10.1016/j.cobeha.2017.05.025

Hamid OH, Wendemuth A, Braun J. 2010. Temporal context and conditional associative learning. *BMC Neurosci* **11:** 45. doi:10.1186/1471-2202-11-45

Jiang Y, Wagner LC. 2004. What is learned in spatial contextual cuing: configuration or individual locations? *Percept Psychophys* **66:** 454–463. doi:10.3758/BF03194893

Kahn AE, Karuza EA, Vettel JM, Bassett DS. 2018. Network constraints on learnability of probabilistic motor sequences. *Nat Hum Behav* **2:** 936–947. doi:10.1038/s41562-018-0463-8

Karuza EA, Kahn AE, Thompson-Schill SL, Bassett DS. 2017. Process reveals structure: how a network is traversed mediates expectations about its architecture. *Sci Rep* **7:** 1–9. doi:10.1038/s41598-017-12876-5

Karuza EA, Kahn AE, Bassett DS. 2019. Human sensitivity to community structure is robust to topological variation. *Complexity* **2019:** 8379321. doi:10.1155/2019/8379321

Kemp C, Tenenbaum JB. 2008. The discovery of structural form. *Proc Natl Acad Sci* **105:** 10687–10692. doi:10.1073/pnas.0802631105

Macmillan NA, Creelman CD. 2004. *Detection theory: a user's guide*, 2nd ed. Lawrence Erlbaum Associates, Mahwah, NJ.

Maslov S, Sneppen K. 2002. Specificity and stability in topology of protein networks. *Science* **296:** 910–913. doi:10.1126/science.1065103

Miyashita Y. 1988. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* **335:** 817–820. doi:10.1038/335817a0

Otsuka S, Saiki J. 2016. Gift from statistical learning: visual statistical learning enhances memory for sequence elements and impairs memory for items that disrupt regularities. *Cognition* **147:** 113–126. doi:10.1016/j.cognition.2015.11.004

Rigotti M, ben Dayan Rubin D, Morrison SE, Salzman CD, Fusi S. 2010. Attractor concretion as a mechanism for the formation of context representations. *Neuroimage* **52:** 833–847. doi:10.1016/j.neuroimage.2010.01.047

Rubinov M, Sporns O. 2010. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52:** 1059–1069. doi:10.1016/j.neuroimage.2009.10.003

Russek EM, Momennejad I, Botvinick MM, Gershman SJ, Daw ND. 2017. Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Comput Biol* **13:** e1005768. doi:10.1371/journal.pcbi.1005768

Saffran JR, Aslin RN, Newport EL. 1996. Statistical learning by 8-month-old infants. *Science* **274:** 1926–1928. doi:10.1126/science.274.5294.1926

Schapiro AC, Turk-Browne NB. 2015. Statistical learning. In *Brain mapping: an encyclopedic reference* (ed. Toga AW), pp. 501–506. Academic Press, New York.

Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, Botvinick MM. 2013. Neural representations of events arise from temporal community structure. *Nat Neurosci* **16:** 486–492. doi:10.1038/nn.3331

Siegelman N, Bogaerts L, Kronenfeld O. 2018. Redefining 'learning' in statistical learning: what does an online measure reveal about the assimilation of visual regularities? *Cogn Sci* **42:** 692–727. doi:10.1111/cogs.12556

Sigala N, Logothetis NK. 2002. Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* **415:** 318–320. doi:10.1038/415318a

Sigala N, Gabbiani F, Logothetis NK. 2002. Visual categorization and object representation in monkeys and humans. *J Cogn Neurosci* **14:** 187–198. doi:10.1162/089892902317236830

Sisk CA, Remington RW, Jiang YV. 2019. Mechanisms of contextual cueing: a tutorial review. *Atten Percept Psychophys* **81:** 2571–2589. doi:10.3758/s13414-019-01832-2

Stanislaw H, Todorov N. 1999. Calculation of signal detection theory measures. *Behav Res Methods Instrum Comput* **31:** 137–149. doi:10.3758/BF03207704

Swallow KM, Zacks JM, Abrams RA. 2009. Event boundaries in perception affect memory encoding and updating. *J Exp Psychol Gen* **138:** 236–257. doi:10.1037/a0015631

Thalmann M, Souza AS, Oberauer K. 2019. How does chunking help working memory? *J Exp Psychol Learn Mem Cogn* **45:** 37–55. doi:10.1037/xlm0000578

Wallis G. 1998. Temporal order in human object recognition learning. *J Biol Syst* **6:** 299–313. doi:10.1142/S0218339098000200

Wallis G, Bülthoff H. 1999. Learning to recognize objects. *Trends Cogn Sci* **3:** 22–31. doi:10.1016/s1364-6613(98)01261-3

Received November 24, 2020; accepted in revised form February 13, 2021.

Visual object recognition is facilitated by temporal community structure

Ehsan Kakaei, Stepan Aleshin and Jochen Braun

Learn. Mem. 2021, **28:**
Access the most recent version at doi:10.1101/lm.053306.120

Chapter 3

Gradual change of cortical representations with growing visual expertise for synthetic shapes

Imaging Neuroscience
an open access journal

Research Article

Gradual change of cortical representations with growing visual expertise for synthetic shapes

Ehsan Kakaei[a,b,c], Jochen Braun[b,c]

[a]European Structural and Investment Funds Graduate School on Analysis, Imaging, and Modelling of Neuronal and Inflammatory Processes, Otto-von-Guericke University, Magdeburg, Germany
[b]Institute of Biology, Otto-von-Guericke University, Magdeburg, Germany
[c]Center for Behavioral Brain Sciences, Otto-von-Guericke University, Magdeburg, Germany

Corresponding Author: Ehsan Kakaei (ehsankakaei91@gmail.com)

ABSTRACT

Objective: Visual expertise for particular categories of objects (e.g., mushrooms, birds, flowers, minerals, and so on) is known to enhance cortical responses in parts of the ventral occipitotemporal cortex. How is such additional expertise integrated into the prior cortical representation of life-long visual experience? To address this question, we presented synthetic visual objects rotating in three dimensions and recorded multivariate BOLD responses as initially unfamiliar objects gradually became familiar.

Main results: An analysis of pairwise distances between multivariate BOLD responses ("representational similarity analysis," RSA) revealed that visual objects were linearly discriminable in large parts of the ventral occipital cortex, including the primary visual cortex, as well as in certain parts of the parietal and frontal cortex. These cortical representations were present from the start, when objects were still unfamiliar, and even though objects were shown from different sides. As shapes became familiar with repeated viewing, the distribution of responses expanded to fill more of the available space. In contrast, the distribution of responses to novel shapes (which appeared only once) contracted and shifted to the margins of the available space.

Conclusion: Our results revealed cortical representations of object shape and gradual changes in these representations with learning and consolidation. The cortical representations of once-viewed shapes that remained novel diverged dramatically from repeatedly viewed shapes that became familiar. This disparity was evident in both the similarity and the diversity of multivariate BOLD responses.

Keywords: object recognition, visual expertise, functional imaging, representational similarity, and multi-voxel activity

1. INTRODUCTION

An essential aspect of visual object recognition is the processing of visual shapes. The neural substrate of shape processing includes the ventral visual pathway, which in humans extends over the ventral occipitotemporal cortex from the occipital pole to the lateral occipital cortex, fusiform gyrus, and beyond (reviewed by Bi et al., 2016; Grill-Spector & Weiner, 2014; Kravitz et al., 2013; Weiner & Zilles, 2016). Functional imaging studies of ventral occipitotemporal cortex reveal intriguing functional anatomy, with responsiveness to specific object categories (e.g., faces, scenes, body parts) changing systematically over the cortical surface along several large-scale anatomical gradients (e.g., animate-inanimate, large-small, feature-whole, or perception-action; Freud et al., 2017; Grill-Spector & Weiner, 2014; Grill-Spector et al., 2004; Konkle & Oliva, 2012; Wurm & Caramazza, 2022; Yildirim et al., 2019).

Received: 11 June 2023 Revision: 4 June 2024 Accepted: 2 July 2024 Available Online: 22 July 2024

Experience and learning improve object recognition performance, and also modify shape processing in the ventral occipitotemporal cortex. Indeed, functional imaging evidence shows that particular visual expertise—being able to identify and categorize visually similar objects of a particular kind—often entails moderate but anatomically distributed changes in the pre-existing responsiveness to shape (reviewed by Bukach et al., 2006; de Beeck & Baker, 2010; Gauthier & Tarr, 2016; Harel et al., 2013). This has been established by comparing novices and experts for identifying particular categories of natural objects (e.g., birds, mushrooms, minerals, degraded images; Cetron et al., 2019; Connolly et al., 2012; Duyck et al., 2021; Freud et al., 2017; Martens et al., 2018; McGugin et al., 2012; Roth & Zohary, 2015), as well as by comparing observers before and after they have learned to categorize initially unfamiliar synthetic shapes (e.g., computer-generated "greebles," "spikies," or "ziggerins"; Brants et al., 2011; de Beeck et al., 2006; Gauthier et al., 1999; A. C.-N. Wong et al., 2009; Y. K. Wong et al., 2012; Yue et al., 2006).

Here, we map the cortical representation of synthetic visual objects and track gradual changes as initially unfamiliar objects become progressively familiar with learning. We wondered how pre-existing shape representations would accommodate and integrate novel synthetic objects. We further wondered whether representational changes would be specific to learned objects or extend also to other objects of the same kind. To explore these questions, we analyzed "representational similarity" of spatiotemporal BOLD patterns (Haxby, 2012; Kriegeskorte, Mur, Ruff, et al., 2008), which offers a potentially sensitive measure for the information encoded in neural activity and may also be related to similarity as perceived by human observers (Charest & Kriegeskorte, 2015; Collins & Behrmann, 2020; Nestor et al., 2016).

Most previous studies of visual expertise identified cortical sites associated with a particular object category by comparing BOLD activity either between novices and experts or before and after learning. We extend this work in three ways: firstly, by establishing representational distance at the level of object exemplars rather than object categories; secondly, by monitoring gradual changes as observers gain familiarity with object exemplars; and thirdly, by analyzing changes in the diversity of multivariate BOLD activity. Few previous studies have attempted to resolve shape representations in such detail (Brants et al., 2016; Duyck et al., 2021; Eger et al., 2008; Visconti di Oleggio Castello et al., 2021). To progress fine-grained analysis of representational geometry, we developed synthetic shapes for which visual expertise is acquired comparatively slowly (Kakaei et al., 2021) and took advantage of a numerically tractable method for linear discriminant analysis in O(10^3)-dimensional multivariate activity (DLDA; Yu & Yang, 2001).

Our results showed view-invariant representations of shape over surprisingly extensive regions of the ventral occipitotemporal cortex, including the fusiform gyrus, lateral occipital areas, and primary visual cortex. Representational distances were high from the start, even before learning, suggesting that new visual expertise was accommodated and encoded within pre-existing representations. However, shapes that appeared repeatedly (and were memorized by observers) and shapes that appeared just once (and were ignored) diverged dramatically, in terms of their cortical representations, while visual expertise was being acquired and consolidated.

2. METHODS

2.1. Observers and behavior

Eight healthy observers (4 female and 4 male; aged 25 to 32 years) took part in behavioral training ("sham experiment," one session per observer), the functional imaging experiment ("main experiment," six scanning sessions per observer), and a final behavioral assessment (two sessions). All observers were paid and gave informed consent. Ethical approval was granted under Chiffre 30/21 by the ethics committee of the Faculty of Medicine of the Otto-von-Guericke University, Magdeburg.

In both sham and main experiments, observers viewed sequences of 200 recurring and non-recurring objects (see below and Fig. 1A) and attempted to classify each object as "familiar" or "novel" (by pressing the appropriate button). Over the course of multiple sessions, observers gradually became familiar with recurring objects and thus became able to distinguish them from non-recurring objects. Objects of the sham experiment were two-dimensional shapes, whereas objects of the main experiment were rotating, three-dimensional shapes (see below and Fig. 1A).

The main experiment extended over 3 successive weeks, with three sessions on separate days of both the 1st and 3rd week (no sessions took place in the 2nd week). The experiments of the 1st and 3rd week differed in four aspects: sequence type (structured or unstructured), the set of recurring objects, object color (red or blue), and responding hand (left or right). All aspects were counterbalanced across observers.

After the three scanning sessions of a week, observers participated in an additional behavioral session to confirm that they had in fact become familiar with every recurring object. Specifically, they performed a spatial search task in which they pointed out recurring target objects among non-recurring distractor objects (Kakaei

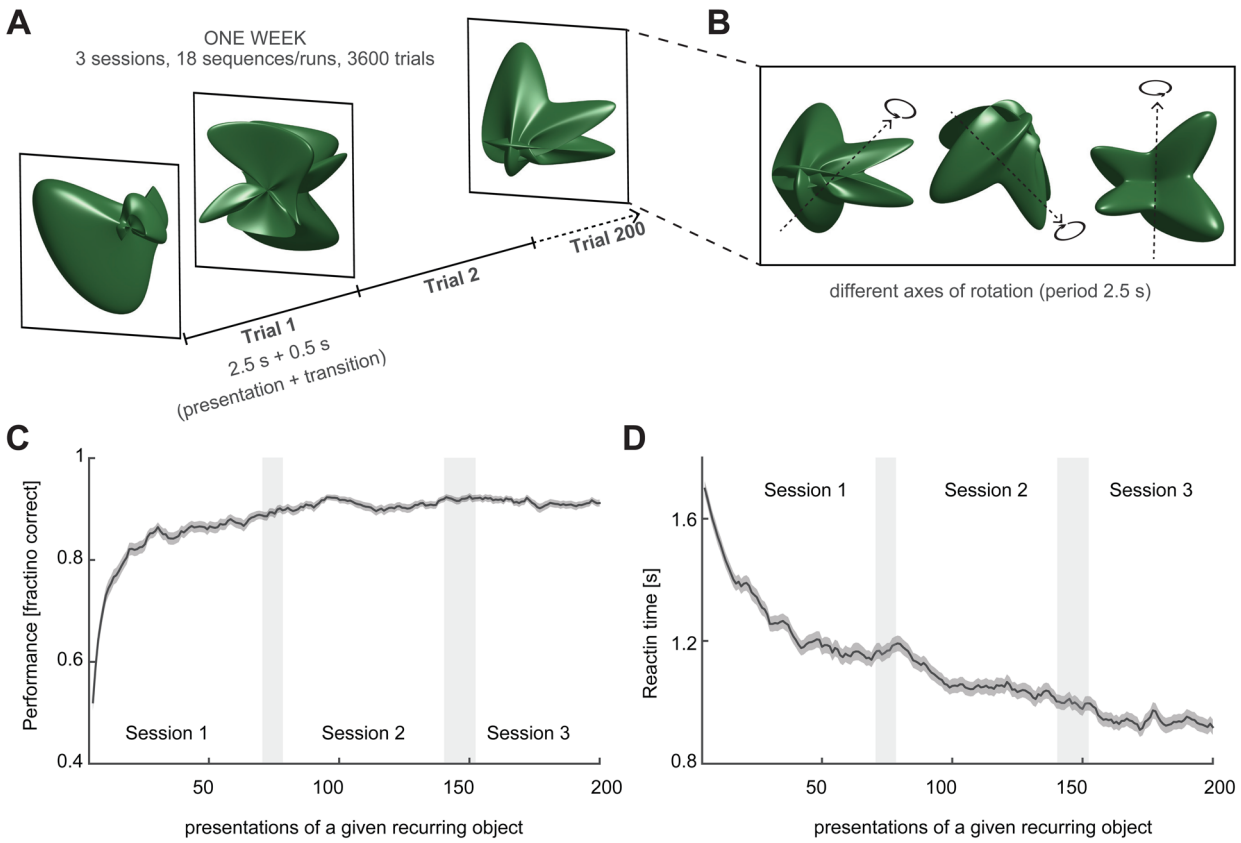

Fig. 1. Experimental paradigm. (A) Complex objects were shown for 2.5 s each, separated by 0.5 s transition, in sequences of 200 presentations, with a total duration of 600 s. Over 1 week, observers participated in 3 sessions, viewing 6 sequences during each session (18 sequences in total). Fifteen objects appeared many times each ("recurring objects"), while other objects appeared exactly once ("non-recurring objects"). Observers were required to categorize each object as either "familiar" or "unfamiliar" (by button press). (B) Objects appeared randomly rotated and revolved for one full turn (clockwise or counter-clockwise about variable axes in the frontal plane, inclination of 0°, 45°, or –45°). (C) Over the course of the week, as observers became familiar with recurring objects, classification performance improved. Here, performance (average and S.E.M.) is shown as a function of the number of presentations for 15 recurring objects, 8 observers, and 2 conditions. The relation between presentations and sessions was probabilistic (indicated by gray shading). (D) Reaction time (average and S.E.M.) as a function of presentation number. With increasing familiarity, reaction times decrease by 50% (from 1.7 *s* to 0.9 *s*) and become considerably shorter than the presentation time.

et al., 2021). In addition, observers were offered the opportunity to voice anything they might have noticed about the experiment.

2.2. Experimental paradigm

Complex three-dimensional objects were computer-generated and presented as described previously (Kakaei et al., 2021). A movie can be viewed under this LINK. All objects were highly characteristic and dissimilar from each other as confirmed computationally in terms of vector distances between depth maps (Kakaei et al., 2021). Objects were presented every 3*s*, with 2.5*s* viewing and 0.5*s* transition time (Fig. 1A). Objects were shown from all sides and, after appearing at an arbitrary angle, revolved smoothly for one full turn (period 2.5*s*, frequency 0.4 *Hz*, angular frequency 144°/*s*) about one of several axes in the frontal plane (–45°, 0°, 45°, clockwise or counter-clockwise). Axes and directions were counterbalanced for each object, and initial viewing angles were chosen randomly (Fig. 1B). All stimuli were generated with MATLAB (The MathWorks, Inc.), presented with the psychophysics toolbox (Brainard, 1997), and viewed in a mirror mounted to the MR head coil (screen resolution 960×720 pixels, frame rate 60 *Hz*, subtending approximately 8°×6° of visual angle, average luminance 50 Cd/m^2, background luminance 5 Cd/m^2). Observers responded with the right or left index finger on an MR-safe response box.

Fifteen objects recurred many times during three sessions ("recurring" objects), whereas other objects appeared exactly once ("non-recurring" or "singular" objects). As mentioned, observers classified every object as either "familiar" or "unfamiliar" by pressing a button during its presentation. Over the course of three sessions,

all observers gradually became familiar with the "recurring objects" (see below). The average time-course of learning, as established by a simplified signal detection and reaction-time (RT) analysis, is shown in Figure 1C.

Every session comprised six sequences ("runs"), each lasting 600*s* and presenting 180 "recurring" and 20 "non-recurring" objects (200 objects in total). As there were 15 different recurrent objects, each such object was seen 12 ± 1.9 times during every sequence. Over the three sessions (or 18 sequences), each recurring object appeared at least 190 times each (mean $\pm$ S.D.: 216 ± 9), whereas non-recurring objects appeared only once. Altogether, there were 3, 240 presentations of recurring objects ($3 \times 6 \times 180$) and 360 presentations of non-recurring objects ($3 \times 6 \times 180$).

Presentation sequences started with a random recurring object and continued randomly to one of the possible next objects, with neither immediate repetitions ($X \rightarrow X$) nor direct returns ($X \rightarrow Y \rightarrow X$) being allowed. Sequences comprised 200 objects, of which 180 were recurring and 20 objects non-recurring and were interspersed at random intervals. Object sequences were post-selected such as to counterbalance the number of appearances of every recurring object in every session.

All observers performed the experiment twice in the scanner, once during the 1st week and again during the 3rd week of the main experiment (so that 8 observers provided 16 data sets). As mentioned, the 2 weeks differed in terms of the recurring objects and the presentation sequence. "Structured" sequences exhibited predictive sequential dependencies (3 possible recurring next objects), whereas "unstructured" sequences did not (14 possible recurring next objects, see Kakaei et al., 2021 for details). As a result, the repetition latency (i.e., the latency of successive presentations of the same object) was 5.5 ± 15 (median and S.D.) for "structured" and 10.5 ± 11 for "unstructured" sequences. Further aspects and effects of sequence structure are reported and discussed in detail in a companion paper.

To verify that recurring objects had become familiar to observers, every observer performed 60 trials of a spatial search task with 3 recurring and 9 non-recurring objects. The 12 objects were positioned randomly in a 3×4 array and were presented for 30 s while rotating in three dimensions (as in the main experiment). After each presentation, observers indicated the recurring object positions with the computer mouse. Performance was consistently above 95% correct.

2.3. MRI acquisition

All magnetic-resonance images were acquired on a 3T Siemens Prisma scanner with a 64-channel head coil. Structural images were T1-weighted sequences (**MPRAGE** TR = 2,500 ms, TE = 2.82 ms, TI = 1,100 ms, 7° flip angle, isotropic resolution $1 \times 1 \times 1$ *mm* and matrix size of $256 \times 256 \times 192$). Functional images were T2*-weighted sequences (TR = 1,000 ms, TE = 30 ms, 65° flip angle, resolution of $3 \times 3 \times 3.6$ *mm* and matrix size of $72 \times 72 \times 36$). Field maps were obtained by gradient dual-echo sequences (TR = 720 ms, TE1 = 4.92 ms, TE2 = 7.38 ms, resolution of $1.594 \times 1.594 \times 2$ *mm* and matrix size of $138 \times 138 \times 72$).

2.4. fMRI pre-processing

Our approach to fMRI analysis was influenced by recent advances in comparing uni- and multivariate responses of corresponding voxels between different observers (e.g., Kumar et al., 2022; Nastase et al., 2019). The *local* correlation structure of voxel response, which is similar in different observers, provided the basis for our functional parcellation (Dornas & Braun, 2018). The parcellation obviated "searchlight" strategies by defining for all observers corresponding brain "parcels" with corresponding episodes of high-dimensional (O(1000)) multivariate activity.

The fMRI pre-processing procedure was similar to that published previously (Dornas & Braun, 2018). First, DICOM files were converted into NIFTI format using MRICRON (MRICRON Toolbox, Maryland, USA, NIH). Then, brain tissues were extracted and segmented using BET (Smith, 2002) and FAST (Zhang et al., 2001). Field map correction, head motion correction, spatial smoothing, high-pass temporal filtering, and registration to structural and standard images were performed with the MELODIC package of FSL (Beckmann & Smith, 2004).

Field map correction and registration to structural image were carried out using Boundary-Based Registration (BBR; Greve & Fischl, 2009). MELODIC uses MCFLIRT (Jenkinson et al., 2002) to correct for head motion. Spatial smoothing was performed with SUSAN (Smith & Brady, 1997), with full width at half maximum set at FWHM = 5 *mm*. To remove low-frequency artifacts, we applied a high-pass filter of the cut-off frequency $f = 0.01$ Hz, that is, oscillations/events with periods of more than 100 s were removed. To register the structural image to Montreal MNI152 standard space with isotropic 2 *mm* voxel size, we used FLIRT (FMRIB's Linear Image Registration Tool; Jenkinson & Smith, 2001; Jenkinson et al., 2002) with 12 degrees of freedom (DOF) and FNIRT (FMRIB's Nonlinear Image Registration Tool) to apply the non-linear registration. To further reduce artifacts arising from head motion, we applied despiking with a threshold of $\lambda = 100$ using BrainWavelet toolbox (Patel et al., 2014). Later, we regressed out the mean CSF activity as well as

12 DOF translation and rotation factors predicted by a motion correction algorithm (MCFLIRT). Afterward, the time series of each voxel was detrended linearly and whitened (with Matlab functions "detrend" and "zscore").

Finally, the 160, 099 voxels of MNI152 space were grouped into 758 functional parcels according to the MD758 atlas (Dornas & Braun, 2018). Each functional parcel is associated with an anatomically labeled region of the AAL atlas (Tzourio-Mazoyer et al., 2002) and comprises approximately 200 voxels or approximately $1.7cm^3$ of gray matter volume (212 ± 70 voxels, range 45 to 462 voxels). Parcels were defined for a small population of observers such as to maximize signal covariance *within* and minimize covariance *between* parcels in the resting state. In contrast to other parcellation schemes, this was based exclusively on the (typically strong) functional correlations within each anatomical region and disregarded the (typically weak) correlations between different anatomical regions. The MD758 parcellation offers superior cluster quality, correlational structure, sparseness, and consistency with fiber tracking, compared to other parcellation schemes of similar resolution (Albers et al., 2021; Dornas & Braun, 2018).

2.5. fMRI data analysis

To study the neural representation of objects, we extracted the multivoxel activity pattern at $N_t = 9$ time points following object onset. In a functional parcel with N_{vox} voxels, this response pattern constituted a point (or vector) in an N_{dim}-dimensional space, where $N_{dim} = N_t \cdot N_{vox}$ (Fig. 2A). To identify parcels with significant selectivity for individual recurring objects, we employed a representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008) (Fig. 2B). This analysis uses the standardized Euclidean (Mahalanobis) distance between responses in a high-dimensional space to examine the separability of neural object representations as a function of learning, or object type (recurring or non-recurring), or both. Over all 758 parcels, response dimensionality was $N_{dim} = 1{,}911 \pm 634$ (mean and standard-deviation), with a range from 405 (Calcarine-L 329, with 45 voxels) to 4,113 (Postcentral-R-484, with 457 voxels).

Our approach to RSA differed from previous work in some respects. Firstly, we analyzed high-dimensional *spatiotemporal* patterns of BOLD activity (200 voxels $\times$ 9 s, or O(10^3) dimensions) in non-overlapping gray matter volumes (758 functional subdivisions of 90 anatomical regions, averaging 1.7 cm^3; Dornas & Braun, 2018). Other studies have used lower-dimensional *spatial* activity patterns in overlapping searchlight volumes ($O(10^2)$ voxels or dimensions, covering 0.25 to 1.0 cm^3; Kriegeskorte et al., 2006). Secondly, we employed multi-class linear discriminant analysis ("direct linear discriminant analysis," DLDA; Yu & Yang, 2001), rather than pairwise discriminability or one-versus-all discriminability (e.g., Hung et al., 2005; Liu et al., 2009). With these modifications, RSA revealed representational geometry at the level of object exemplars, as well as gradual changes in this geometry over sessions and runs.

2.5.1. Linear discriminant analysis

To analyze the response variance that discriminates $\kappa = 15$ recurring objects, at most $(\kappa - 1)$-dimensions are required. Restricting the analysis to 14 principal components of the response could potentially have neglected smaller but more discriminating components. Accordingly, we performed a Linear Discriminant Analysis (LDA), which amounts to a "supervised" principal component analysis (PCA) and yields the $(\kappa - 1)$-dimensional orthonormal subspace $\mathbb{S}$ that optimally discriminates the κ response classes. Here, optimality is defined as simultaneously minimizing within-class variance and maximizing between-class variance of responses.

The results of LDA and PCA showed considerable commonality. Over the 758 parcels, the first 14 principal components captured $53 \pm 7\%$ (mean and S.D.) of the total response variance, whereas the 14-dimensional subspaces $\mathbb{S}$ captured $33 \pm 7\%$ of the total variance (or $61 \pm 6\%$ of the principal component variance). Almost all of the subspace variance overlapped with the principal component variance (i.e., $88 \pm 5\%$ of subspace variance projected into the space of the first 14 principal components, while the remaining $12 \pm 5\%$ projected into the space of the remaining principal components).

Similar numbers were obtained for the 124 identity-selective parcels. The first 14 principal components captured $57 \pm 6\%$ (mean and S.D.) of the total response variance, and subspaces $\mathbb{S}$ captured $38 \pm 6\%$ of the total variance (or $67 \pm 4\%$ of the principal component variance). Almost all of the subspace variance ($91 \pm 3\%$) overlapped with the first 14 principal components. In summary, Linear Discriminant Analysis captured the useful (discriminating) part of correlated variance and distributed this variance more uniformly over its 14 orthonormal dimensions ($6 \pm 3\%$ per dimension) than principal component analysis could ($4 \pm 6\%$ per dimension).

A numerically tractable procedure for identifying the optimal subspace $\mathbb{S}$ is available in terms of "direct LDA" or DLDA (Ye et al., 2006; Yu & Yang, 2001). Briefly, this method first diagonalizes between-class variance to identify $\kappa - 1$ discriminative eigenvectors with non-zero eigenvalues, next diagonalizes within-class variance, and finally yields a rectangular matrix for projecting activity

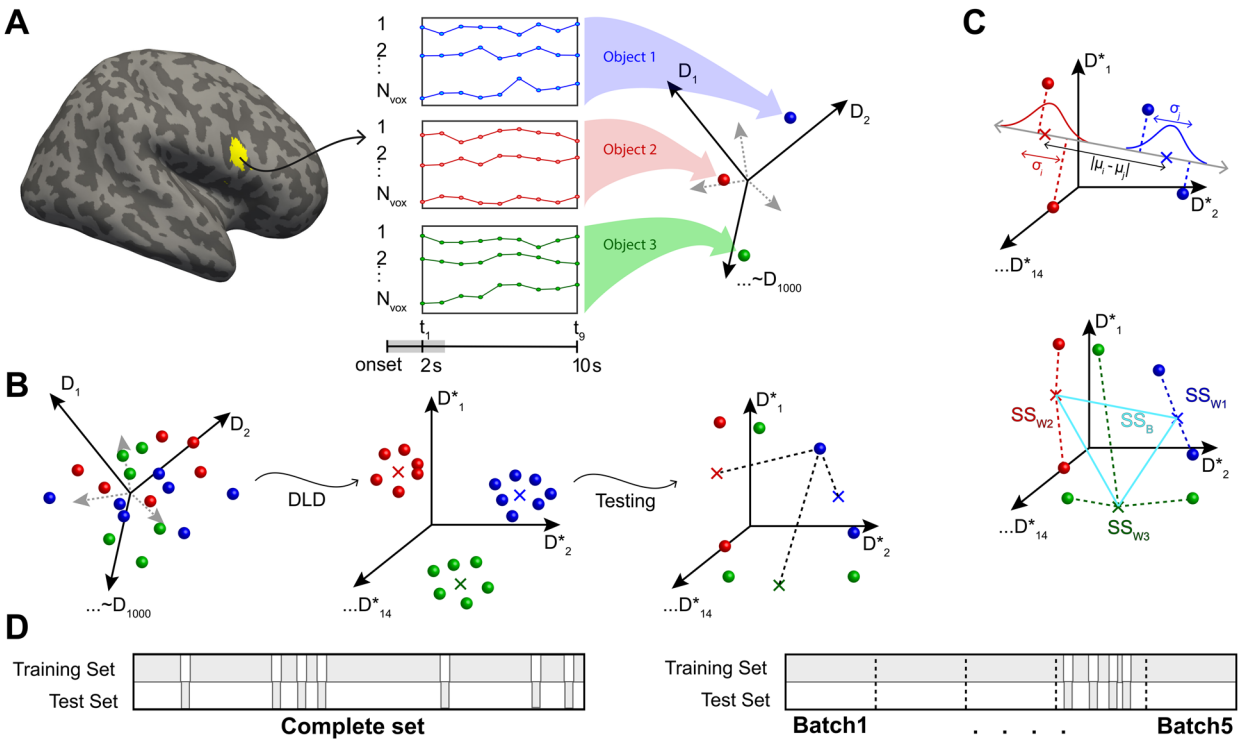

Fig. 2. Analysis of fMRI activity with direct linear discriminant analysis, or DLDA. For each functional parcel, DLDA identified the 14-dimensional space that optimally discriminated the 15 classes of activity patterns associated with 15 *recurring* objects. Other activity patterns, such as those associated with *nonrecurring* objects, were also analyzed in this space. (A) For a given parcel with N_{vox} voxels (e.g., yellow region Frontal-Inf-R-8), activity was recorded over 9 *s* during and following object presentation (2 to 11 *s* after onset). Each such activity pattern corresponds to a point (or vector) in a $9 \cdot N_{vox}$-dimensional space (right). Here, activity patterns associated with three object presentations are represented schematically (red, green, and blue spheres). (B) To cross-validate discriminability, recurrent object presentations were divided randomly into a training set (90%) and a test set (10%). From the training set, the DLD subspace $\mathbb{S}$ was established. Here, exemplars (solid spheres) and class centroids (crosses) are represented schematically. Next, the projections into this space of test set patterns were compared to class centroids. (C) Projection onto the line connecting class centroids *i* and *j* revealed the pairwise discriminability/dissimilarity $\delta_{i,j}$ of object classes *i* and *j* (top), and the distances to class centroids yielded the within-class and between-class variance of representations, SS_W and SS_B, and the associated variance ratio $F = SS_B/SS_W$ (bottom). Additionally, a matrix of (mis-)classification probabilities $P(reported\ i \mid true\ j)$ (a.k.a confusion matrix) could be obtained (not shown). (D) To assess object representation generally, test presentations were drawn randomly from the complete set of object presentations (left). To assess changes over the duration of the experiment, the set of presentations was divided into five successive "batches" and test presentations were drawn from one of these batches (bottom). In either case, the training set comprised all remaining presentations (i.e., the complement of the test set).

patterns from the original activity space (dimensionality N_{dim}) to the maximally discriminative subspace $\mathbb{S}$ and back. As this method is linear and relies on all available degrees of freedom, its results are deterministic. An important feature of this particular algorithm is that within-class variance is maintained near unity for all classes, by means of a suitable scaling of the subspace dimensions. The link github.com/cognitive-biology/DLDA provides a Matlab implementation of DLDA.

2.5.2. *Amplitudes, distances, and correlations*

Activity patterns x_{jk} associated with trials *k* were analyzed in the maximally discriminative subspace $\mathbb{S}$. The normalized amplitude $a_k = \sqrt{\frac{1}{\kappa-1}\sum_{j=1}^{\kappa-1} x_{jk}^2}$ of such patterns exhibited an average value of $\langle a \rangle = 0.99$. The normalized distance $d_{kl} = \sqrt{\frac{1}{\kappa-1}\sum_{j=1}^{\kappa-1}\left(x_{jk} - x_{jl}\right)^2}$ between patterns associated with trials *k* and *l* measured on average $\langle d \rangle = 1.40$, consistent with distance expected between random patterns of this amplitude ($\sqrt{2}$). Averaging over trials *k* produced normalized response amplitudes $A = \langle a_k \rangle_k$. Averaging over pairs of trials *k*, *l* separated by a given latency $l - k$, produced normalized response distances $D(l-k) = \langle d_{kl} \rangle_{k,l}$.

The patterns from successive trials exhibited a weak temporal correlation, with approximately 5% smaller

distances at delays below 4 trials and approximately 2% larger distances at delays ranging from 6 to 15 trials (see Supplementary Fig. S1A, B). Comparing pairs of trials with different types of objects, we observed approximately 3% *larger* response distances D (at all latencies) for the same recurring objects than for either different recurring or non-recurring objects (Supplementary Fig. S1C). Differential response amplitudes A increased marginally with latency, because response amplitudes tended to increase slightly over the course of each run (Supplementary Fig. S1D). This trend was evident for all types of objects and with both "structured" and "unstructured" sequences. In other words, the effect of object type on multivariate hemodynamic responses was limited to response distances and did not extend to response amplitudes. Thus, our data provided no evidence for "repetition suppression."

For certain analyses (Sections 2.5.8 and 2.5.9), we established for each parcel w the average delay-dependent distance $T_w(\Delta k) = \langle d_{w,u,r}(\Delta k)\rangle_{u,r}$ between patterns with a relative delay of Δk trials, where the average was taken over subjects u and runs r. The time-course T_w allowed us to discount temporal correlations by computing $d^{corrected}_{w,u,r}(\Delta k) = d_{w,u,r}(\Delta k) - T_w(\Delta k) + \langle T_w(\Delta k)\rangle_{\Delta k}$, where $\langle T_w(\Delta k)\rangle_{\Delta k}$ is the average value over delays Δk.

2.5.3. Representation of shape "identity" for recurring objects

Our observations comprised approximately 200 activity patterns for each of the 15 recurring object classes (per observer and condition). To allow for cross-validation, we randomly divided these patterns in a larger "training set" (90% or 190 ± 7.7 per object class) and a smaller test set (10% or 22 ± 0.9 per object class) (Fig. 2B). Note that the "training set" comprised exclusively activity patterns associated with recurring objects. To reduce the variability introduced by random test sets, this selection was repeated $N_r = 20$ times and all statistical measures described below represent the average over repetitions. As illustrated in Figure 2C, in the discriminative subspace $\mathbb{S}$, we compared the n_i test set exemplars x_{ki} (where $k = 1,\ldots,n_i$) of class i to the centroids c^{train}_j established for the training exemplars of class j. To compute Mahalanobis distances and variance ratios (see below), we compared test set exemplars x_{ki} of class i to the centroids c^{test}_j of test set exemplars of class j.

We used three measures for this comparison, all with comparable results. Firstly, the nearest class centroid c^{train}_i to each pattern exemplar x_{ki} was identified to establish a matrix of classification probabilities $P(j|i)$ (probability that an exemplar of class i is nearest to the centroid of class j), also known as "confusion matrix," as well as the "classification accuracy" $\alpha = \sum_i P(i|i)P(i)$, which is the probability that the nearest centroid is the correct one.

Secondly, for each pair of object classes (i, j), object exemplars x_{ki} and x_{kj} from the test set were projected onto the line connecting the two test set centroids, c^{test}_i and c^{test}_j, and a pairwise discriminability/dissimilarity/Mahalanobis distance $\delta_{i,j}$ was computed from the means, μ_i and μ_j, and variances, σ^2_i and σ^2_j, of these projections, as $\delta_{i,j} = \frac{|\mu_i - \mu_j|}{\sqrt{\frac{1}{2}(\sigma^2_i + \sigma^2_j)}}$. The average over all pairs of object classes was computed as $\delta = \frac{2}{\kappa(\kappa-1)}\sum_{i,j}\delta_{i,j}$.

Thirdly, given class centroids c^{test}_i and overall centroid c^{test}, we computed the Euclidean distances $d_{ki} = \left\|x_{ki} - c^{test}_i\right\|$ between exemplars x_{ki} and class centroid c^{test}_i and, for each object class i, the "sum of squares" as $SS_{Wi} = \sum_{k=1}^{n_i} d^2_{ki}$. The "within-class" variance of all classes was computed as $SS_W = \frac{1}{N}\sum_{i=1}^{\kappa} SS_{Wi}$, where $N = \sum_{i=1}^{\kappa} n_i$. Similarly, from the Euclidean distances $d_i = \left\|c^{test}_i - c^{test}\right\|$ between individual and overall centroids, we computed "between-class" variance $SS_B = \frac{1}{N}\sum_{i=1}^{\kappa} n_i d^2_i$. From the Euclidean distances $d_{ki} = \left\|x^{test}_{ki} - c^{test}\right\|$ between exemplars and overall centroid, we computed "total" variance $SS_T = \frac{1}{N}\sum_{i=1}^{\kappa}\sum_{k=1}^{n_i} d^2_{ki}$. Variances SS_W, SS_B, and SS_T are also denoted, respectively, SS_{same}, SS_{diff}, and SS_{fam} further below. To quantify the discriminability of classes, the variance ratio $F_{identity} = SS_B(N-\kappa)/SS_W(\kappa-1)$ provided a non-parametric multivariate statistic (PERMANOVA; Anderson, 2001). The average within-class and between-class dispersion per dimension could be estimated as $\sigma_W = \sqrt{SS_W/(N-\kappa)}$ and $\sigma_B = \sqrt{SS_B/(\kappa-1)}$, respectively.

2.5.4. Minimum statistic

To test for statistical significance, we computed average classification performance (in terms of both classification accuracy α_{obs} and f-ratio F_{obs}) over N_r test sets, as well as over 10^3 first-level permutations of object identities (in each of the N_r test sets). In principle, we could have tested an "individual null" hypothesis for every parcel and every data set, namely, the probability of obtaining the observed performance α_{obs} (or F_{obs}) purely by chance. Instead, we computed the "minimum statistic" $m = min_k \alpha_k$ (or $m = min_k F_k$) over data sets k, as well as over 10^5 second-level permutations (drawn randomly from the first level permutations) and tested

the "global null" hypothesis, namely, the probability $p_n(m)$ of obtaining the observed minimum performance over n data sets purely by chance. This computation was performed separately for each of the 2 conditions (8 data sets from 8 observers per condition) as well as for the union of conditions (16 data sets from 8 observers). When the "global null" hypothesis could be rejected, we inferred statistically significant classification performance in at least *some* data sets. Our threshold for significance was $p_n^\star(m) < 0.05$ after correction for multiple comparisons (758 parcels and 2 conditions) (Allefeld et al., 2016).

2.5.5. Prevalence analysis

To summarize the results from all observers and conditions, we used a "prevalence analysis" (Allefeld et al., 2016). Prevalence γ_{true} is the fraction of significant performance over $n = 16$ data sets. To test the "prevalence null" hypothesis that γ_{true} is below a threshold $\gamma_0 = 0.5$, an upper bound for $P(\gamma_{true} < \gamma_0)$ was obtained from the probability $p_n^\star(m)$ of the minimum statistic over $n = 16$ data sets, after correction for multiple comparisons:

$$P(\gamma_{true} < \gamma_0) \le p_n(m|\gamma) = \left[(1-\gamma_0)\sqrt[n]{p_n^\star(m)} + \gamma_0 \right]^n$$

This was the criterion used to label parcels as "identity selective." Threshold prevalence $\gamma \simeq 0.5$ corresponded to corrected probability $p_n^\star(m) \simeq 0.0012$ and *minimal* accuracy of 6.67% (i.e., near chance).

Additionally, we computed γ_{est} as the largest value for which the "prevalence null" hypothesis could be rejected from

$$\gamma_{est} = \frac{\sqrt[n]{\alpha} - \sqrt[n]{p_n^\star(m)}}{1 - \sqrt[n]{p_n^\star(m)}}$$

where $p_n^\star(m)$ is the corrected minimum probability, $n = 16$ the number of data sets, and $\alpha = 0.05$ the significance threshold.

2.5.6. Representation of shape "novelty" for non-recurring objects

Although recurring and non-recurring objects were comparable and generated in the same way, it seemed possible that neural representations might discriminate the class of 15 recurring objects from the class of 360 non-recurring objects. Indeed, the two classes became discriminable after observers had learned to classify recurring objects as "familiar" and non-recurring objects as "novel." Accordingly, we considered this discriminability a representation of "novelty."

To assess the neural representation of "novelty," we divided non-recurring and recurring objects into two sets of unequal size (approximately $N = 216 \times 15$ recurrent or "familiar" exemplars vs. $M = 360$ non-recurrent or "novel" exemplars). From the Euclidean distances $d_k = ||x_k - c||$ between test set exemplars x_k and centroids $c_{fam} = \frac{1}{N}\sum_{k=1}^{N} x_k$ or $c_{nov} = \frac{1}{M}\sum_{k=1}^{M} x_k$, we obtained "within-class" variance $SS_W = SS_{fam} + SS_{nov}$, where $SS_{fam} = \frac{1}{N+M}\sum_{k=1}^{N} d_{k,fam}^2$ and $SS_{nov} = \frac{1}{N+M}\sum_{k=1}^{M} d_{k,nov}^2$. From distances $d_{fam} = ||c_{fam} - c_{tot}||$ and $d_{nov} = ||c_{nov} - c_{tot}||$ between class centroids and overall centroid $c_{tot} = \frac{N}{N+M}c_{fam} + \frac{M}{N+M}c_{nov}$, we obtained "between-class" variance $SS_B = SS_{novfam} = \frac{N}{N+M}d_{fam}^2 + \frac{M}{N+M}d_{nov}^2 = \frac{NM}{(N+M)^2}(c_{fam} - c_{nov})^2$. Finally, from distances $d_k = ||x_k - c_{tot}||$ between exemplars and overall centroid, we obtained total variance $SS_T = \frac{1}{N+M}\sum_{k=1}^{N+M} d_k^2$. To quantify the discriminability of non-recurring and recurring objects, we formed the variance ratio $F_{novelty} = SS_B(N+M-2)/SS_W$ (Anderson, 2001). Average within-class and between-class dispersion per dimension was obtained from $\sigma_W = \sqrt{SS_W/(N+M-2)}$ and $\sigma_B = \sqrt{SS_B}$, respectively.

2.5.7. Changes of representation analyzed in "batches"

To assess changes in neural representations over the course of the experiment, while also allowing for cross-validation, we divided all recurring object presentations into five successive "batches" $B_1, B_2, \ldots$, each with 20% of the presentations (Fig. 2D). In this way, we could select "test sets" for cross-validated DLDA from one particular batch, while retaining all other presentations as a "training set." As every recurrent object was presented 210 ± 9 times over all sessions, a batch would comprise 42 ± 1.8 presentations, a test set 21 ± 0.9, and a training set 189 ± 8.1 presentations. To reduce the variance deriving from test set selection, we repeated the random selection $N_r = 20$ times and averaged over repetitions.

To quantify representational changes over the course of learning, we computed the variance ratios $F_{m,w,u}^{identity}$ for each temporal window or batch m, identity-selective parcel w, and data sets $u \in \{1, \ldots, 16\}$. We formed the average

ratio over 16 data sets, $F^{identity}_{m,w} = \left\langle F^{identity}_{m,w,u} \right\rangle_u$, and assessed statistical significance by shuffling (10^3 permutations) the identity of recurring objects to obtain the distribution of variance ratios due to chance or data structure. The mean $\mu_{m,w}$ and variance $\sigma^2_{m,w}$ of this distribution could also be used to convert $F^{identity}_{m,w}$ into z-score values $Z^{identity}_{m,w} = \left(F^{identity}_{m,w} - \mu_{m,w}\right) \big/ \sigma_{m,w}$.

Additionally, we performed a regression analysis and quantified representational changes in terms of linear trends. Specifically, we determined a "rate" parameter $\beta^{identity}_w$ by fitting a linear mixed-model $F^{identity}_{m,w,u} = \beta_{0,w} + \beta^{identity}_w m + \xi_{0,w,u} + \xi_{1,w,u} m + \epsilon_{m,w,u}$ with data sets u as the grouping variable, where $\beta_{0,w}$ was a fixed-effect coefficient, $\xi_{0,w,u}$ and $\xi_{1,w,u}$ were random effect coefficients, and $\epsilon_{m,w,u}$ was residual error.

Similarly, to assess whether neural representations of non-recurring objects change with learning, we divided all object presentations (recurring and non-recurring) into five successive "batches" $B_1, B_2, \ldots$, each with 20% of the presentations (Fig. 2D), to obtain variance ratios $F^{novelty}_{m,w,u}$ for each temporal window or batch m, identity-selective parcel w, and data sets $u \in \{1, \ldots, 16\}$. After averaging over 16 data sets, $F^{novelty}_{m,w} = \left\langle F^{novelty}_{m,w,u} \right\rangle_u$, we assessed statistical significance by shuffling (10^3 permutations) the identity of recurring and non-recurring objects to obtain the distribution of variance ratios due to chance or data structure. The mean $\mu_{m,w}$ and variance $\sigma^2_{m,w}$ of this distribution were used to convert $F^{novelty}_{m,w}$ into z-score values $Z^{novelty}_{m,w} = \left(F^{novelty}_{m,w} - \mu_{m,w}\right) \big/ \sigma_{m,w}$.

Additionally, we performed a regression analysis to establish linear trends. Changes in the representation of object "novelty" were assessed by fitting the "rate" parameter $\beta^{novelty}_w$ in a linear mixed-model $F^{novelty}_{m,w,u} = \beta_{0,w} + \beta^{novelty}_w m + \xi_{0,w,u} + \xi_{1,w,u} m + \epsilon_{m,w,u}$, with data sets u as the grouping variable, where $\beta_{0,w}$ was a fixed-effect coefficient, $\xi_{0,w,u}$ and $\xi_{1,w,u}$ were random effect coefficients, and $\epsilon_{m,w,u}$ was a residual error.

To establish linear trends $F_m = \langle F_{m,w,u} \rangle_{w,u}$ (of either identity and novelty) that average over both parcels w and data sets u, we obtained a rate parameter β_1 by fitting linear mixed-model $F_{m,w,u} = \beta_0 + \beta_1 m + \xi_{0,w,u} + \xi_{1,w,u} m + \epsilon_{m,w,u}$ with both parcels and data sets as grouping variables.

2.5.8. Geometry of representations

In the cross-validated analyses described above, subspaces $\mathbb{S}$ differed slightly between different batches (and training sets). To analyze the geometry of neural representations in a stable framework, we repeated some analyses in fixed subspaces $\mathbb{S}$ that reflected all observations (i.e., all recurring activity patterns x_k). In the fixed subspace, we calculated the normalized amplitude $a_k = \left\|x_k\right\| / \sqrt{\kappa - 1} = \sqrt{\sum_{j=1}^{\kappa-1} x_{jk}^2} \,/ \sqrt{\kappa - 1}$ of individual patterns k and the normalized pairwise distance $d_{kl} = \left\|x_k - x_l\right\| / \sqrt{\kappa - 1} = \sqrt{\sum_{j=1}^{\kappa-1} (x_{jk} - x_{jl})^2} \,/ \sqrt{\kappa - 1}$ between two patterns k and l.

For each parcel w, data set u, and run r, we obtained the average amplitude $A^{tot}_{w,u,r} = \frac{1}{N+M} \sum_{k=1}^{N+M} a_k$ of all patterns, the average amplitude $A^{fam}_{w,u,r} = \frac{1}{N} \sum_{k=1}^{N} a_k$ of *recurring* patterns, and the average amplitude $A^{nov}_{w,u,r} = \frac{1}{M} \sum_{k=1}^{M} a_k$ of *non-recurring* patterns. Similarly, we obtained the average pairwise distance $D^{tot}_{w,u,r} = \frac{2}{(N+M)(N+M-1)} \sum_{k=1}^{N+M} \sum_{l=k}^{N+M} d_{kl}$ between all patterns, the average distance $D^{nov}_{w,u,r} = \frac{2}{M(M-1)} \sum_{k=1}^{M} \sum_{l=k}^{M} d_{kl}$ between non-recurring patterns, the average distance $D^{fam}_{w,u,r} = \frac{2}{N(N-1)} \sum_{k=1}^{N} \sum_{l=k}^{N} d_{kl}$ between recurring patterns, and the average distance $D^{novfam}_{w,u,r} = \frac{1}{MN} \sum_{k=1}^{M} \sum_{l=1}^{N} d_{kl}$ between pairs comprising one recurring and one non-recurring pattern. For recurring patterns, we further obtained the average distance $D^{same}_{w,u,r} = \frac{2}{N(N/\kappa - 1)} \sum_{i=1}^{\kappa} \sum_{k=1}^{n_i} \sum_{l=k}^{n_i} d_{kl}$ between pairs of recurring patterns in the *same* class and the average distance $D^{diff}_{w,u,r} = \frac{1}{N(N - N/\kappa)} \sum_{i=1}^{\kappa} \sum_{k=1}^{n_i} \sum_{l=1}^{N-n_i} d_{kl}$ between pairs in *different* classes. All distances were corrected for the temporal auto-correlation by subtracting the time course of $T_w(i, j)$, as described above.

As described further above, the distances between individual activity patterns and different centroids—such as c_{tot}, c_{nov}, and c_{fam}—yielded total variance $SS_T = SS_{tot}$, within-class variance $SS_W = SS_{fam} + SS_{nov}$, and between-class variance $SS_B = SS_{novfam}$. For recurring patterns, distances to individual class centroids c_i and overall centroid c_{fam} yielded total variance $SS_T = SS_{fam}$, within-class variance $SS_W = SS_{same}$, and between-class variance $SS_B = SS_{diff}$.

These values were computed for each parcel w, observer u, and run r, in order to obtain variance fractions $F^{fam}_{w,u,r} = SS_{fam}/SS_{tot}$, $F^{nov}_{w,u,r} = SS_{nov}/SS_{tot}$, $F^{novfam}_{w,u,r} = SS_{novfam}/SS_{tot}$, $F^{same}_{w,u,r} = SS_{same}/SS_{fam}$, and $F^{diff}_{w,u,r} = SS_{diff}/SS_{fam}$, as well as variance ratios $R^{identity}_{w,u,r} = SS_{diff}(N - \kappa)/SS_{same}(\kappa - 1)$ and $R^{novelty}_{w,u,r} = SS_{novfam}(N + M - 2)/(SS_{nov} = SS_{fam})$.

2.5.9. Changes with learning analyzed by "runs"

Fixed subspaces permitted us to assess representational changes between successive "runs." To this end, we computed average amplitudes $A_{w,u,r}$, distances $D_{w,u,r}$, variances $SS_{w,u,r}$, and variance ratios $F_{w,u,r}$, as described above, for each parcel w, data set $u \in \{1,\ldots, 16\}$, and run r. Within each session s, we assessed the changes of these parameters $Y \in \{A, D, SS, F\}$ over runs $r' \in s$ by determining a "rate" parameter β_s for identity-selective w and non-selective parcels w'. Each β_s coefficient was acquired from a linear mixed-model $Y_{r',w,u} = \beta_{0,s} + \beta_s r' + \xi_{0,w,u} + \xi_{1,w,u} r' + \epsilon_{r',w,u}$ with observers and parcels as grouping variables, where $\beta_{0,s}$ was a fixed-effect coefficient, $\xi_{0,w,u}$ and $\xi_{1,w,u}$ were random effect coefficients, and ϵ was residual error. The same approach was used to assess gradual changes over runs in the centroid-to-centroid distances $D_{same}(r)$, $\Delta D_{same}(r)$, $D_{nov}(r)$, and $\Delta D_{nov}(r)$. This served to test the statistical significance of linear rates β_s in each session. Sessions with significant rates are marked by stars in Figure 6.

2.5.10. Stability of shape identity and novelty representations

We also assessed the stability of the representation of the 16 response classes (15 recurring and 1 non-recurring) over the course of the experiment. To this end, we compared the average representation in individual runs r (centroids C_r of responses to exemplars) to the average representation over all runs (centroids C_{ave}). For observer u, identity-selective parcel w, and object class i, we calculated the Euclidean distance $D_{u,w,i,r}$ between the relevant C_r and C_{ave}, and also the difference $\Delta D_{u,w,i,r}$ between the relevant centroids from successive runs, C_r and C_{r+1}. After averaging over observers u, identity-selective parcels w, and object classes i, we obtained $D_{same}(r)$ and ΔD_{same} for recurring objects and by $D_{nov}(r)$ and $\Delta D_{nov}(r)$ for non-recurring objects.

As a baseline for comparison, we also computed the distances $D_{u,w,i,r}$ and differences $\Delta D_{u,w,i,r}$ that may be expected purely on the basis of response variance. To this end, we permuted the sequence of all 3,600 trials, separately within each of the 16 response classes (15 recurring and 1 non-recurring) such as to obtain 18 "pseudo-runs" with 200 trials each. Expectation values were obtained by repeating this $N_r = 1{,}000$ times.

We note that, in an n-dimensional hypersphere of unit radius, the average Euclidean distance between two random points is

$$d_{ave} = \frac{2^n}{\sqrt{\pi}} \frac{\Gamma^2\left(\frac{n+1}{2}\right)}{\Gamma\left(n+\frac{1}{2}\right)}$$

with $d_{ave} \approx 1.4017$ for $n = 14$.

2.5.11. Dimensional reduction

To visualize representational geometry in two dimensions, we randomly sampled 50 response patterns to each of the recurring and non-recurring objects within the first and the last sessions and calculated a $1{,}600 \times 1{,}600$ pair-wise distance matrix ($D_{w,u}$) for each identity-selective parcel w and subject u. We did not wish to average distance matrices over observers, as we did not expect the activity patterns of different observers to be comparable. To sidestep this difficulty, we permuted the order of recurring objects 100 times and for each subject obtained an average matrix $\bar{D}$ over permutations, which was then averaged over subjects. To visualize the representational geometry of identity in the first and the last session, we used multidimensional scaling (Matlab function *mdscale*, metric stress) to map the distances matrices for recurring objects (50 exemplars from the first session and 50 exemplars from the last session) into a two-dimensional space. To visualize the representational geometry of novelty, we restricted the distance matrix to non-recurring objects (50 exemplars from the first session and 50 exemplars from the last session) and just 3 of the 15 recurring objects (20 exemplars from either session).

3. RESULTS

Observers viewed sequences of computer-generated objects, with each object shown for 2.5 *s* while rotating in three dimensions (Fig. 1A, B, a movie may be viewed HERE). Over three sessions, observers viewed 3,600 objects in total, of which 3,240 were presentations of *recurring* objects (15 different objects, each appearing approximately 216 times) and 360 were presentations of *non-recurring* objects (360 different objects, each appearing once). The display was intended to be sufficiently intriguing to remain interesting over 3 successive days. To this end, presentations never repeated exactly. Observers were required to classify each object as "familiar" (recurring) or "novel" (non-recurring). The task performance improved as observers became increasingly familiar with recurring objects, as illustrated in Figure 1C. Over the first 600 presentations, classification performance improved approximately from 50% correct

(chance) to 85% correct, and reaction times decreased approximately from 1.65s to 1.25s. Over the remaining 3,000 presentations, performance improved further to approximately 90% correct and reaction times decreased further to approximately 0.95s. After three sessions, all observers were "familiar" with all recurring objects and could pick them out from an array of distractor objects.

All sessions were performed in an MRI scanner while whole-brain functional imaging data were being collected. In the following, we report the results of three types of analyses. First, we describe the cortical areas in which multivariate BOLD activity encodes information about the identity of recurring objects ("object identity"), as determined by cross-validated analyses of entire data sets (3 sessions per observer). Second, we describe changes in cortical representations over coarse time intervals, by means of cross-validated analyses of successive parts of the data sets (3 sessions divided into 5 batches). These changes pertain to the encoding of both recurring objects and the distinction between recurring and non-recurring objects ("object novelty"). Third, we describe changes in representations over finer time intervals (3 sessions divided into 18 runs), by foregoing cross-validation and adopting a fixed reference frame. These finer intervals confirm the results from coarse intervals but reveal more details about the geometry of neural representations and their development over time.

3.1. Cross-validated representation of object identity

To assess the extent to which multivariate neural responses to recurring objects encoded object identity, we relied on optimal linear classifiers combined with cross-validation ("direct linear discriminant analysis," DLDA, see Methods for details). Specifically, we quantified the "identity" information in multivariate responses of every parcel $w \in \{1,\ldots, 758\}$ and data set $u \in \{1,\ldots, 16\}$ in terms of classification accuracy $\alpha_{w,u}$, average pairwise dissimilarity $\delta_{w,u}$, and the ratio of between-class and within-class variance $F_{w,u}$. All three measures proved highly correlated and supported similar conclusions. For example, Figure 3B illustrates the correlation of classification accuracy $\alpha_{w,u}$ and variance ratio $F_{w,u}$ ($\rho = 0.94$, $p < 0.001$). The correlations of $a_{w,u}$ and $\delta_{w,u}$ ($\rho = 0.95$, $p < 0.001$), and of $\delta_{w,u}$ and $F_{w,u}$ ($\rho = 0.98$, $p < 0.001$) were comparably strong. The results of individual observers from the two experimental conditions (structured and unstructured object sequences) were highly similar as well, demonstrating test-retest consistency (Supplementary Fig. S2).

For most parcels, the results from different observers showed considerable variability. Whereas a few parcels exhibited significant accuracy $\alpha_{w,u}$ and variance ratio $F_{w,u}$ in all data sets (e.g., Calcarine 331), in many parcels the representation of object identity was significant only in some data sets (e.g., Parahippocampus 325) (Fig. 3B). Global significance was assessed by comparing the *minimal* accuracy or variance ratio over the 8 data sets from one condition (structured or unstructured) to the minimal values obtained with shuffled data (red ellipse in Fig. 3C, see Methods for details).

Minimal classification accuracy α_w was significant in 17% of all parcels (128 of 758 parcels) in the structured sequence condition and in 19% of parcels (146 of 748) in the unstructured condition ($p^\star$ 0.05, corrected for multiple comparisons), when compared to null-distributions obtained from shuffled object identities. For minimal variance ratios $F_{w,u}$, the corresponding values were 18% and 17%, respectively (136 and 130 parcels). To combine the results from both conditions, we used a "prevalence" analysis to determine parcels in which "identity" was represented significantly in a majority of all 16 data sets (prevalence $\gamma \geq 0.5$), once again comparing the observed minimal values to the minimal values obtained with shuffled data (red ellipse in Fig. 3C, see Methods for details).

Figure 3A illustrates the 124 parcels identified as significantly "identity-selective" by the prevalence criterion $\gamma \geq 0.5$ and Supplementary Figure S3 shows the same information in terms of a sliced brain. Among these were 70 parcels in the occipital cortex, 29 in the parietal cortex, 18 in the fusiform or temporal cortex, and 7 in the frontal cortex. The average prevalence of identity-selectivity in these parcels was 0.663 ± 0.016 (mean and S.D.), and the minimal value was 0.58. As the prevalence criterion (based on 16 data sets) was marginally more conservative than the accuracy criterion (based on 8 data sets), 120 of the 124 parcels were significantly "identity-selective" in terms of both criteria. The four exceptions (identified only by prevalence, but not by accuracy) were Frontal-superior-R 56, Occipital-superior-R 393, Occipital-middle-L 403, and Parietal-superior-R 510. Appendix Table A1 lists the statistical significance of all three criteria for all "identity-selective" parcels.

Overall, there was a pronounced posterior-anterior gradient. Whereas many parcels at the posterior pole of the brain exhibited high classification accuracy, this tended to progressively decrease at more anterior locations (Fig. 3A; Supplementary Fig. S3; Appendix Table A1). To formalize this trend, we assigned 66 of the 124 identity-selective parcels to the 25 topographic visual areas defined by Wang et al. (2015) and, additionally, to the anterior inferior temporal cortex (AIT) and to the inferior frontal cortex (IFC). Supplementary Figure S6 provides an overview of all topographically assigned and non-assigned parcels selective for identity. As illustrated in Figure 8A, this assignment showed that accuracy was comparable in

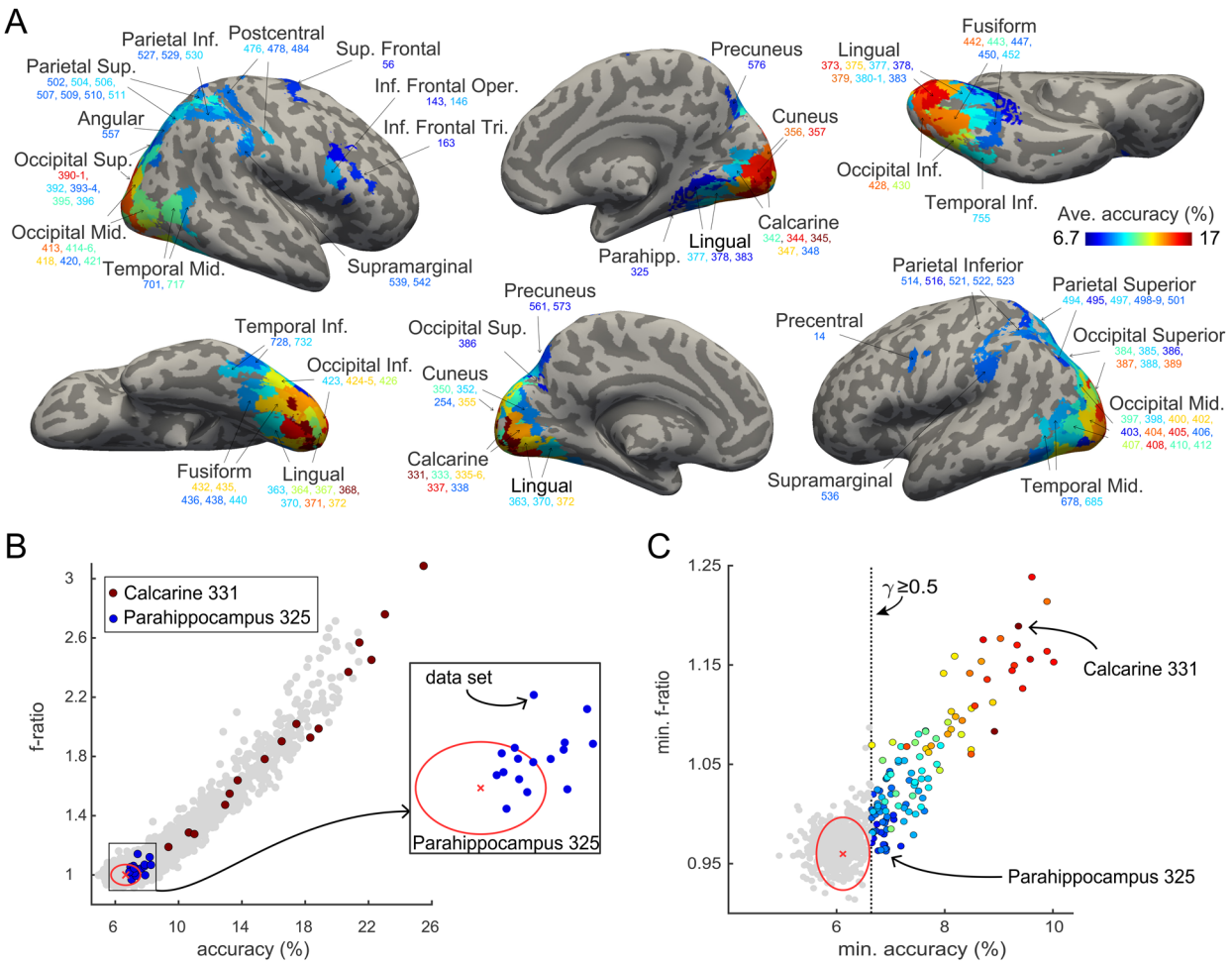

Fig. 3. Neural representation of object identity. (A) Identity-selective parcels are shown in color (124 of 758 parcels) on an inflated standard brain and are found in the occipital (70 parcels), parietal (29), temporal/fusiform (18), and frontal cortex (7). Color indicates classification α_w^{ave} (average over 16 data sets), and ranges from chance to the largest observed value (6.67% to 17%). Parcels are identified by AAL region and number (in color), as detailed in Appendix Table A1. (B) Classification $\alpha_{w,u}$ and variance ratio $F_{w,u}$ for all 758 parcels w and 16 data sets u. Both values differ highly significantly from the values obtained with shuffled object identities (red cross and ellipse, representing mean ± 3 S.D.). Two particular parcels are highlighted (Calcarine-L 331 in red, Parahippocampus-R 325 in blue, and magnified in the inset) to illustrate the variability of data sets. (C) Minimum values a_w^{min} and F_w^{min} for all parcels over 16 data sets. Identity-selective parcels are colored according to a_w^{ave} as in (A). A minimum above chance 6.67% corresponds to a prevalence γ above 0.5 (dotted vertical line). The distributions obtained with shuffled identities are indicated as well (red cross and ellipse).

early visual areas (V1-hV4) and in the posterior-ventrolateral regions of the temporal lobe, whereas accuracy was lower in the anterior temporal cortex, the inferior frontal cortex, and in parietal cortical areas.

3.2. Cross-validated changes with learning

To assess changes with learning, we separately analyzed five successive and non-overlapping sets of trials ("batches") with linear classifiers and cross-validation (see Methods for details). Specifically, we established ratios of between- and within-class variance for both object identity (15 classes formed by responses to 15 recurring objects) and for object novelty (2 classes formed by responses to recurring and non-recurring objects, respectively). These two variance ratios measured the neural representation of "identity" and "novelty."

Variance ratios were converted to z-score values (with respect to the mean and variance of the corresponding shuffle distribution) before being averaged over data sets and/or over parcels. Figure 4A summarizes the results in terms of a grand average over all identity selective parcels. The average identity and novelty ratios were highly significant in all batches ($p < 0.001$). Over successive batches, the average identity ratio weakened slightly but significantly ($p < 0.05$), whereas the average novelty ratio strengthened considerably, especially between batches $m = 1$ and $m = 2$ ($p < 0.001$).

As expected, it was the between class-variances $SS_B^{identity}$ and $SS_B^{novelty}$ that changed significantly over successive batches m ($p < 0.05$ and $p < 0.001$, respectively), whereas the within-class variances $SS_W^{identity}$ and $SS_W^{novelty}$ remained essentially the same ($p = n.s.$), as illustrated by Figure 4B. This was owing to the DLDA algorithm, which maintained within-class variance near unity. Nevertheless, over successive batches, the neural representations of recurring objects tended to become slightly more similar to each other, but more dissimilar to the representations of non-recurring objects.

To ascertain that these overall trends hold true also for individual parcels, we carried out more conventional regression analyses of variance ratios $F_{m,w,u}^{identity}$ and $F_{m,w,u}^{novelty}$ over batches m, parcels w and data sets u. Specifically, we fitted linear mixed-models in order to estimate "rate" parameters $\beta_w^{identity}$ and $\beta_w^{novelty}$ for each identity-selective parcel w. The results revealed negative rates $\beta_w^{identity}$ and positive rates $\beta_w^{identity}$ for almost all parcels, confirming the overall trends in Figure 4C. The variability over parcels was numerically larger for $\beta_w^{novelty}$ (0.15 ± 0.1, mean and S.D.) than for $\beta_w^{identity}$ (0.022 ± 0.015), with both rates weakly correlated ($\rho = 0.30$, $p < 0.001$). Classification accuracy $\alpha_w^{identity}$ correlated negatively with $\beta_w^{novelty}$ ($\rho = -0.22$, $p < 0.05$) and with $\beta_w^{identity}$ ($\rho = -0.74$, $p < 0.001$).

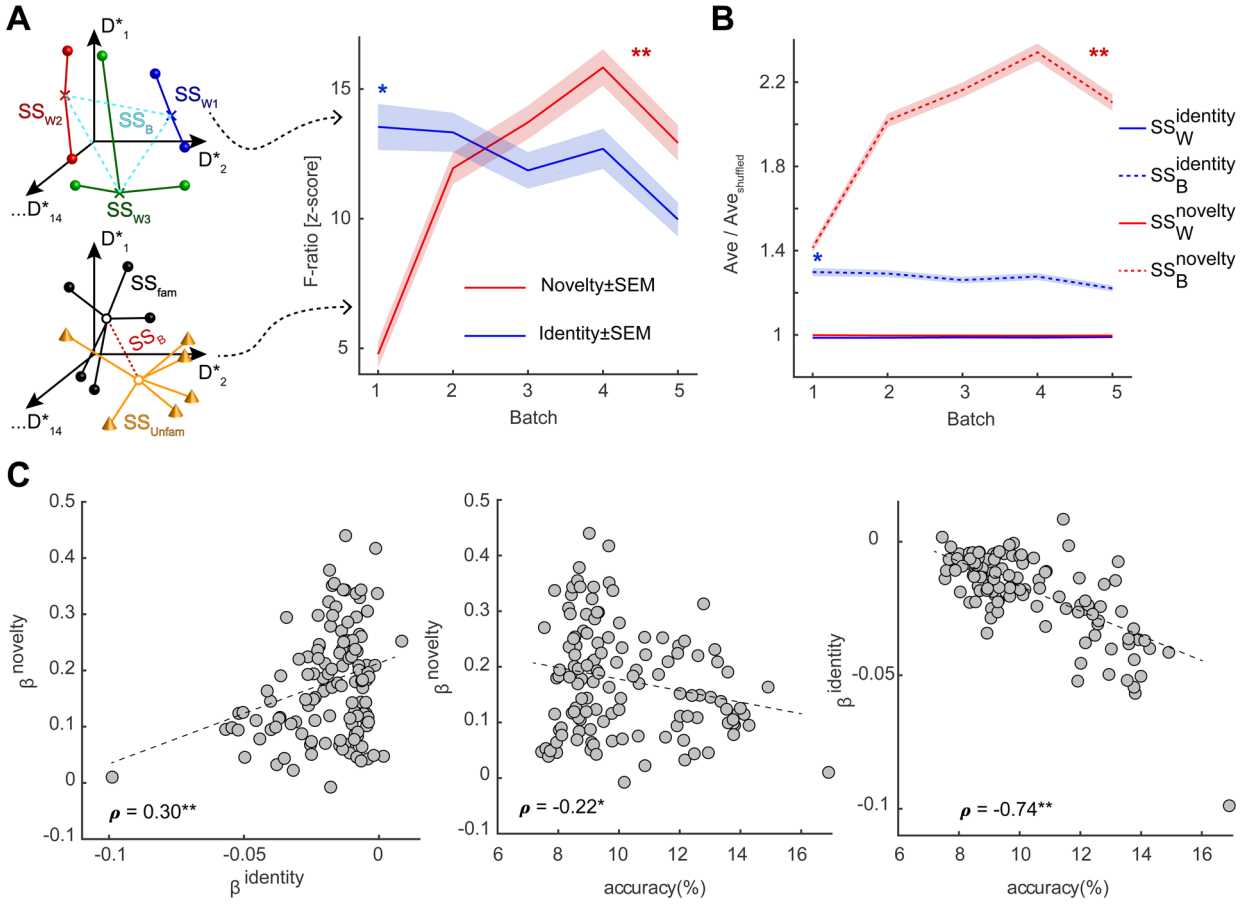

Fig. 4. Changes in the representation of "identity" and "novelty" over successive "batches" of trials. (A) Ratio of within- and between-class variance for object "identity" ($\kappa = 15$ classes, inset top left) and object "novelty" (2 classes, inset bottom left). Average variance ratios $F_m^{identity}$ (blue, mean ± S.E.M.) and $F_m^{novelty}$ (red, mean ± S.E.M.), as a function of batch number m. While $F_m^{identity}$ decreases slightly over time ($p < 0.05$), $F_m^{novelty}$ increases considerably ($p < 0.001$), especially initially. All values are averages over data sets in z-score units. (B) Average within- and between-class variances (mean ± S.E.M.), as a function of batch number m. Whereas between-class variances decrease ($SS_{B,m}^{identity}$, $p < 0.05$) or increase ($SS_{B,m}^{novelty}$, $p < 0.001$), within-class variances remain unchanged. All values are averages over data sets, relative to shuffled averages. (C) Results of regression analysis for 124 identity-selective parcels w. Linear "rate" parameters $\beta_w^{identity}$ and $\beta_w^{novelty}$ compared to each other and to classification α_w. Novelty and identity rates correlate weakly over parcels (left, $\rho = 0.298$, $p < 0.001$), as do novelty rate and classification accuracy $\alpha_w^{identity}$ (middle, $\rho = -0.22$, $p < 0.05$). Identity rates β_w and accuracies α_w correlate strongly and negatively (right, $\rho = -0.74$, $p < 0.001$). Significance of linear trends is indicated by * for $p < 0.05$ and ** for $p < 0.001$.

To take a closer look at the interaction between "novelty" and "identity," we divided the identity-selective parcels into "novelty terciles" (high, medium, and low, defined by $\beta^{novelty}$) before comparing representations of novelty ($F^{novelty}$) and identity (accuracy α) (Fig. 5B). The results differed substantially between batches and terciles. In early batches, $F^{novelty}$ and α correlated for all terciles, suggesting that initially the representations of non-recurring and recurring objects were linked. However, in successively later batches, this correlation waned in the upper tercile. This may suggest that pronounced representations of non-recurrent objects progressively detached from representations of recurrent objects.

Figure 5A illustrates the degree to which identity-selective parcels express the overall novelty trend, as quantified by fitted rate $\beta_w^{novelty}$, and Supplementary

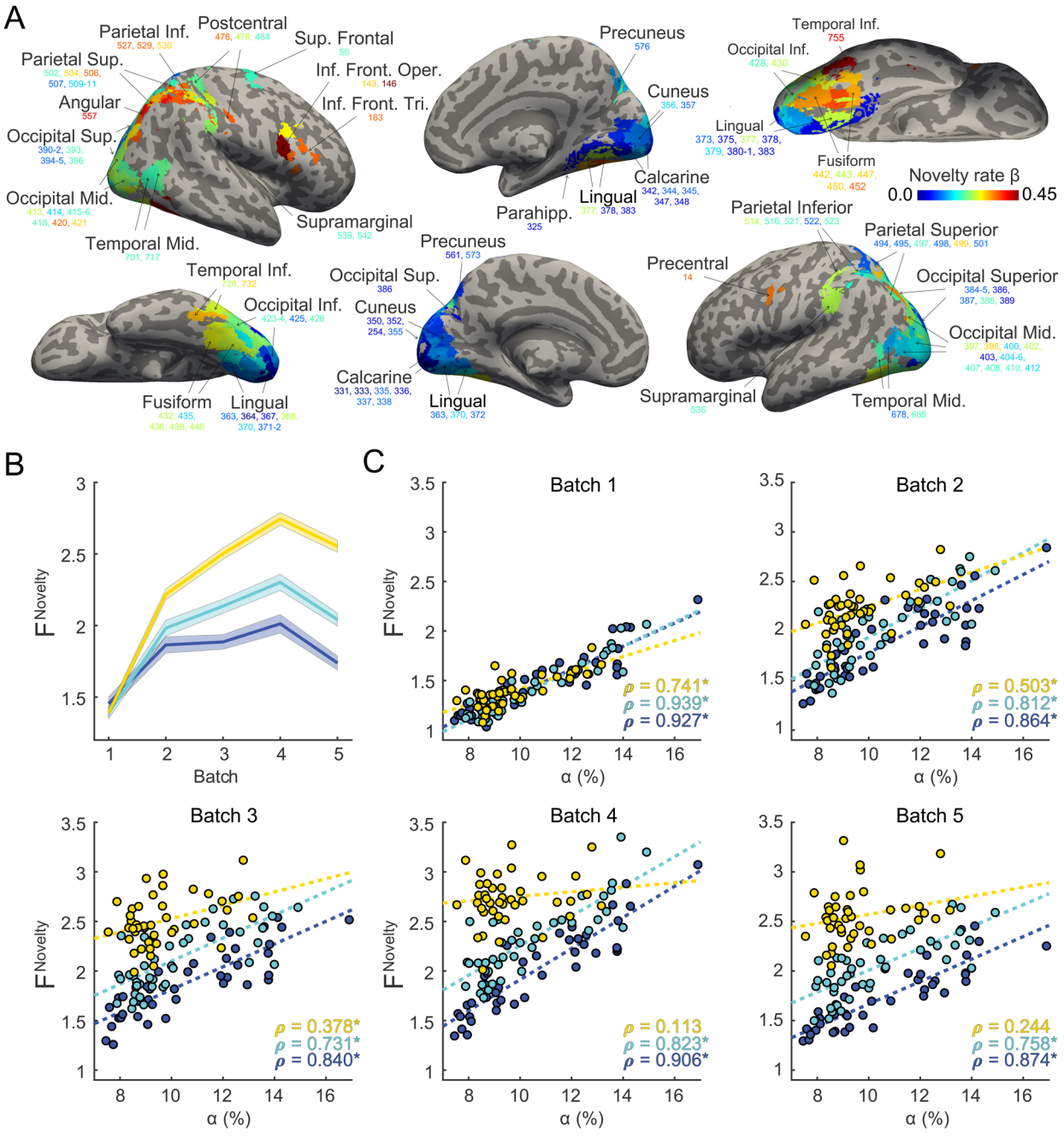

Fig. 5. Neural representation of "novelty" in terms of the variance ratio $F^{novelty}$ and its development over successive batches. (A) Identity-selective parcels and their individual rate parameters $\beta^{novelty}$ (color scale), estimated by fitting linear-mixed models to the $F^{novelty}$ values (from all batches and data sets). (B) Development of $F^{novelty}$ (mean ± S.E.M.) for different "novelty terciles" (upper, middle, and lower tercile of parcels defined by $\beta^{novelty}$). (C) Correlation between $F_w^{novelty}$ and accuracy α_w for different batches and novelty terciles. The parcels of each tercile are distinguished by color, with individual regression lines (dashed) and correlation coefficients ρ (* indicates $p < 0.05$).

Figure S4 shows the same information in terms of brain slices. An anterior-posterior gradient is evident, with a more pronounced representation of novelty at anterior than at posterior locations. This gradient is also apparent when parcels are assigned to topographic visual areas, as illustrated in Figure 8B. Appendix Table A1 lists the rates $\beta_w^{novelty}$ for all identity-selective parcels.

3.3. Geometry of identity and novelty representations

Next, we present results from alternative analyses relying on fixed subspaces $\mathbb{S}$ for each data set (3,600 trials). Fixed subspaces reveal a more detailed geometry of neural representations and allow any changes in this geometry to be tracked over successive runs (200 trials each). The disadvantage of this approach is that it precludes cross-validation. Our aim was to establish not just between- and within-class variances, but also the distances underlying the variances, and the response amplitudes underlying the distances. For the representation of object "identity," the within- and between-class geometry was defined by response pairs to *same* and to *different* recurring objects, respectively. For the representation of object "novelty," the within-class geometry reflected responses either to pairs of *familiar* (recurring) or to pairs of *novel* (non-recurring) objects, whereas the between-class geometry concerned responses to mixed pairs of objects (*novel-familiar*).

We analyzed multivariate responses in terms of variances, distances, and amplitudes and averaged the results over all data sets and all 124 identity-selective parcels, to obtain separate mean values (and standard errors) for each of the 18 successive runs. Additionally, we averaged the results over the remaining 634 (non-identity-selective) parcels of the brain. We hoped that this would help distinguish more general effects and trends (e.g., habituation, attention, alertness) from learning-related changes in shape representations. All distances in these analyses were residual distances, to minimize the influence of temporal auto-correlations (Supplementary Fig. S1; see Methods for details).

The analyzed quantities—response amplitudes A, response distances D, and variances SS—are illustrated schematically in Figure 6A, and the results are presented in Figure 6B–D in terms of the mean values and standard errors for every run. In identity-selective parcels, response amplitudes A_{fam} to recurring patterns decreased during the first session (runs 1 to 6, $p < 0.05$), but not in the second and third session (runs 7 to 12, runs 13 to 18, $p > 0.5$). Response amplitudes A_{nov} to non-recurring patterns showed no significant change (p n.s.) in any session (Fig. 6B). In non-selective parcels, response amplitudes decreased in all sessions, consistent with general habituation. In identity-selective parcels, response distances D_{diff} between different recurring objects declined similarly during the first session ($p < 0.05$), but not during subsequent sessions ($p > 0.6$) (Fig. 6C). Also, response distances D_{same} between the same recurring objects did not change significantly during any session (p n.s.). In contrast, response distances D_{nov} between non-recurring objects declined disproportionately during the first session ($p < 0.05$) but increased during the third session ($p < 0.05$). Response distances D_{novfam} between recurring and non-recurring objects, on the other hand, did not change significantly over sessions (p n.s.).

A first conclusion is that response amplitudes and response distances are consistently larger for recurring objects (blue traces in Fig. 6B, C) than for non-recurring objects (red traces). Importantly, in the very first run, response distances are comparable between different recurring objects (D_{diff}) and different non-recurring objects (D_{nov}), demonstrating that both recurring and non-recurring objects were represented comparably well. Over subsequent runs, response distances decrease far more between different non-recurring objects (D_{nov}) than different recurring objects (D_{diff}), demonstrating that a comparative advantage for *recurring* objects develops gradually (i.e., a kind of repetition enhancement). A second conclusion is that the observed development differs between identity-selective and non-selective parcels. Whereas amplitudes and distances stabilize in the former group of parcels, they habituate progressively in the latter group (both within and between sessions). Thus, the responsiveness of identity-selective parcels remains stable over sessions. A third conclusion is that response distances D_{nov} between different non-recurring objects become comparatively small (already during the first session), not only smaller than the distances D_{diff} between *different* recurring objects but even smaller than the distances D_{same} between the *same* recurring objects.

The results for response variances confirmed the trends observed earlier in the batch analysis of cross-validated variance ratios (Fig. 4A, B). Between-class variance SS_{diff} for recurring objects declined over the course of sessions ($p < 0.005$), whereas between-class variance SS_{novfam} for non-recurring objects increased over the first session ($p < 0.005$), only to decline again during the third session ($p < 0.05$). Within-class variances SS_{same} and SS_{nov} remained largely unchanged. The close correspondence between the trends observed over runs and over batches is illustrated also in Supplementary Figure S5. Surprisingly, non-identity-selective parcels mirrored the trends observed for identity-selective parcels in

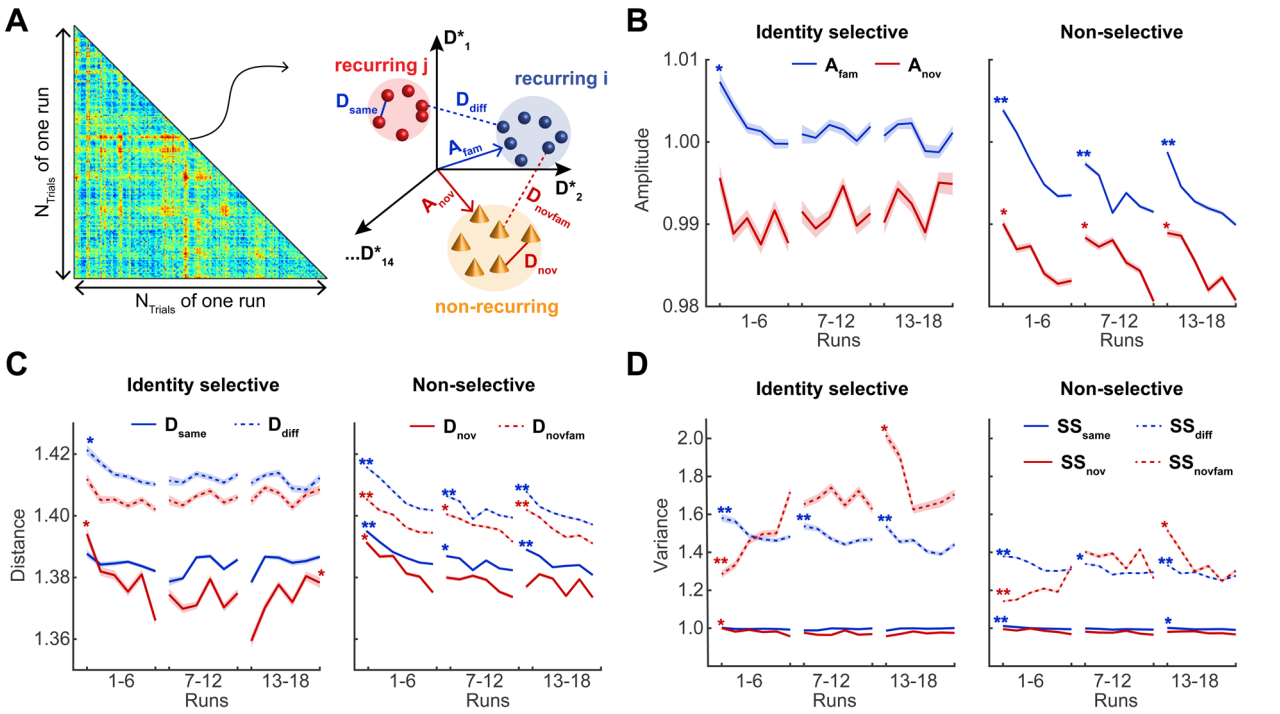

Fig. 6. Geometry of identity and novelty representation over successive sessions and runs. (A) For each run with N_{trials} = 200 trials, we collected all individual response amplitudes *a* and all pairwise response distances *d* (triangular area with color scale) in the maximally discriminating space and computed average amplitudes A_{fam} and A_{nov} (for recurring and non-recurring objects, respectively) and average distances D_{same} and D_{diff} (for same and different recurring objects, respectively), as well as average distances D_{nov} and D_{novfam} (for non-recurring objects and between recurring and non-recurring objects, respectively). (B) Response amplitude A_{nov} (red, mean, and S.E.M.) and A_{fam} (blue, mean, and S.E.M.), over 18 runs grouped into three sessions, for identity-selective (left) and non-selective parcels (right). (C) Pairwise response distance D_{same} (solid blue), D_{diff} (dashed blue), D_{nov} (solid red), and D_{novfam} (dashed red)), over runs and sessions, for both groups of parcels. (D) Variance of response distances SS_{same} (solid blue), SS_{diff} (dashed blue), SS_{nov} (solid red), and SS_{novfam} (dashed red), over runs and sessions, for identity-selective and non-selective parcels. Stars indicate a significant linear trend during a session (see text). All plots show mean (traces) and S.E.M. (shading).

attenuated form. The fact that between- and within-class variances differ systematically suggests that even non-identity-selective parcels represent object identity to some degree.

It is natural to compare these results to the time-course of behavioral performance (fraction correct and reaction time) in our observers Fig. 1C, D). The changes in the representation of *recurring* objects (between class distances D_{diff} and variances SS_{diff}) show a gradual *decrease* in the quality of representation and thus do *not* correspond to improving performance in terms of fraction correct. However, the changes in the representation of *non-recurring* objects, including the decrease of within-class distances D_{nov} and variances SS_{nov} and the increase of between-class variances SS_{novfam} and variance ratio $R_{novelty}$, do correspond to the rapid improvement in fraction correct over the first few runs. Thus, the neural changes over the course of learning point to diverging representations of "novel" (non-recurring) and "familiar" (recurring) objects.

3.4. Stability of identity and novelty representations

Relying on fixed subspaces $\mathbb{S}$ to analyze each data set also permitted us to assess the *stability* of neural representations over successive runs. With this in mind, we established the centroids of response classes for each run and examined the displacement of centroids between successive runs. As this calculation concerned centroid-to-centroid distances (rather than exemplar-to-exemplar distances), we could not correct for temporal auto-correlations.

The computation of centroids for particular response classes is illustrated schematically in Figure 7A. Given the centroids C_{r-1} and C_r for successive runs $r-1$ and r and the average centroid C_{aver} over all runs, we computed absolute centroid-to-centroid distances $DC_r = ||C_r - C_{ave}||$ as well as relative centroid-to-centroid distances $\Delta DC_r = ||C_r - C_{r-1}||$. The 16 response classes were formed by each recurring object (15 classes, DC_{same} and ΔDC_{same}) and by the non-recurring objects (1 class,

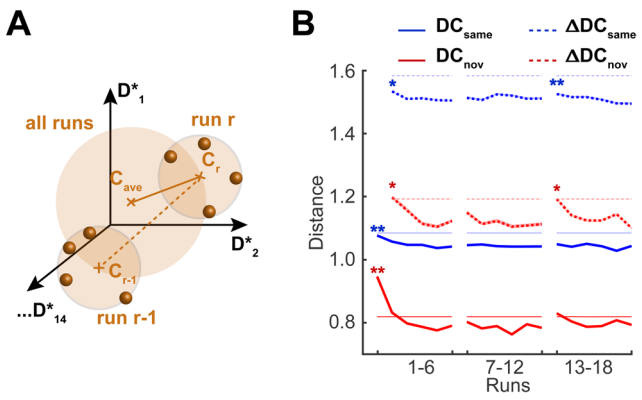

Fig. 7. Stability of identity and novelty representations over successive sessions and runs (A) For each recurring and all non-recurring objects, we calculated the response centroids C_r for each run r and the average centroid C_{ave} for all runs and obtained both absolute distances $DC_r = ||C_r - C_{ave}||$ and relative distances $\Delta DC_r = ||C_r - C_{r-1}||$. (B) Centroid-to-centroid distances (mean ± S.E.M.) for all identity-selective parcels and all data sets. Distances DC_{same} and ΔDC_{same} for the same recurring objects (blue) and distances DC_{nov} and ΔDC_{nov} for non-recurring objects (red) are compared to the corresponding values obtained from shuffled data sets (thin, pale lines). Stars indicate a significant linear trend during a session (see text).

DC_{nov} and ΔDC_{nov}). To compare the displacements expected from sampling noise, we also computed the centroid-to-centroid distances after permuting the responses in each class and regrouping them into 18 "pseudo-runs" (see Methods for details).

The results are shown in Figure 7B. For both recurring and non-recurring objects, average absolute distances $DC_{same}(r)$ and $DC_{nov}(r)$ diminished during the first session (runs 1 to 6, $p < 0.005$), but remained stable during the second and third sessions (runs 7 to 12, and 13 to 18, $p > 0.2$). Notably, absolute distances $DC_{nov}(r)$ of novel objects decreased to a much lower average level. Relative distances $\Delta DC_{same}(r)$ and $\Delta DC_{nov}(r)$ between successive runs declined during the first session (runs 1 to 6, $p < 0.05$), remained stable during the second session (runs 7 to 12, $p > 0.2$), only to decline once again the last session (13 to 18, $p < 0.005$ for recurring and $p < 0.05$ for non-recurring objects). Absolute distances were far larger for recurring than for non-recurring classes, corroborating the substantial "response enhancement" already noted above. Both absolute and relative distances were slightly smaller than predicted by sampling noise (thin, pale lines, $p < 0.001$), demonstrating that responses of true runs were distributed slightly more compactly and consistently than those of pseudo-runs. Note also that relative distances approached the values expected for fully random displacements in a 14-dimensional hypersphere—specifically, relative distances ΔDC were approximately 1.4 times larger than absolute distances DC – again underlining the dominant influence of sampling noise.

4. DISCUSSION

We studied the cortical representation of synthetic visual objects over multiple days of repeated viewing, while observers learned to classify initially unfamiliar objects as "familiar." Relying on "representational similarity analysis" (RSA), we established distances between spatiotemporal hemodynamic (BOLD) responses to exemplars of different *recurring* objects, as well as to exemplars of *non-recurring* objects. Response distances between the same and different recurring objects quantified the neural representation of object *identity*. Response distances between recurring and non-recurring objects measured the neural representation of object *novelty*. The results showed that object identity was neurally represented from the start, in the ventral occipitotemporal cortex and beyond. With growing familiarity, the quality of this neural representation remained high, but its geometry expanded to fill the available representational space. In contrast, the neural representation of non-recurring objects (which remained "novel" by definition) improved over time, but its geometry contracted and shifted to the margins of the representational space.

4.1. Cortical representation of object identity

To permit a fine-grained analysis of representational geometry, we generated complex and three-dimensional shapes that were highly characteristic and distinguishable and presented these shapes from various points of view and in various states of rotation (always for one complete turn) (Kakaei et al., 2021). Thus, observers had to recognize an object from all sides in order to classify it as "familiar." Within the category of our synthetic shapes, every recurring object constituted strictly speaking an "exemplar," with individual presentations providing different "instantiations." However, we chose to term objects "classes" and individual presentations "exemplars," as this terminology conforms better to RSA conventions.

The selectivity of cortical parcels for object identity was assessed in optimized 14-dimensional subspaces $\mathbb{S}$ of the much higher-dimensional space of multivariate responses ($O(10^3)$ dimensions). Specifically, we computed a cross-validated "classification accuracy" (Kriegeskorte, Mur, & Bandettini, 2008) and used a prevalence analysis to combine results from different conditions and observers (Allefeld et al., 2016). Essentially identical results were obtained with alternative measures such as "linear discriminability" and "variance ratio" (of between- and within-class variance; Anderson, 2001).

When spatiotemporal responses to different objects are linearly discriminable, they form a neural representation of object identity. As exemplars of each object were presented from various sides, any such neural representation was by definition view-invariant. The obvious caveats are *(i)* that object rotation may have exposed the same characteristic features in many or most presentations and *(ii)* that multivariate hemodynamic responses over 9 s can only distantly reflect the neuronal activity evoked during each 2.5 s presentation. Nevertheless, hemodynamic signals exhibited significant invariance to the various modes of presentation of a given object (e.g., the initial perspective, the axis, and the sense of rotation).

In contrast to many other studies, we did not observe suppressed responses when objects were repeated (i.e., no "repetition suppression") but rather a small enhancement of responses both with longer delays and later trial numbers (Supplementary Fig. S1). This may simply reflect the fact that the object presentations were highly variable and never repeated exactly. Recall that we designed a highly variable display such as to retain the observers' interest over 3 successive days.

The 124 of 758 parcels that were identified as "identity-selective" on this basis were situated mostly in the ventral occipitotemporal cortex, but some parcels were also located in the parietal or frontal cortex, as illustrated in Figure 3A. The degree of selectivity exhibited a clear gradient, being stronger at the posterior pole and becoming progressively weaker in more anterior and more dorsal regions, as summarized in Figure 8A. These results are consistent with previous findings that multivariate activity distinguishing different exemplars of a particular class of objects (e.g., faces) is present in the ventral and lateral occipital cortex, on the fusiform gyrus, and in the ventral temporal cortex (Brants et al., 2016; Eger et al., 2008; Visconti di Oleggio Castello et al., 2021).

In general, it is thought that progressively "higher" levels of visual processing represent progressively "larger" visual sets, beginning with image features, and widening gradually to object features, object exemplars, object categories, and finally to supercategories such as animate or inanimate objects, or objects and landscapes (Grill-Spector & Weiner, 2014). Accordingly, the discriminability of exemplars within a category is expected to diminish at more anterior locations, which correspond to "higher" levels of visual processing (Eger et al., 2008; Grill-Spector & Weiner, 2014). Moreover, it has been hypothesized that the spatial scale of neural representations increases with the level of abstraction, in the sense that exemplars are represented at smaller scales than categories (Grill-Spector & Weiner, 2014). Thus, if this trend is exacerbated in the more anterior parts of the ventral pathway, exemplar representations may become progressively less discriminable at the spatial resolution of BOLD signals.

A previous study of visual expertise for synthetic shapes (Brants et al., 2016) reported a gradual enhancement of neural representations in object-selective areas, whereas we observed a moderate decline. This difference may have been due to task design. Brants and

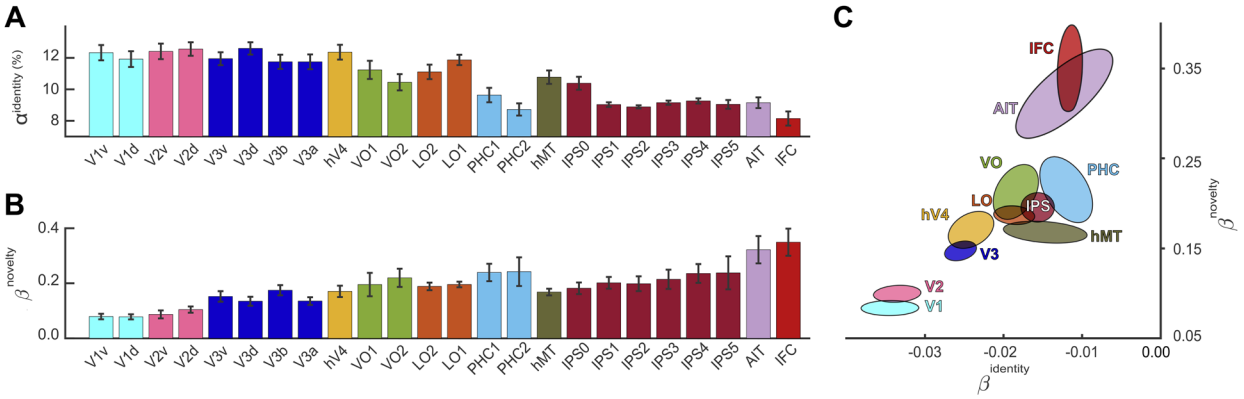

Fig. 8. Shape identity and shape novelty representations in 26 topographical regions. (A) Identity representation as indexed by classification accuracy $\alpha^{identity}$ (mean ± S.E.M.). Posterior regions (V1-hV4, VO, LO) exhibit higher accuracy than more anterior or more dorsal regions (IPS, AIT, IFC). (B) Novelty representation as indexed by rate $\beta^{novelty}$ of novelty gain (mean ± S.E.M.). More anterior or more dorsal regions (IPS, AIT, IFC) exhibit a higher slope parameter than the posterior visual cortex (V1-hV4). (C) Comparison of identity and novelty representations as indexed by rate $\beta^{identity}$ of identity loss (negative values) and rate $\beta^{novelty}$ of novelty gain (positive values). Groups of regions are distinguished by color, with ellipsoids indicating mean and standard error. Note the positive correlation between novelty *gain* and identity *loss*. List of abbreviations: visual cortex (V1, V2, V3, hV4), ventral occipital cortex (VO1, VO2), lateral occipital cortex (LO1, LO2), parahippocampal cortex and fusiform gyrus (PHC), medial temporal areas (hMT), intraparietal sulcus (IPS), anterior inferior temporal cortex (AIT), and inferior frontal cortex (IFC).

colleagues used barely discriminable shapes and emphasized perceptual load, whereas we used highly distinguishable shapes and emphasized memory load.

We also observed identity-selectivity in frontoparietal regions that are typically associated with the dorsal visual pathway and the right frontoparietal "attention network." This is consistent with previous findings on the presence of object- and/or face-selective representations in dorsal areas (Freud et al., 2017; Jeong & Xu, 2016; Konen & Kastner, 2008; Poirier et al., 2006; Visconti di Oleggio Castello et al., 2021). However, the interpretation of this selectivity is not straightforward. Particularly the clusters associated with the "attention network" are often found to express functional correlations with ventral visual areas in both resting and task states (Dornas & Braun, 2018; Mutlu et al., 2022; Smith et al., 2013). Thus, it seems possible that multivariate functional correlations could have propagated identity-selectivity feedforward throughout the "attention network" and beyond.

Finally, we observed pronounced identity-selectivity in the primary visual cortex (calcarine sulcus, left and right), where neuronal activity encodes basic visual features (orientation, spatial frequency, direction of movement, and so on) (Grill-Spector & Weiner, 2014; Haxby et al., 2001). It is possible that multivariate hemodynamic responses in the primary visual cortex could have reflected this visually evoked neuronal activity sufficiently well to have encoded object identity, especially as the rotation may have exposed the same low-level features in many or most presentations. Additionally, hemodynamic responses could have been driven by spatiotemporal patterns of feedback from higher areas of the visual cortex. There is some evidence to suggest that feedback can dominate the hemodynamics of the early visual cortex under continuous viewing conditions (as used here) (Blake & Braun, 2009).

4.2. Cortical representation of novel object shapes

We also investigated the representation of "novel" object shapes that were encountered only once (and never recurred). Note that "novelty" is here not meant to imply "surprise" for the observer in the sense of a violation of expectations (e.g., Uddin, 2015). Rather, it simply denotes the more heterogeneous class of *non-recurring* objects (with 360 exemplars, each from a different object), as distinct from the 15 more homogeneous classes of *recurring* objects (with approximately 200 exemplars each, all from the same object). As mentioned, "novelty" was measured in terms of the linear discriminability of hemodynamic responses to non-recurring and recurring objects in 14-dimensional subspaces $\mathbb{S}$, more specifically, by comparing pairwise response distances between classes (recurring and non-recurring) and within classes (either recurring or non-recurring).

All 124 "identity-selective" parcels were also "novelty-selective," in the sense that hemodynamic responses discriminated non-recurring and recurring objects to some degree, as illustrated in Figure 5A. As discriminative subspaces were optimized for recurring objects—that were generated in the same way as non-recurring objects—some degree of discriminability was to be expected. Moreover, as non-recurring objects were more numerous (360 objects) than recurring objects (15 objects), some discriminability was expected purely by chance, particularly in a 14-dimensional space. However, as discussed further below, the linear discriminability of non-recurring objects increased over successive runs and sessions, mirroring observers' improving ability to classify objects as "novel" or "familiar." Because of this dynamic aspect, we quantified the novelty-selectivity of cortical parcels in terms of an "improvement rate," $\beta^{novelty}$ (Fig. 4). Interestingly, there was an anterior-posterior gradient in that novelty-selectivity was more pronounced in more frontal, parietal, and anterior temporal areas than more posterior temporal and occipital areas, as summarized in Figure 8B. In other words, the representational disparity between familiar object shapes and novel objects shapes tended to be larger in the higher-level (more anterior) visual cortex than in the lower-level (more posterior) cortex, suggesting that learning effects were more pronounced.

4.3. Representational changes with learning

As representational changes with learning were the main objective of our study, we addressed this issue with several complementary approaches. First, we divided our observations from 18 runs into five successive "batches" and established the neural representation of both "identity" and "novelty" separately for each batch with cross-validated statistics, while aggregating over all identity-selective parcels (Fig. 4B). Second, to assess changes in individual parcels, we performed a regressional analysis of the same cross-validated data and obtained "rates" of representational changes for every identity-selective parcel (Fig. 4C). Third, we adopted stable discriminative subspaces $\mathbb{S}$ and sacrificed cross-validation in order to analyze representational geometry over individual runs (Fig. 6). All three approaches yielded comparable results.

Already in the first run and the first batch, without time for plasticity or learning, the neural representations of identity were *maximally* differentiated (Figs. 4A and 6D; Supplementary Fig. S5). This initial identity representation was most pronounced in known object processing areas, including the ventral occipitotemporal cortex and early

visual cortex. Apparently, pre-existing representations based on life-long experience were sufficient to immediately provide a view-independent representation of synthetic shapes, which we had designed to be highly characteristic and discriminable. In contrast, neural representations of novelty were *minimally* differentiated in the first run and the first batch. As there was no systematic difference between recurring and non-recurring objects (and without time for plasticity), any residual initial discriminability of novelty must be attributed to chance.

Over subsequent runs and batches, the neural representation of object identity remained pronounced, but its quality declined steadily over time (Figs. 4A and 6D; Supplementary Fig. S5). Some decline in BOLD activity is not untypical for learning studies over multiple days and is commonly ascribed to repetition suppression, sparsification of responses, and/or diminishing attention or effort (e.g., Poldrack, 2000). However, while our results are consistent with such a scenario in non-identity-selective parcels, they do not support a general decline of activity in identity-selective parcels, as the response amplitudes and distances in these parcels declined only initially and subsequently remained stable (Fig. 6B, C).

In contrast, the neural representation of object novelty improved substantially over subsequent runs and batches. The time course was similar in both analyses (batch-by-batch and run-by-run), with the steepest improvement occurring over the first few runs (Figs. 4A and 6D; Supplementary Fig. S5). However, the detailed results revealed that this "improvement" (in discriminating non-recurring and recurring objects) actually reflected a deterioration in the representation of non-recurring objects (i.e., diminishing response distances, Fig. 6C).

In absolute terms, response amplitudes and distances were already larger for recurring objects and smaller for non-recurring objects during the first run and the difference increased over the next few runs (Fig. 6B, C). Apparently, recurring objects benefited from a "repetition enhancement," as the only immediate and systematic difference between recurring and non-recurring objects was the frequency of recurrence. Interestingly, this enhancement was comparable for "structured" and "unstructured" sequences, even though the repetition latencies were quite different (Supplementary Fig. S1B, C). Accordingly, we hypothesize that the enhancement was not merely a passive effect but rather a consequence of task relevance and cognitive engagement (Supplementary Fig. S1B, C).

As mentioned, the rates of change of identity and novelty representations differed systematically between cortical regions (Fig. 8C). Intriguingly, the rates of novelty *gain* and identity *loss* varied inversely over the cortical hierarchy: in early visual areas (V1, V2, V3, hV4), identity declined rapidly, whereas novelty grew slowly. At the opposite end, in the inferior frontal cortex (IFC) and anterior ventral temporal cortex (AIT), identity declined slowly, but novelty grew rapidly. In the higher visual cortex (VO, LO), both rates were intermediate.

It is informative to visualize the observed representational changes in two dimensions (Fig. 9), while approximately preserving the *relative* pairwise distances in the discriminative subspaces $\mathbb{S}$. This visualization makes clear

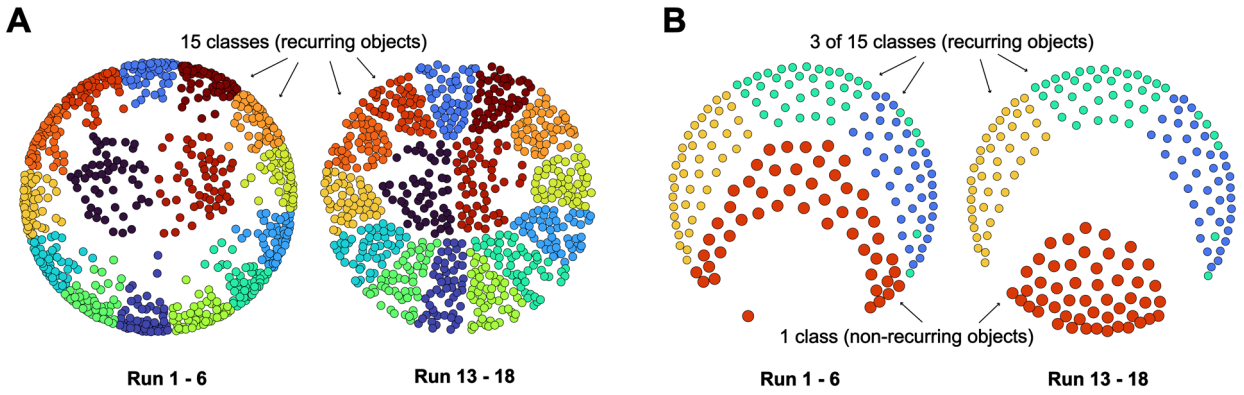

Fig. 9. Changes in the geometry of shape identity and novelty representations, visualized with multi-dimensional scaling. Symbols (colored circles) represent neural response patterns in a 14-dimensional space $\mathbb{S}$. Symbols are positioned such that pairwise distances reflect pairwise distances in $\mathbb{S}$. Response classes are distinguished by color and are represented by 50 randomly selected responses each. (A) Fifteen response classes to recurring objects in the first session (left, run 1–6) and the third session (right, run 13–18). Note that *recurring* response classes expand with learning to fill the available space. (The regions occupied by classes depend on the selected responses. "Inside" and "outside" classes can exchange positions). (B) Three response classes to recurring objects and one response class to non-recurring objects (larger symbols), in the first session (left, run 1–6) and the third session (right, run 13–18). Note that the *non-recurring* response class contracts with learning and shifts to the margins of the available space. Class positions are similar for other triplets of recurring classes.

that the neural representation of recurring objects expands between the beginning and the end of the experiment, filling the available representational space (Fig. 9A). The expansion explains our observation that the linear discriminability of object classes degrades but remains high. In contrast, the neural representation of non-recurring objects contracts between the beginning and the end of the experiment while also shifting to the margins of representational space, which explains why the linear discriminability of non-recurring objects improved over time (Fig. 9B). These two opposite developments may reflect both cognitive engagement and repetition frequency: representations may expand for objects that observers attempt to memorize and/or that recur frequently, but contract for objects that observers learn to ignore and/or that are rare.

In addition to relative changes in representational geometry indexed by linear discriminability, we established absolute changes in representational geometry, indexed by distances between response centroids in successive runs (see Fig. 7). The results were dominated by sampling noise, and the displacement of centroids was comparable to random jumps in a hypersphere while maintaining a given distance from its center. However, both absolute and relative centroid distances were slightly (and significantly) smaller than predicted by sampling noise, indicating that the representations were slightly more consistent and compact. The most interesting result of this analysis was that centroid distances were approximately 30% smaller for non-recurring than for recurring objects, highlighting again the representational disparity noted above.

4.4. Behavioral and cognitive changes with learning

The behavioral changes over three sessions of viewing sequences of objects included both increased classification performance ("familiar" or "novel") and decreased reaction times. Both behavioral measures changed rapidly during the first three runs of the first session and more slowly during the second and third sessions (Fig. 1). As described elsewhere (Kakaei et al., 2021), the classification of a particular object typically changed from (mostly) "novel" to (mostly) "familiar" at one identifiable point in time during the sessions, which we termed "onset of familiarity." This objective observation was consistent with the subjective reports of observers that they memorized all three-dimensional shapes one by one, such that every object became recognizable from all sides. Some observers also mentioned having assigned linguistic labels to individual recurring objects. After the three sessions, all observers were "familiar" with all recurring objects and could pick them out from an array of distractor objects.

Only some of these behavioral changes have obvious counterparts in the neural changes discussed above. First, the decrease of reaction times from under 2 s to under 1 s implies that observers spend less time actively evaluating the stimulus and more time passively observing it. However, the neural response of identity-selective parcels does not mirror this trend, as both response amplitudes and response differences stabilize after the first few runs (Fig. 6B, C). In the rest of the brain (non-identity-selective parcels), the neural responses do show a progressive decrease, but any attribution would be speculative.

Second, the increase in objective performance and in subjective "familiarity" was not mirrored directly in neural responses to recurring objects, as multivariate responses were sufficiently rich to identify such objects from the very start. However, multivariate responses were dispersed over the three sessions such as to fill more of the available space (see above). This growing response diversity is a plausible correlate of memory consolidation, that is, the formation of stable long-term memories in visually responsive cortical areas. When such memories are consolidated, one would expect that increased connectivity would enhance pattern completion over additional levels of representation, rendering network activity more complex (e.g., Steinberg & Sompolinsky, 2022). It is worth noting that this development was observed for both types of presentation sequences ("structured" and "unstructured"), suggesting that neural consolidation was due to task relevance and not merely to repetition latency.

Third, the increase in objective performance was mirrored indirectly in neural responses to *non-recurring* objects. Whereas these responses were initially comparable to recurring responses, they contracted over three sessions into a smaller part of the available space, thus becoming more stereotypical. As this part was comparably distant from all recurring responses, it lay at the margins of the representational space. The time course of classification performance corresponded best to this particular development in neural representations. Accordingly, this development was a plausible *indirect* correlate of memory consolidation, in the sense that visually responsive areas grew *less responsive* to other objects that failed to match the newly formed long-term memories.

5. CONCLUSION

We analyzed the cortical representation of visual objects in the multivariate hemodynamic responses of 758 brain parcels. For each parcel, we used linear discriminant analysis to map the $O(10^3)$-dimensional responses into a lower-dimensional subspace that optimally discriminated the 15 stimulus classes (recurring objects). Optimal subspaces

captured a large part of the correlated variance and overlapped substantially with the principal components of the responses. Typically, 2/3 of the principal component variance discriminated between stimulus classes (and thus coincided with the optimal subspace), while the remaining 1/3 was shared between stimulus classes. Our analyses revealed where and how the cortical representations of visual objects changed as visual expertise was being acquired and consolidated by the observers.

Our results were broadly consistent with other recent studies of visual expertise, which have highlighted the roles of three pathways or networks (Kravitz et al., 2011, 2013), an occipitotemporal pathway ("ventral pathway"), an occipitoparietal pathway ("dorsal pathway"), and a right frontoparietal network ("attention system"). Several studies linked behavioral performance to enhanced activity and/or representation in the frontoparietal network (Duyck et al., 2021; Poirier et al., 2006; Visconti di Oleggio Castello et al., 2021), as well as in the more anterior parts of the occipitotemporal pathway and the more dorsal parts of the occipitoparietal pathway (Christophel et al., 2017).

Due to our focus on object shape, our results do not speak directly to the modulation of cortical responses by expectation, such as "expectation suppression" or "surprise signalling" (Barron et al., 2016; Bell et al., 2016; Mayrhauser et al., 2014; Vinken et al., 2018). Moreover, in our paradigm, object presentations were never repeated exactly and every object presentation contained elements of surprise, as neither the object, nor the point of view, nor the direction of rotation could be anticipated by observers.

The most robust representations of object shape for both recurring objects ("identity") and non-recurring objects ("novelty") were observed in the ventral occipitotemporal cortex, at the intermediate levels of the shape processing hierarchy (Grill-Spector & Weiner, 2014; Perry & Fallah, 2014). Additionally, we found representations of object shape in "dorsal stream" cortical areas, consistent with the view that these areas encode goal- and task-related object features (Perry & Fallah, 2014).

The most novel aspect of our findings was changes in the geometry of cortical representations as visual expertise for recurring objects was being acquired and consolidated. In relative terms, distances between response classes decreased, and/or distances within classes increased, while observers repeatedly viewed and became familiar with the corresponding stimulus classes. This modest *decline* in stimulus encoding was however associated with an expansion (or *diversification*) in the distribution of responses within classes, so that responses of all classes taken together scattered more uniformly over the available representational space. Changes in cortical representations were quite different for stimuli that appeared only once and that observers did not attempt to memorize (non-recurring objects). Here, again in relative terms, distances between classes (non-recurring and recurring) increased and/or distances within classes (non-recurring) decreased. This steep *growth* in class encoding was associated with a substantial contraction (or *stereotypisation*) in the distribution of responses, in the sense that responses to non-recurring objects shifted to the margin of the available representational space.

We conclude that hemodynamic responses to novel object shapes immediately represent the differences between these shapes, even prior to learning, presumably reflecting life-long prior experience. When object shapes grow familiar with learning, hemodynamic responses to the same shapes become more diverse, whereas responses to different shapes remain comparably dissimilar from each other. Responses to control objects that are always novel develop quite differently in that they become less diverse relative to each other, but also more dissimilar from responses to familiar objects.

DATA AND CODE AVAILABILITY

Direct linear discriminant analysis and prevalence inference is available on github.com/cognitive-biology/DLDA. MR data will be made available upon request.

AUTHOR CONTRIBUTIONS

Ehsan Kakaei: Conceptualization, data curation, formal analysis, visualization, and writing of the original draft. Jochen Braun: Conceptualization, linear algebra, formal analysis, supervision, and reviewing & editing.

ACKNOWLEDGMENTS

We thank Claus Tempelmann, Martin Kanowski, and Denise Scheermann at the Magnetic Resonance Imaging Laboratory of the Department of Neurology of Otto-von-Guericke University, Magdeburg. We are grateful to Oliver Speck for providing essential support and a balanced perspective. We also thank Stepan Aleshin for helpful discussions and constructive comments. This study was funded by the federal state Saxony-Anhalt and the European Structural and Investment Funds (ESF, 2014–2020), project number ZS/2016/08/80645, as part of the doctoral program ABINEP (Analysis, Imaging and Modelling of Neuronal Processes).

DECLARATION OF COMPETING INTEREST

The authors are not aware of any competing interest.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available with the online version here: https://doi.org/10.1162/imag_a_00255

REFERENCES

Albers, K. J., Ambrosen, K. S., Liptrot, M. G., Dyrby, T. B., Schmidt, M. N., & Mørup, M. (2021). Using connectomics for predictive assessment of brain parcellations. *NeuroImage*, *238*, 118170. https://URL.org/10.1016/j.neuroimage.2021.118170

Allefeld, C., Görgen, K., & Haynes, J.-D. (2016). Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *NeuroImage*, *141*, 378–392. https://doi.org/10.1016/j.neuroimage.2016.07.040

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*(1), 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Barron, H. C., Garvert, M. M., & Behrens, T. E. J. (2016). Repetition suppression: A means to index neural representations using bold? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1705), 20150355. https://doi.org/10.1098/rstb.2015.0355

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, *23*(2), 137–152. https://doi.org/10.1109/TMI.2003.822821

Bell, A. H., Summerfield, C., Morin, E. L., Malecek, N. J., & Ungerleider, L. G. (2016). Encoding of stimulus probability in macaque inferior temporal cortex. *Current Biology*, *26*(17), 2280–2290. https://doi.org/10.1016/j.cub.2016.07.007

Bi, Y., Wang, X., & Caramazza, A. (2016). Object domain and modality in the ventral visual pathway. *Trends in Cognitive Sciences*, *20*(4), 282–290. https://doi.org/10.1016/j.tics.2016.02.002

Blake, R., & Braun, J. (2009). Visual perception: Tracking the elusive footprints of awareness. *Current Biology*, *19*(1), R30–R32. https://doi.org/10.1016/j.cub.2008.11.009

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brants, M., Bulthé, J., Daniels, N., Wagemans, J., & de Beeck, H. P. O. (2016). How learning might strengthen existing visual object representations in human object-selective cortex. *NeuroImage*, *127*, 74–85. https://doi.org/10.1016/j.neuroimage.2015.11.063

Brants, M., Wagemans, J., & Op de Beeck, H. P. (2011). Activation of fusiform face area by greebles is related to face similarity but not expertise. *Journal of Cognitive Neuroscience*, *23*(12), 3949–3958. https://doi.org/10.1162/jocn_a_00072

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trends in Cognitive Sciences*, *10*(4), 159–166. https://doi.org/10.1016/j.tics.2006.02.004

Cetron, J. S., Connolly, A. C., Diamond, S. G., May, V. V., Haxby, J. V., & Kraemer, D. J. (2019). Decoding individual differences in STEM learning from functional MRI data. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-10053-y

Charest, I., & Kriegeskorte, N. (2015). The brain of the beholder: Honouring individual representational idiosyncrasies. *Language, Cognition and Neuroscience*, *30*(4), 367–379. https://doi.org/10.1080/23273798.2014.1002505

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The distributed nature of working memory. *Trends in Cognitive Sciences*, *21*(2), 111–124. https://doi.org/10.1016/j.tics.2016.12.007

Collins, E., & Behrmann, M. (2020). Exemplar learning reveals the representational origins of expert category perception. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(20), 11167–11177. https://doi.org/10.1073/pnas.1912734117

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., Abdi, H., & Haxby, J. V. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, *32*(8), 2608–2618. https://doi.org/10.1523/JNEUROSCI.5547-11.2012

de Beeck, H. P. O., & Baker, C. I. (2010). The neural basis of visual object learning. *Trends in Cognitive Sciences*, *14*(1), 22–30. https://doi.org/10.1016/j.tics.2009.11.002

de Beeck, H. P. O., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *Journal of Neuroscience*, *26*(50), 13025–13036. https://doi.org/10.1523/JNEUROSCI.2481-06.2006

Dornas, J. V., & Braun, J. (2018). Finer parcellation reveals detailed correlational structure of resting-state fMRI signals. *Journal of Neuroscience Methods*, *294*, 15–33. https://doi.org/10.1016/j.jneumeth.2017.10.020

Duyck, S., Martens, F., Chen, C.-Y., & Op de Beeck, H. (2021). How visual expertise changes representational geometry: A behavioral and neural perspective. *Journal of Cognitive Neuroscience*, *33*(12), 2461–2476. https://doi.org/10.1162/jocn_a_01778

Eger, E., Ashburner, J., Haynes, J.-D., Dolan, R. J., & Rees, G. (2008). fMRI activity patterns in human loc carry information about object exemplars within category. *Journal of Cognitive Neuroscience*, *20*(2), 356–370. https://doi.org/10.1162/jocn.2008.20019

Freud, E., Culham, J. C., Plaut, D. C., & Behrmann, M. (2017). The large-scale organization of shape processing in the ventral and dorsal pathways. *eLife*, *6*, e27576. https://doi.org/10.7554/eLife.34464

Gauthier, I., & Tarr, M. J. (2016). Visual object recognition: Do we (finally) know more now than we did? *Annual Review of Vision Science*, *2*, 377–396. https://doi.org/10.1146/annurev-vision-111815-114621

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*(6), 568–573. https://doi.org/10.1038/9224

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*(5), 555–562. https://doi.org/10.1038/nn1224

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747

Harel, A., Kravitz, D., & Baker, C. I. (2013). Beyond perceptual expertise: Revisiting the neural substrates of expert object recognition. *Frontiers in Human Neuroscience*, *7*, 885. https://doi.org/10.1167/14.10.820

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage*, *62*(2), 852–855. https://doi.org/10.1016/j.neuroimage.2012.03.016
Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736
Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, *310*(5749), 863–866. https://doi.org/10.1126/science.1117593
Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132
Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6
Jeong, S. K., & Xu, Y. (2016). Behaviorally relevant abstract object identity representation in the human parietal cortex. *Journal of Neuroscience*, *36*(5), 1607–1619. https://doi.org/10.1523/JNEUROSCI.1016-15.2016
Kakaei, E., Aleshin, S., & Braun, J. (2021). Visual object recognition is facilitated by temporal community structure. *Learning & Memory*, *28*(5), 148–152. https://doi.org/10.1101/lm.053306.120
Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience*, *11*(2), 224–231. https://doi.org/10.1038/nn2036
Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, *74*(6), 1114–1124. https://doi.org/10.1016/j.neuron.2012.04.036
Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, *12*(4), 217–230. https://doi.org/10.1038/nrn3008
Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011
Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103
Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008
Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043
Kumar, M., Anderson, M. J., Antony, J. W., Baldassano, C., Brooks, P. P., Cai, M. B., Chen, P.-H. C., Ellis, C. T., Henselman-Petrusek, G., Huberdeau, D., Hutchinson, J. B., Li, Y. P., Lu, Q., Manning, J. R., Mennen, A. C., Nastase, S. A., Richard, H., Schapiro, A. C., Schuck, N. W., … Norman, K. A. (2022). BrainIAK: The brain imaging analysis kit. *Aperture Neuro*, *2021*(4), 1–19. https://doi.org/10.52294/31bb5b68-2184-411b-8c00-a1dacb61e1da
Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*, *62*(2), 281–290. https://doi.org/10.1016/j.neuron.2009.02.025
Martens, F., Bulthé, J., van Vliet, C., & de Beeck, H. O. (2018). Domain-general and domain-specific neural changes underlying visual expertise. *NeuroImage*, *169*, 80–93. https://doi.org/10.1016/j.neuroimage.2017.12.013
Mayrhauser, L., Bergmann, J., Crone, J., & Kronbichler, M. (2014). Neural repetition suppression: Evidence for perceptual expectation in object-selective regions. *Frontiers in Human Neuroscience*, *8*, 225. https://doi.org/10.3389/fnhum.2014.00225
McGugin, R. W., Gatenby, J. C., Gore, J. C., & Gauthier, I. (2012). High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(42), 17063–17068. https://doi.org/10.1073/pnas.1116333109
Mutlu, M. C., Kakaei, E., & Braun, J. (2022). Candidate areas for initiating spontaneous reversals of kinetic depth: Inferior frontal cortex and insula. In *Bernstein Conference 2022* (PIII–64). Berlin, Germany. https://doi.org/10.12751/nncn.bc2022.208
Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience*, *14*(6), 667–685. https://doi.org/10.1093/scan/nsz037
Nestor, A., Plaut, D. C., & Behrmann, M. (2016). Feature-based face representations and image reconstruction from behavioral and neural data. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(2), 416–421. https://doi.org/10.1073/pnas.1514551112
Patel, A. X., Kundu, P., Rubinov, M., Jones, P. S., Vértes, P. E., Ersche, K. D., Suckling, J., & Bullmore, E. T. (2014). A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *NeuroImage*, *95*, 287–304. https://doi.org/10.1016/j.neuroimage.2014.03.012
Perry, C. J., & Fallah, M. (2014). Feature integration and object representations along the dorsal stream visual hierarchy. *Frontiers in Computational Neuroscience*, *8*, 84. https://doi.org/10.3389/fncom.2014.00084
Poirier, C. C., De Volder, A. G., Tranduy, D., & Scheiber, C. (2006). Neural changes in the ventral and dorsal visual streams during pattern recognition learning. *Neurobiology of Learning and Memory*, *85*(1), 36–43. https://doi.org/10.1016/j.nlm.2005.08.006
Poldrack, R. A. (2000). Imaging brain plasticity: Conceptual and methodological issues—A theoretical review. *NeuroImage*, *12*(1), 1–13. https://doi.org/10.1006/nimg.2000.0596
Roth, Z. N., & Zohary, E. (2015). Fingerprints of learned object recognition seen in the fMRI activation patterns of lateral occipital complex. *Cerebral Cortex*, *25*(9), 2427–2439. https://doi.org/10.1093/cercor/bhu042
Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062
Smith, S. M., & Brady, J. M. (1997). Susan—A new approach to low level image processing. *International Journal of Computer Vision*, *23*(1), 45–78. https://doi.org/10.1023/A:1007963824710

Smith, S. M., Vidaurre, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M., Miller, K. L., Nichols, T. E., Robinson, E. C., Salimi-Khorshidi, G., Woolrich, M. W., Barch, D. M., Uğurbil, K., & Van Essen, D. C. (2013). Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences*, *17*(12), 666–682. https://doi.org/10.1016/j.tics.2013.09.016

Steinberg, J., & Sompolinsky, H. (2022). Associative memory of structured knowledge. *Scientific Reports*, *12*(1), 21808. https://doi.org/10.1038/s41598-022-25708-y

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*, *16*(1), 55–61. https://doi.org/10.1038/nrn3857

Vinken, K., Op de Beeck, H. P., & Vogels, R. (2018). Face repetition probability does not affect repetition suppression in macaque inferotemporal cortex. *The Journal of Neuroscience*, *38*(34), 7492–7504. https://doi.org/10.1523/jneurosci.0462-18.2018

Visconti di Oleggio Castello, M., Haxby, J. V., & Gobbini, M. I. (2021). Shared neural codes for visual and semantic information about familiar faces in a common representational space. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(45), e2110474118. https://doi.org/10.1073/pnas.2110474118

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, *25*(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277

Weiner, K. S., & Zilles, K. (2016). The anatomical and functional specialization of the fusiform gyrus. *Neuropsychologia*, *83*, 48–62. https://doi.org/10.1016/j.neuropsychologia.2015.06.033

Wong, A. C.-N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for facelike expertise with objects: Becoming a ziggerin expert—But which type? *Psychological Science*, *20*(9), 1108–1117. https://doi.org/10.1111/j.1467-9280.2009.02430.x

Wong, Y. K., Folstein, J. R., & Gauthier, I. (2012). The nature of experience determines object representations in the visual system. *Journal of Experimental Psychology: General*, *141*(4), 682. https://doi.org/10.1037/a0027822

Wurm, M. F., & Caramazza, A. (2022). Two 'what' pathways for action and object recognition. *Trends in Cognitive Sciences*, *26*(2), 103–116. https://doi.org/10.1016/j.tics.2021.10.003

Ye, J., Xiong, T., & Madigan, D. (2006). Computational and theoretical analysis of null space and orthogonal linear discriminant analysis. *Journal of Machine Learning Research*, *7*(7), 1183–1204. http://jmlr.org/papers/v7/ye06a.html

Yildirim, I., Wu, J., Kanwisher, N., & Tenenbaum, J. (2019). An integrative computational architecture for object-driven cortex. *Current Opinion in Neurobiology*, *55*, 73–81. https://doi.org/10.1016/j.conb.2019.01.010

Yu, H., & Yang, J. (2001). A direct LDA algorithm for high-dimensional data—With application to face recognition. *Pattern Recognition*, *34*(10), 2067–2070. https://doi.org/10.1016/S0031-3203(00)00162-X

Yue, X., Tjan, B. S., & Biederman, I. (2006). What makes faces special? *Vision Research*, *46*(22), 3802–3811. https://doi.org/10.1016/j.visres.2006.06.017

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

(*Appendix* follows on next page)

APPENDIX

Appendix Table A1. List of identity-selective parcels and their anatomical region.

AAL region	Parcel	MNI			α (%)	β	$p^\star$	$p^\star$	$p^\star$	Topog.
	No.	x	y	z	Identity	Novelty	Struct.	Unstruct.	Both	Assign.
Precentral	14	-51	7	36	8.4	0.33	0.006	10^{-5}	10^{-5}	-
Superior frontal	*56*	25	-9	64	7.9	0.18	n.s.	n.s.	4×10^{-5}	-
Inferior frontal (opercular)	143	38	11	31	7.5	0.27	5×10^{-5}	5×10^{-3}	0.02	IFC
	146	52	10	22	9	0.44	0.05	n.s.	10^{-5}	IFC
Inferior frontal (triangular)	163	51	23	20	7.9	0.34	10^{-5}	10^{-5}	10^{-5}	IFC
Parahippocampal	325	26	-38	-9	7.5	0.05	0.007	2×10^{-4}	10^{-5}	-
Calcarine	331	-2	-91	-1	16.9	0.01	10^{-5}	10^{-5}	10^{-5}	V1v
	333	-2	-94	-4	10.2	-0.01	10^{-5}	5×10^{-3}	10^{-5}	V1v
	335	-4	-86	7	12.6	0.11	10^{-5}	10^{-5}	10^{-5}	V1d
	336	-5	-99	-8	12.2	0.03	10^{-5}	10^{-5}	10^{-5}	V1v
	337	-12	-99	-5	13.8	0.10	10^{-5}	10^{-5}	10^{-5}	V1d
	338	-7	-75	10	8.7	0.08	10^{-5}	10^{-5}	10^{-5}	-
	342	17	-83	11	9.8	0.07	10^{-5}	10^{-5}	10^{-5}	-
	344	9	-85	6	13.8	0.10	10^{-5}	10^{-5}	10^{-5}	V1v
	345	15	-91	1	14.3	0.09	10^{-5}	10^{-5}	10^{-5}	V1d
	347	18	-99	-1	12.1	0.07	10^{-5}	10^{-5}	10^{-5}	V1d
	348	11	-72	11	9.1	0.06	n.s.	10^{-5}	2×10^{-4}	-
Cuneus	350	-7	-85	26	10.1	0.07	10^{-5}	10^{-5}	10^{-5}	-
	352	0	-94	25	9.1	0.05	10^{-5}	10^{-5}	10^{-5}	V2d
	354	-2	-79	22	8.6	0.07	0.003	10^{-5}	10^{-5}	-
	355	-4	-92	25	11.9	0.09	n.s.	10^{-5}	0.02	V2d
	356	12	-91	18	13	0.15	10^{-5}	10^{-5}	10^{-5}	V2d
	357	15	-97	10	13.8	0.08	10^{-5}	10^{-5}	10^{-5}	V2d
Lingual	363	-12	-65	-5	9.5	0.10	10^{-5}	10^{-5}	10^{-5}	V3v
	364	-15	-95	-16	10.9	0.02	10^{-4}	10^{-5}	10^{-5}	V2v
	367	-29	-89	-16	11.5	0.07	2×10^{-5}	10^{-5}	10^{-5}	hV4
	368	-17	-85	-12	14.9	0.26	10^{-5}	10^{-5}	10^{-5}	V2v
	370	-22	-65	-5	9.7	0.16	10^{-5}	10^{-5}	10^{-5}	VO2
	371	-12	-79	-8	13.9	0.12	10^{-5}	10^{-5}	10^{-5}	V2v
	372	-6	-74	2	12	0.11	10^{-5}	10^{-5}	10^{-5}	V1v
	373	16	-81	-7	14.1	0.11	10^{-5}	10^{-5}	10^{-5}	V3v
	375	11	-72	-4	12.5	0.04	10^{-5}	10^{-5}	10^{-5}	V2v
	377	21	-58	-3	9.3	0.24	10^{-5}	10^{-5}	10^{-5}	-
	378	13	-52	2	7.6	0.05	n.s.	10^{-5}	2×10^{-4}	-
	379	16	-88	-10	13.4	0.14	10^{-5}	10^{-5}	10^{-5}	V2v
	380	27	-91	-16	9.2	0.06	n.s.	0.01	0.01	-
	381	17	-98	-10	9.3	0.04	10^{-5}	n.s.	0.005	V1v
	383	14	-56	-6	7.9	0.05	2×10^{-5}	0.01	10^{-5}	-
Occipital (superior)	384	-18	-84	25	10.6	0.08	10^{-5}	10^{-5}	10^{-5}	V3a
	385	-16	-85	41	9.8	0.12	10^{-5}	10^{-5}	10^{-5}	IPS0
	386	-18	-69	29	7.7	0.04	n.s.	10^{-5}	0.005	-
	387	-16	-95	23	13.5	0.12	10^{-5}	10^{-5}	10^{-5}	V3a
	388	-22	-75	34	9.1	0.21	10^{-5}	10^{-5}	10^{-5}	IPS1
	389	-11	-95	9	12.9	0.05	10^{-5}	10^{-5}	10^{-5}	V2d
	390	22	-90	24	13.6	0.10	10^{-5}	10^{-5}	10^{-5}	V3a
	391	21	-98	14	13.8	0.11	10^{-5}	10^{-5}	10^{-5}	V2d
	392	27	-85	40	9.2	0.12	10^{-5}	10^{-5}	10^{-5}	IPS0
	393	*25*	-67	33	8	0.18	n.s.	n.s.	0.005	-
	394	24	-75	21	8.7	0.11	10^{-5}	10^{-5}	10^{-5}	-
	395	23	-79	33	10	0.12	10^{-5}	10^{-5}	10^{-5}	-
	396	29	-70	43	9.2	0.21	0.03	10^{-5}	10^{-5}	IPS1
Occipital (middle)	397	-28	-77	27	10.1	0.23	10^{-5}	10^{-5}	10^{-5}	IPS0
	398	-28	-72	34	9.1	0.32	n.s.	10^{-5}	10^{-5}	IPS1
	400	-38	-86	4	12	0.15	10^{-5}	10^{-5}	10^{-5}	LO2
	402	-33	-87	19	12	0.24	10^{-5}	10^{-5}	10^{-5}	V3b

Appendix Table A1. (*Continued*)

AAL region	Parcel	MNI			α (%)	β	$p^\star$	$p^\star$	$p^\star$	Topog.
	No.	x	y	z	Identity	Novelty	Struct.	Unstruct.	Both	Assign.
	403	-30	-78	2	7.7	0.05	n.s.	n.s.	0.02	-
	404	-27	-94	1	12.7	0.15	10^{-5}	10^{-5}	10^{-5}	V3d
	405	-16	-100	1	14	0.13	10^{-5}	10^{-5}	10^{-5}	V2d
	406	-40	-75	15	8.9	0.14	10^{-5}	10^{-5}	10^{-5}	-
	407	-27	-83	14	11.3	0.19	10^{-5}	10^{-5}	10^{-5}	IPS0
	408	-24	-93	13	13.6	0.19	10^{-5}	10^{-5}	10^{-5}	V3d
	410	-38	-83	23	9.9	0.21	10^{-5}	10^{-5}	10^{-5}	-
	412	-46	-77	5	10.1	0.14	10^{-5}	10^{-5}	10^{-5}	hMT
	413	33	-88	7	13.2	0.23	10^{-5}	10^{-5}	10^{-5}	LO1
	414	33	-96	4	10.7	0.17	10^{-5}	10^{-5}	10^{-5}	V3d
	415	39	-81	14	10.6	0.19	10^{-5}	10^{-5}	10^{-5}	V3b
	416	44	-78	5	10.9	0.22	10^{-5}	10^{-5}	10^{-5}	LO2
	418	32	-86	23	11.9	0.22	10^{-5}	10^{-5}	10^{-5}	V3b
	420	34	-69	32	8.5	0.34	n.s.	10^{-5}	0.02	-
	421	32	-76	27	10	0.28	10^{-5}	10^{-4}	10^{-5}	IPS0
Occipital (inferior)	423	-50	-68	-14	9.4	0.22	10^{-5}	10^{-5}	10^{-5}	-
	424	-31	-83	-8	12.6	0.22	10^{-5}	10^{-5}	10^{-5}	-
	425	-22	-95	-9	12.3	0.11	10^{-5}	10^{-5}	10^{-5}	-
	426	-42	-73	-8	11.6	0.21	10^{-5}	10^{-5}	10^{-5}	-
	428	36	-85	-7	13.2	0.21	10^{-5}	10^{-5}	10^{-5}	-
	430	42	-73	-9	11.4	0.25	10^{-5}	10^{-5}	10^{-5}	-
Fusiform	432	-27	-71	-11	12.2	0.25	10^{-5}	10^{-5}	10^{-5}	VO2
	435	-33	-77	-17	12.4	0.15	10^{-5}	10^{-5}	10^{-5}	hV4
	436	-31	-53	-13	8.8	0.26	10^{-5}	n.s.	10^{-5}	PHC1
	438	-41	-56	-17	8.9	0.26	n.s.	10^{-5}	0.01	-
	440	-36	-63	-16	9.6	0.25	0.001	10^{-5}	10^{-5}	-
	442	28	-74	-11	12.8	0.31	10^{-5}	10^{-5}	10^{-5}	hV4
	443	36	-71	-16	10.8	0.25	0.02	10^{-5}	10^{-5}	hV4
	447	29	-47	-14	8.3	0.31	0.002	10^{-5}	10^{-5}	PHC2
	450	41	-48	-20	8.4	0.30	0.006	10^{-5}	0.02	-
	452	28	-59	-12	9.7	0.35	10^{-5}	10^{-5}	10^{-5}	VO2
Postcentral	476	60	-18	37	9.2	0.34	10^{-5}	10^{-5}	10^{-5}	-
	478	42	-31	49	8.5	0.25	10^{-5}	0.02	10^{-5}	-
	484	30	-40	61	8.5	0.20	10^{-5}	10^{-5}	10^{-5}	-
Parietal (superior)	494	-26	-61	61	9.1	0.09	10^{-5}	10^{-5}	10^{-5}	-
	495	-27	-53	68	8	0.09	0.02	0.05	2×10^{-5}	-
	497	-21	-67	47	9.3	0.19	10^{-5}	10^{-5}	10^{-5}	IPS1
	498	-15	-69	50	8.5	0.12	0.01	10^{-5}	10^{-5}	IPS2
	499	-28	-69	50	8.9	0.29	0.001	3×10^{-5}	10^{-5}	-
	501	-17	-79	50	8.7	0.11	10^{-4}	3×10^{-5}	10^{-5}	IPS1
	502	30	-60	64	8.8	0.18	0.002	5×10^{-4}	10^{-5}	IPS3
	504	34	-62	57	9.3	0.30	5×10^{-5}	0.003	10^{-5}	-
	506	33	-50	58	9.8	0.34	10^{-5}	10^{-5}	10^{-5}	-
	507	21	-57	74	8.1	0.09	n.s.	10^{-5}	5×10^{-4}	-
	509	31	-73	53	8.6	0.13	n.s.	10^{-5}	0.001	-
	510	21	-74	55	8.8	0.14	0.06	0.07	2×10^{-5}	IPS1
	511	20	-65	54	9.2	0.17	10^{-5}	10^{-5}	10^{-5}	IPS2
Parietal (inferior)	514	-45	-30	42	8.6	0.24	0.002	10^{-5}	10^{-5}	-
	516	-32	-75	44	8	0.19	0.01	5×10^{-4}	10^{-5}	-
	521	-39	-47	42	8.6	0.17	*n.s.*	10^{-4}	0.005	-
	522	-32	-46	50	8.5	0.12	10^{-5}	10^{-5}	10^{-5}	-
	523	-31	-53	46	8.4	0.18	10^{-5}	n.s.	2×10^{-5}	-
	527	46	-39	50	8.6	0.35	n.s.	10^{-5}	0.001	-
	529	37	-49	46	8.7	0.34	0.06	0.003	2×10^{-5}	-
	530	35	-44	51	9.2	0.29	5×10^{-4}	10^{-5}	10^{-5}	-
Supramarginal	536	-61	-28	34	8.4	0.24	0.003	10^{-5}	10^{-5}	-
	539	44	-34	41	8.7	0.20	0.001	0.001	10^{-5}	-
	542	63	-24	37	8.5	0.22	10^{-5}	n.s.	10^{-4}	-

Appendix Table A1. (*Continued*)

AAL region	Parcel	MNI			α (%)	β	$p^\star$	$p^\star$	$p^\star$	Topog.
	No.	x	y	z	Identity	Novelty	Struct.	Unstruct.	Both	Assign.
Angular	557	34	-60	44	8.7	0.38	n.s.	10^{-5}	0.005	-
Precuneus	561	-5	-77	53	7.9	0.06	10^{-5}	2×10^{-5}	10^{-5}	-
	573	-9	-71	54	8.1	0.08	10^{-5}	0.003	10^{-5}	-
	576	14	-71	45	7.9	0.12	10^{-5}	10^{-5}	10^{-5}	-
Temporal (middle)	678	-45	-67	11	8.9	0.12	10^{-5}	n.s.	10^{-4}	-
	685	-49	-62	0	9.3	0.19	10^{-5}	10^{-5}	10^{-5}	-
	701	52	-59	3	8.9	0.20	10^{-5}	10^{-5}	10^{-5}	-
	717	49	-69	4	10.5	0.18	10^{-5}	10^{-5}	10^{-5}	-
Temporal (inferior)	728	-54	-58	-11	8.5	0.25	2×10^{-4}	10^{-5}	10^{-5}	AIT
	732	-45	-52	-13	9.3	0.30	10^{-5}	10^{-5}	10^{-5}	AIT
	755	46	-53	-11	9.7	0.42	10^{-5}	10^{-5}	10^{-5}	AIT

Parcel ID, geometrical centroid x/y/z in MNI coordinates, average classification accuracy α, average novelty rate β, corrected significance $p^\star$ in structured or unstructured conditions ($n = 8$), corrected significance $p^\star$ in both conditions ($n = 16$), and topographical assignment, if any.

Chapter 4

Incidental learning of predictive temporal context within cortical representations of visual shape

Imaging Neuroscience
an open access journal

Research Article

Incidental learning of predictive temporal context within cortical representations of visual shape

Ehsan Kakaei[a,b,c], Jochen Braun[b,c]

[a]European Structural and Investment Funds Graduate School on Analysis, Imaging, and Modelling of Neuronal and Inflammatory Processes, Otto-von-Guericke University, Magdeburg, Germany
[b]Institute of Biology, Otto-von-Guericke University, Magdeburg, Germany
[c]Center for Behavioral Brain Sciences, Otto-von-Guericke University, Magdeburg, Germany

Corresponding Author: Ehsan Kakaei (ehsankakaei91@gmail.com)

ABSTRACT

Objective: Incidental learning of spatiotemporal regularities and consistencies—also termed 'statistical learning'—may be important for discovering the causal principles governing the world. We studied statistical learning of temporal structure simultaneously at two time-scales: the presentation of synthetic visual objects (3 s) and predictive temporal context (30 s) in the order of appearance of such objects.
Methods: Visual objects were complex and rotated in three dimensions about varying axes. Observers viewed fifteen (15) objects recurring many times each, intermixed with other objects that appeared only once, while whole-brain BOLD activity was recorded. Over three successive days, observers grew familiar with the recurring objects and reliably distinguished them from others. As reported elsewhere (Kakaei & Braun, 2024), representational similarity analysis (RSA) of multivariate BOLD activity revealed 124 'object-selective' brain parcels with selectivity for recurring objects, located mostly in the ventral occipitotemporal cortex and the parietal cortex.
Main results: Here, we extend RSA to the representation of predictive temporal context, specifically "temporal communities" formed by objects that tended to follow each other. After controlling for temporal proximity, we observed 27 'community-sensitive' brain parcels, in which pairwise distances between multivariate responses reflected community structure, either *positively* (smaller distances within than between communities) or *negatively* (larger distances within). Among object-selective parcels, 11 parcels were *positively* community-sensitive in the primary visual cortex (2 parcels), the ventral occipital, lingual, or fusiform cortex (8 parcels), and the inferior temporal cortex (1 parcel). Among non-object-selective parcels, 12 parcels were *negatively* community-sensitive in the superior, middle, and medial frontal cortex (6 parcels), the insula (2 parcels), the putamen (1 parcel), and in the superior temporal or parietal cortex (3 parcels).
Conclusion: We conclude that cortical representations of object shape and of predictive temporal context are largely coextensive along the ventral occipitotemporal cortex.

Keywords: incidental learning, statistical learning, functional imaging, representational similarity, multi-voxel activity

1. INTRODUCTION

Even when sensory stimuli are experienced passively—without task or reward—they can modify the underlying neural pathways and alter subsequent sensory performance and behavior (e.g., Conway & Christiansen, 2005; Lengyel et al., 2019; Li & DiCarlo, 2012). This incidental and automatic type of plasticity had been termed 'statistical learning' or 'implicit learning' (for reviews, see Aslin, 2017; Fiser & Lengyel, 2022; Perruchet, 2019; Perruchet & Pacton, 2006; Saffran & Kirkham, 2018; A. Schapiro &

Received: 13 June 2024 Revision: 22 July 2024 Accepted: 5 August 2024 Available Online: 12 August 2024

Turk-Browne, 2015). Some theories of cognitive development hypothesize that incidental learning during everyday experience captures the causal processes and relationships underlying sensory observations at a more abstract level (Kemp & Tenenbaum, 2008; Tenenbaum et al., 2011). If so, statistical learning might contribute to higher cognitive function by acquiring the quality of a "structural learning" that could underpin learning from examples, generalizing between domains, or gaining causal insight and understanding (Lake et al., 2017; Shafto et al., 2011).

A well-studied instance of incidental learning is the view-invariance of visual object recognition (for reviews, see DiCarlo et al., 2012; Gauthier & Tarr, 2016; Logothetis & Sheinberg, 1996). Humans and non-human primates typically recognize visual objects from different viewing directions and distances, presumably relying on characteristic features and/or their spatial relationships. This perceptual invariance can be modified rapidly by the experience of contiguous sequences of different views, demonstrating dependence on learning (Tian & Grill-Spector, 2015; Wallis & Bülthoff, 2001; Wallis et al., 2009). The neural representation of visual shape in the ventral occipitotemporal cortex is similarly view-invariant and equally subject to modification by the recent experience of (natural or unnatural) sequences of views (Jia et al., 2021; Li & DiCarlo, 2008, 2010, 2012; Op de Beeck & Baker, 2010; Van Meel & Op de Beeck, 2018, 2020).

Incidental learning is not limited to individual objects but extends also to spatiotemporal configurations of multiple objects. When human observers experience temporal sequences or spatial arrays of visual objects, task-irrelevant statistical regularities and contingencies are learned rapidly (within minutes), as can be revealed by subsequent behavioral tests (Fiser & Aslin, 2001, 2002, 2005; Kakaei et al., 2021; Sáringer et al., 2022; Turk-Browne et al., 2005, 2009). In non-human primates, the experience of task-irrelevant temporal dependencies modifies object-specific responses of neurons in visual areas of the ventral temporal cortex, but also in multimodal areas of the medial temporal lobe (Erickson & Desimone, 1999; Kaposvari et al., 2018; Meyer et al., 2014; Miyashita, 1988; Sakai & Miyashita, 1991). In human observers, functional imaging evidence reveals that task-irrelevant temporal dependencies can modulate BOLD responses in visually selective areas of the ventral occipital cortex, as well as in multimodal areas such as the medial temporal lobe, hippocampus, and basal ganglia (Gheysen et al., 2011; Giorgio et al., 2018; Hindy et al., 2016; Hsieh et al., 2014; Karlaftis et al., 2019; Turk-Browne et al., 2009, 2010; A. C. Schapiro et al., 2012; R. Wang et al., 2017). Statistical learning goes beyond first-order dependencies (between immediate temporal neighbors) and extends to higher-order dependencies (between more distant neighbors). For example, A. C. Schapiro et al. (2013, 2016) demonstrated statistical learning of clusters of dependencies ("temporal communities") and observed BOLD correlates of this predictive temporal context in associative areas of the frontal and temporal lobes and in the hippocampus, but not in visually selective areas of the ventral occipitotemporal cortex.

Here, we investigate statistical learning by human observers with temporal sequences of visual objects, seeking to compare neural correlates of learning at the levels of individual visual objects and of higher-order temporal dependencies. Unlike previous work, we focus on the visual pathways in the ventral occipitotemporal cortex, the major neural substrate of visual experience and long-term memory (reviewed by Bi et al., 2016; Grill-Spector & Weiner, 2014; Kravitz et al., 2013; Weiner & Zilles, 2016). We hypothesize that learning of visual shapes (i.e., spatiotemporal relationships of characteristic features) might interact with the learning of the context in which such shapes appear (i.e., spatiotemporal configurations of distinct shapes) (Miyashita, 1988).

Our visual stimuli were synthetic, three-dimensional objects of unique and characteristic shape that rotated slowly about varying axes. Over three successive sessions/days, observers viewed 15 'recurring' objects approximately 200 times each, as well as 360 'non-recurring' objects once each, while attempting to classify each presented object as either 'familiar' (recurring) or 'novel' (non-recurring). As reported previously (Kakaei & Braun, 2024; Kakaei et al., 2021), observers quickly gained familiarity with recurring objects and learned to recognize their characteristic shape from all points of view.

We recorded whole-brain BOLD activity during all three sessions/days and analyzed this activity in terms 758 functionally defined brain parcels (Dornas & Braun, 2018), which on average comprised approximately 200 voxels and $1.7cm^3$ of gray matter. 'Representational similarity analysis' (RSA; Haxby, 2012; Kriegeskorte et al., 2008) was used to quantify the information encoded by the BOLD activity of each brain parcel, specifically, the $9s$ of multivariate activity following object presentation. For every brain parcel, this analysis was carried out in a lower-dimensional subspace chosen to maximize the differences between BOLD responses ('optimal subspace'; Yu & Yang, 2001). A cross-validated analysis identified 124 (of 758) brain parcels that were 'identity-selective' in the sense that object identity could be decoded from BOLD activity in the majority of observers: 90 parcels in the ventral occipitotemporal cortex, 28 parcels in the parietal cortex, and 6 parcels in frontal and other regions. The detailed results are reported in a companion study (Kakaei & Braun, 2024).

To investigate the effect of predictive temporal context, we manipulated the order in which objects were

presented, adapting the paradigm of A. C. Schapiro et al. (2013). For three sessions/days, the sequence of recurring objects was generated such as to form 'temporal communities' in the sense that every object was likely to be followed by other objects from the same community ("structured condition"). For three further sessions/days, the presentation sequence was fully random so that every object was equally likely to be followed by any other object ("unstructured condition").

To ascertain the effect of 'temporal communities' on multivariate BOLD activity, we analyzed pairwise distances in the 'optimal subspace' and established 'community sensitivity' in terms of the ratio of average response distances to objects in the same community and in different communities. After controlling for effects of temporal proximity, we established significant 'community sensitivity' in 27 of 758 cortical parcels. In particular, we observed *positive* community sensitivity (i.e., smaller distances within than between communities) in 11 'identity-selective' parcels in the ventral occipitotemporal cortex, as well as *negative* community sensitivity (i.e., larger distances within than between communities) in 12 non-identity-selective parcels of the frontal cortex, the insula, the putamen, and the superior temporal or parietal cortex. A further 4 parcels exhibited other combinations of community-sensitivity and identity-selectivity.

We conclude that the ventral occipitotemporal cortex harbors largely coextensive representations of both the identity of objects and of statistical regularities in the order of their appearance.

2. METHODS

The experimental paradigm and procedure are described in detail elsewhere (Kakaei & Braun, 2024). Here, we only summarize the most pertinent aspects.

2.1. Observers and behavior

Eight healthy observers (4 female and 4 male; aged 25 to 32 years) took part in behavioral training ('sham experiment', two sessions per observer), the functional imaging experiment ('main experiment', six scanning sessions per observer), and a final behavioral assessment (two sessions). All participants were paid and gave informed consent. Ethical approval was granted under Chiffre 30/21 by the ethics committee of the Faculty of Medicine of the Otto-von-Guericke University, Magdeburg.

In both sham and main experiments, observers viewed sequences of 200 recurring and non-recurring objects (see below and Fig. 1A) and attempted to classify each object as 'familiar' or 'novel' (by pressing the appropriate button). Over the course of multiple sessions, observers gradually became familiar with recurring objects and thus became able to distinguish them from non-recurring objects. Objects of the sham experiment were two-dimensional shapes, whereas objects of the main experiment were rotating, three-dimensional shapes (see below and Fig. 1A).

The main experiment extended over three successive weeks, with three sessions on separate days of both the first and third week (no sessions took place in the second week). The experiments of the first and third weeks differed in four aspects: sequence type (structured or unstructured), the set of recurring objects, object color (red or blue), and responding hand (left or right). All aspects were counterbalanced across observers. With either responding hand, the index finger responded 'familiar' and the middle finger responded 'unfamiliar'. Observers were not informed about the difference in sequence structure.

After the three scanning sessions of a week, observers participated in an additional behavioral session to confirm that they had, in fact, become familiar with every recurring object. Specifically, they performed a spatial search task in which they pointed out recurring target objects among non-recurring distractor objects (Kakaei et al., 2021). In addition, observers were offered the opportunity to voice anything they might have noticed about the experiment.

2.2. Experimental paradigm

Complex three-dimensional objects were computer-generated and presented as described previously (Kakaei et al., 2021). A movie can be viewed under this LINK: https://learnmem.cshlp.org/content/suppl/2021/04/09/28.5.148.DC1/Supplemental_Movie_S1.mp4. All objects were highly characteristic and dissimilar from each other (as confirmed by computational means). Objects were presented every 3*s*, with 2.5 *s* viewing and 0.5 *s* transition time (Fig. 1A). Objects were shown from all sides and, after appearing at an arbitrary angle, revolved smoothly for one full turn (period 2.5 *s*, frequency 0.4 *Hz*, angular frequency 144°/*s*) about one of several axes in the frontal plane (–45°, 0°, 45°, clockwise or counter-clockwise). Axes and directions were counterbalanced for each object, and initial viewing angles were chosen randomly (Fig. 1B). All stimuli were generated with MATLAB (The MathWorks, Inc.), presented with the psychophysics toolbox (Brainard, 1997), and viewed in a mirror mounted to the MR head coil (screen resolution 960×720 pixels, frame rate 60*Hz*, subtending approximately $8^\circ \times 6^\circ$ of visual angle, average luminance $50 Cd/m^2$, background luminance $5 Cd/m^2$). Observers responded with the right or left index finger on an MR-safe response box.

Fifteen objects recurred many times during three sessions ('recurring' objects), whereas other objects appeared exactly once ('non-recurring' or 'singular'

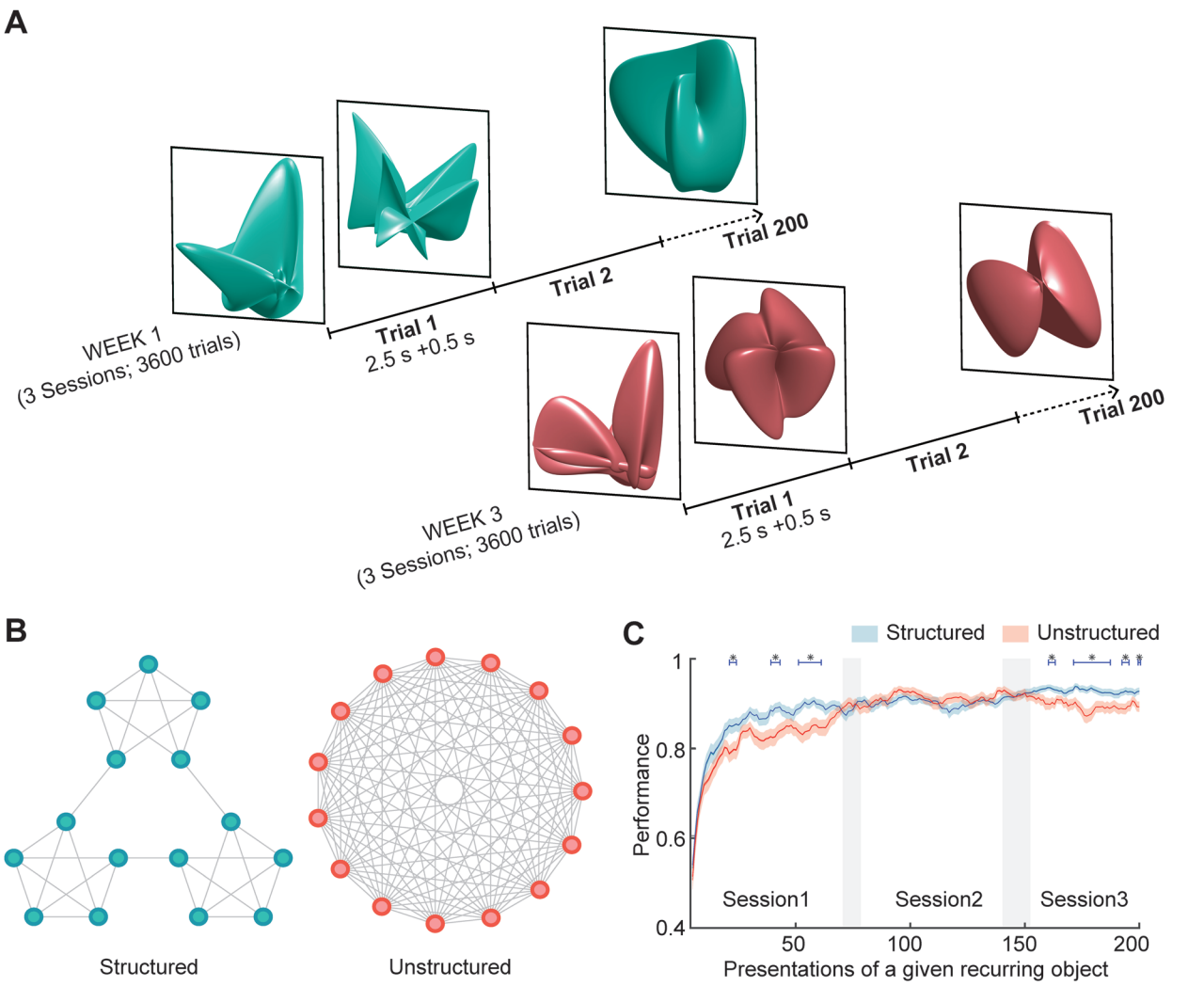

Fig. 1. Experimental paradigm. (A) Observers viewed complex, three-dimensional objects that rotated slowly, presented as sequences with 200 trials (2.5 s presentation time and 0.5 s transition time). During the transition, the previous object vanished to the right while the next object approached from the left (please see https://learnmem.cshlp.org/content/suppl/2021/04/09/28.5.148.DC1/Supplemental_Movie_S1.mp4). Most objects (180 of 200) recurred multiple times within and between sequences ('recurring' objects). The others (20 of 200) were presented only once ('non-recurring' objects). Over 3 days/sessions, observers viewed 18 sequences and attempted to classify each object as either 'familiar' or 'unfamiliar'. (B) Sequences were generated from quasi-random walks on either a sparse and modular graph, or a fully-connected and non-modular graph (nodes represent recurring objects, and links represent possible successions). Sequences from the left graph exhibit clustered sequential dependencies ('structured'). Sequences from the right graph lack such dependencies ('unstructured'). (C) Over three sessions, observers learned to classify recurring and non-recurring objects as 'familiar' and 'unfamiliar', respectively. Performance was slightly better with structured than with unstructured sequences. For the present group of observers, the difference was significant during the first and the last sessions (*$p < 0.05$; FDR corrected). With structured sequences, performance continued to improve slightly from the second to the third session.

objects). As mentioned, observers classified every object as either 'familiar' or 'unfamiliar' by pressing either the left or right button (counterbalanced) during its presentation. Over the course of three sessions, all observers gradually became familiar with the 'recurring objects' (see below). The average time-course of learning, as established by a simplified signal detection analysis, is shown in Figure 1C.

Every session comprised six sequences ('runs'), each lasting 600*s* and presenting 180 'recurring' and 20 'non-recurring' objects (200 objects in total). As there were 15 different recurrent objects, each such object was seen 12 ± 1.9 times during every sequence. Over the three sessions (or 18 sequences), each recurring object appeared at least 190 times each (mean ± S.D: 216 ± 9), whereas non-recurring objects appeared only once. Altogether, there were 3,240 presentations of recurring objects ($3 \times 6 \times 180$) and 360 presentations of non-recurring objects ($3 \times 6 \times 180$).

2.3. Presentation order

To create conditions with and without predictive temporal context ('structured' and 'unstructured'), sequences were generated as quasi-random walks on graphs representing the 15 recurring objects as nodes and possible continuations as edges (Fig. 1B). Each sequence started at a random node and continued with equal probability on any one of the available edges, except that immediate repetition ($X \to X$) and direct returns ($X \to Y \to X$) were not allowed. Although generated randomly, sequences were post-selected to counterbalance the number of appearances of both objects and object pairs (Kakaei et al., 2021). Non-recurring objects were interspersed at random sequence locations.

Structured sequences were generated from the modular graph depicted left in Figure 1B. Note that each object is linked to exactly four other objects (i.e., may be preceded or followed by four other objects). Additionally, links are clustered such as to form three "communities" with five objects each. As a result, the objects from a community tended to follow each other: on average, 9 ± 2 successive objects derived from the same community, so that these "community episodes" lasted 27 ± 6 s on average. Moreover, the same objects tended to repeat at short intervals and the expected repetition latency of 5.5 ± 15 (median and S.D.) was comparatively short.

In structured sequences, the 105 possible pairings of 15 objects could be divided into four groups, as illustrated further below in Figure 6A. There were 27 pairs from the *same* community and *adjacent* on the graph (*SameAdjacent* pairs), 3 pairs from *different* communities and *adjacent* on the graph (DA pairs), as well as 3 pairs from the *same* community and *non-adjacent* on the graph (SN pairs). Finally, 72 pairs were from *different* communities and *non-adjacent* on the graph (DN pairs).

Note that only *SA* pairs and *DA* pairs actually occurred in structured sequences, in the sense that one member occasionally followed the other. Counterbalancing ensured that all objects and all possible object pairs occurred comparably often in presentation sequences (probability approximately $1/60$). See Kakaei et al., 2021 for further details about the statistics of presentation sequences.

Unstructured sequences were generated from the graph depicted right in Figure 1B. In this graph, each object was linked to all other objects (i.e., it may be preceded or followed by any one of the other objects), so that no sequential dependencies arose. As a result, same objects rarely repeated at short intervals and the expected repetition latency of 10.5 ± 11 (median and S.D.) was comparatively long.

2.4. MRI acquisition

All magnetic-resonance images were acquired on a 3T Siemens Prisma scanner with a 64-channel head coil. Structural images were T1-weighted sequences (**MPRAGE** TR = 2,500 ms, TE = 2.82 ms, TI = 1,100 ms, 7° flip angle, isotropic resolution $1 \times 1 \times 1$ *mm*, and matrix size of $256 \times 256 \times 192$). Functional images were T2*-weighted sequences (TR = 1,000 ms, TE = 30 ms, 65° flip angle, resolution of $3 \times 3 \times 3.6$ *mm*, and matrix size of $72 \times 72 \times 36$). Field maps were obtained by gradient dual-echo sequences (TR = 720 ms, TE1 = 4.92 ms, TE2 = 7.38 ms, resolution of $1.594 \times 1.594 \times 2$ *mm*, and matrix size of $138 \times 138 \times 72$).

2.5. fMRI pre-processing

Our approach to fMRI analysis was influenced by recent advances in comparing uni- and multivariate responses of corresponding voxels between different observers (Kumar et al., 2022; Nastase et al., 2019). The *local* correlation structure of voxel response is surprisingly similar in different observers and provides a solid basis for functional parcellation (Dornas & Braun, 2018). Such a parcellation obviates 'searchlight' strategies and can define high-dimensional multivariate activity in corresponding 'parcels' for different observers.

The fMRI pre-processing procedure was similar to that published previously (Dornas & Braun, 2018). Brain tissues were extracted and segmented using BET (Smith, 2002) and FAST (Zhang et al., 2001). Fieldmap correction, head motion correction, spatial smoothing, high-pass temporal filtering, and registration to structural and standard images were performed with the MELODIC package of FSL (Beckmann & Smith, 2004). Field map correction and registration to structural image were carried out using Boundary-Based Registration (BBR; Greve & Fischl, 2009). MELODIC uses MCFLIRT (Jenkinson et al., 2002) to correct for head motion. Spatial smoothing was performed with SUSAN (Smith & Brady, 1997), with full width at half maximum set at FWHM = 5 *mm*. To remove low-frequency artifacts, we applied a high-pass filter of the cut-off frequency $f = 0.01$ Hz, that is, oscillations/events with periods of more than 100 s were removed. To register the structural image to Montreal MNI152 standard space with isotropic 2*mm* voxel size, we used FLIRT (FMRIB's Linear Image Registration Tool; Jenkinson & Smith, 2001; Jenkinson et al., 2002) with 12 degrees of freedom (DOF) and FNIRT (FMRIB's Nonlinear Image Registration Tool) to apply the non-linear registration. To further reduce artifacts arising from head motion, we applied despiking with a threshold of $\lambda = 100$ using BrainWavelet toolbox (Patel et al., 2014). Later, we

regressed out the mean CSF activity as well as 12 DOF translation and rotation factors predicted by a motion correction algorithm (MCFLIRT). Afterward, the time series of each voxel was whitened and detrended. This resulted in a temporal signal-to-noise ratio (average over time-series, divided by standard deviation over time-series) of approximately 200, with a standard deviation of ±30 over voxels and of ±90 over observers.

Finally, the 160,099 voxels of MNI152 space were grouped into 758 functional parcels according to the MD758 atlas (Dornas & Braun, 2018). Each functional parcel is associated with an anatomically labeled region of the AAL atlas (Tzourio-Mazoyer et al., 2002) and comprises approximately 200 voxels or approximately $1.7cm^3$ of gray matter volume (212 ± 70 voxels, range 45 to 462 voxels). Parcels were defined for a small population of observers such as to maximize signal covariance within, and minimize covariance between parcels in the resting state. In contrast to other parcellation schemes, this was based exclusively on the (typically strong) functional correlations within each anatomical region and disregarded the (typically weak) correlations between different anatomical regions. The MD758 parcellation offers superior cluster quality, correlational structure, sparseness, as well as consistency with fiber tracking, compared to other parcellations of similar resolution (Albers et al., 2021; Dornas & Braun, 2018).

2.6. fMRI data analysis

To study the effect of sequence structure on the neural representation of object shape, we extracted the multi-voxel activity pattern at $N_t = 9$ time points following object onset. In a functional parcel with N_{vox} voxels, this response pattern constituted a point (or vector) in an N_{dim}-dimensional space, where $N_{dim} = N_t \cdot N_{vox}$ (Fig. 2A). Our objective was to compare distances between response patterns to the same objects, to different objects in the same community, and to different objects in different communities, in other words, to analyze representational similarity or dissimilarity in terms of the standardized Euclidean (Mahalanobis) distance between responses in a high-dimensional space (RSA; Kriegeskorte & Diedrichsen, 2019). Over all 758 parcels, response dimensionality was $N_{dim} = 1{,}911 \pm 634$ (mean and standard deviation), with a range of 405 to 4,158.

To analyze the response variance that discriminates the 15 recurring objects, we reduced dimensionality with Fisher's Linear Discriminant Analysis (LDA) for multiple classes to identify the (at most) $(\kappa - 1)$-dimensional subspace $\mathbb{S}$ that optimally discriminates $\kappa = 15$ classes of activity patterns (i.e., responses to the 15 recurring patterns). Here, optimality is defined as simultaneously minimizing within-class variance and maximizing between-class variance of activity patterns. This approach corresponded to a 'supervised' principal component analysis and yielded $(\kappa - 1)$ informative dimensions.

To interpret the results, it is important to appreciate the commonality with principal component analysis (PCA). Over all 758 parcels, the first 14 principal components captured $64 \pm 5\%$ to the the total response variance following an object presentation. However, about one-third of this variance was shared between presentations and thus uninformative about the identity of the presented

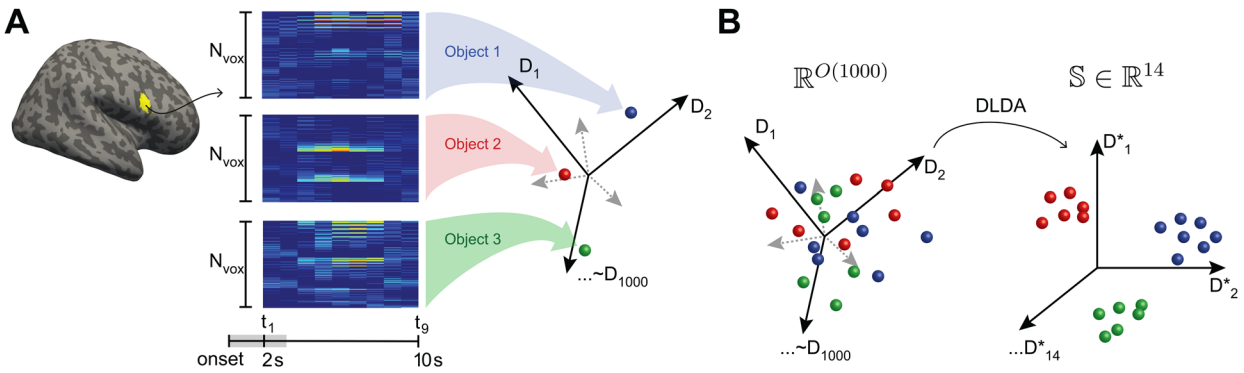

Fig. 2. Direct linear discriminant analysis (DLDA) of multivariate BOLD signals. For each observer and functional parcel, we identified a 14-dimensional space that optimally discriminated the 15 classes of activity patterns associated with *recurring* objects. Typically, this space was contained largely within the space of the 14 principal components ($88 \pm 5\%$ of variance), but excluded shared variance associated with all object presentations. (A) For a given parcel with N_{vox} voxels (yellow: Inf-Front-Oper-R, parcel 146), activity was recorded over 9**s** during and after object presentation (from 2 to 11s after onset). Each such activity pattern corresponded to a point in a $9 \cdot N_{vox}$ -dimensional vector space (right), here represented schematically by spheres (red, green, and blue). Images exemplify average responses to three objects with a color scale. (B) In the optimally discriminative subspace, $\mathbb{S} \in \mathbb{R}^{14}$, Euclidean distance measures the representational similarity of different responses to the same object.

object. The 14-dimensional subspaces $\mathbb{S}$ identified by LDA captured the remaining two-thirds (66 ± 5%) of the PCA variance, which were informative about the objects present. In fact, almost all of the subspace variance (88 ± 5%) overlapped with the space of the 14 leading principal components. Moreover, the subspaces $\mathbb{S}$ tended to distribute variance more uniformly over dimensions (3 ± 3% per dimension) than principal components did (4 ± 6% per dimension).

This commonality between LDA and PCA explained why subspaces $\mathbb{S}$ captured response variance under all conditions (non-recurring objects, non-selective parcels), not just the conditions for which they had been optimized. A numerically tractable procedure for identifying the optimal subspace $\mathbb{S}$ is available in terms of 'direct LDA' or DLDA (Ye et al., 2006; Yu & Yang, 2001). The link github.com/cognitive-biology/DLDA provides a Matlab implementation of DLDA.

The generic nature of subspaces $\mathbb{S}$ permitted us to investigate also the representation of "temporal context" in this way. Specifically, we analyzed the representation of temporal communities with data from structured sequences (8 observers) but performed identical analyses on data from unstructured sequences (8 observers) for comparison. As detailed further below, spurious 'effects' of community structure can be observed due to systematic and/or unsystematic fluctuations of responsiveness over time. To guard against such spurious effects, we removed the effects of temporal proximity and verified that our analyses yielded null results with data from unstructured sequences.

2.6.1. Amplitudes, distances, and temporal correlations

Note that the straightforward approach of decoding community identity (i.e., "community selectivity") would have been confounded by object identity, as any selectivity for "object identity" would necessarily have entailed also some degree of selectivity for "community identity". To sidestep this issue, we devised a somewhat weaker yet independent measure—"community sensitivity"—which compared pairwise distances between responses to objects within and between communities, as detailed further below.

Activity patterns x_{jk} associated with trials k were analyzed in the maximally discriminative subspace $\mathbb{S}$. The average normalized amplitude $a_k = \sqrt{\frac{1}{\kappa-1}\sum_{j=1}^{\kappa-1} x_{jk}^2}$ was $\langle a \rangle = 0.99$, and the average normalized distance $d_{kl} = \sqrt{\frac{1}{\kappa-1}\sum_{j=1}^{\kappa-1}\left(x_{jk}-x_{jl}\right)^2}$ between patterns from trials k and l was $\langle d \rangle = 1.40$. This value corresponds to the normalized distance expected between random patterns, as the average Euclidean distance between two random points, on an n-dimensional hypersphere of unit radius, is

$$d_{ave} = \frac{2^n}{\sqrt{\pi}} \frac{\Gamma^2\left(\frac{n+1}{2}\right)}{\Gamma\left(n+\frac{1}{2}\right)} \quad (1)$$

with $d_{ave} \approx 1.4017$ for $n = 14$.

On successive trials, activity patterns exhibited a weak temporal correlation, with approximately 5% smaller distances at delays below 4 trials and approximately 2% larger distances at delays ranging from 6 to 15 trials. Supplementary Figure S2 shows the delay-dependent distance between response pairs, as well as the pairwise distance within runs, averaged over all parcels and observers. The delay-dependence of response distances to the same objects was comparable in identity-selective and non-selective parcels, although the delay-dependence of distances to different objects was slightly more pronounced in non-selective parcels. In contrast to multivariate response *distances*, we did not observe any effect of delay on multivariate response *amplitudes* (i.e., we observed neither repetition suppression nor repetition facilitation).

To correct for this temporal correlation, we established for each parcel w the average delay-dependent distance $T_w(\Delta i) = \langle d_{w,u,r}(\Delta i) \rangle_{u,r}$ between patterns with relative delay Δi, where the average was taken over subjects u and runs r. The time-course T_w allowed us to subtract the average effect of temporal correlation by computing residual distance $d_{w,u,r}^{corrected}(\Delta i) = d_{w,u,r}(\Delta i) - T_w(\Delta i) + \langle T_w(\Delta i) \rangle_{\Delta i}$, where $\langle T_w(\Delta i) \rangle_{\Delta i}$ is the average value over delays Δi.

2.6.2. Measure of identity-selectivity

Selectivity for object identity was quantified in terms of "classification accuracy", $\alpha^{identity}$, which was defined as the probability that a multivariate response was classified correctly on the basis of distance to class centroids. To test for statistical significance, we relied on the "minimum accuracy" over all observers or data sets (Allefeld et al., 2016). Further details are provided in the companion paper (Kakaei & Braun, 2024).

2.6.3. Geometry of temporal community representations

To assess the representation of community structure, we compared pairwise distances between responses to objects within and between communities for each parcel w. Specifically, we first obtained pairwise

distances d_{ij} and sorted them into two groups: within-community distances with average $D_w^W = \langle d_{ij}(ij \mid i,j \in L) \rangle$ and between-community distances with average $D_w^B = \langle d_{ij}(ij \mid i \in L, j \in K, L \neq K) \rangle$. Then, we established the signed difference $\Delta_w^{BW} = \langle D_w^B \rangle - \langle D_w^W \rangle$, which we termed "community sensitivity", and assessed the statistical significance of Δ_w^{BW} with a two-sample t-test. After correcting for false discovery (Benjamini & Hochberg, 1995), we summarized the results for each parcel in terms of t-statistics t^{BW}.

A similar procedure was used to assess differences between classes of object pairs. Specifically, for every parcel, we established the average pairwise distance D_w (averaged over all pairs and all observers) for different classes of object pairs: same community & adjacent (SA), same community & non-adjacent (SN), different communities & adjacent (DA), and different communities & non-adjacent (DN). The resulting values were termed D_w^{SA}, D_w^{DA}, D_w^{SN}, and D_w^{DN}. The statistical significance was assessed by comparing the observed values to the pairwise distance D_w^{diff}, which contains pairwise distances of all 4 types of object pairs, by a two-sampled t-test. The results were summarized in terms of t-statistics t_w^{SA}, t_w^{DA}, t_w^{SN}, and t_w^{DN}. The behavioral evidence (Kakaei et al., 2021) informed our a-prior hypothesis that SA pairs might be more similar, and NA pairs more dissimilar, than the overall average. A further a-priori hypothesis was that DA pairs might be more dissimilar, as they involve the "linking objects" that mark transitions between different communities.

Note that response distances within and between communities are confounded by temporal proximity because responses *within* communities tend to have shorter relative latencies than responses *between* communities (A. C. Schapiro et al., 2013). To assess the degree to which temporal proximity contaminates the observed community signal, we repeated the analysis of community representations for different ranges of temporal latencies. Specifically, we recalculated the average pairwise distances $D_w^{between}$ and D_w^{within}, and the corresponding t_w^{BW} for object pairs i,j whose relative latencies τ_{ij} where bounded from below by $\tau_{LB} \leq \tau_{ij}$ and from above by sequence termination, with the lower bound ranging over $\tau_{LB} \in \{1,\ldots,30\}$. The t-statistics of response pairs with bounded latencies and their corresponding p-values, corrected for false discovery rate, will be denoted as $t^{BW}(\tau_{LB})$ and $P^{BW}(\tau_{LB})$, respectively.

To assess whether community representations are consistent over different latency ranges, we examined how $t^{BW}(\tau_{LB})$ changes with its lower bound τ_{LB}. Specifically, for each parcel, we defined a consistency measure τ_{sig} as the highest lower bound at which $t^{BW}(\tau_{sig})$ remains significant. We considered a parcel as *'community sensitive'* only if $\tau_{sig} \geq 30$. In other words, a 'community sensitive' parcel exhibited significant between-community separability t^{BW} for all lower bounds $\tau_{LB} \in \{1,\ldots,30\}$. This ruled out the possibility that community sensitivity was a spurious effect of temporal proximity (which was strongest at shorter latencies).

2.6.4. Statistical power

The representational similarity analysis concerning object identity described in Kakaei and Braun (2024) was based on approximately 216 object responses (18 sequences with approximately 12 recurrences of each object) from each of 16 data sets, affording approximately 370,000 representational distances for each of the 105 object pairs. In contrast, the assessment of representational similarity concerning community was based on approximately 120 community episodes (18 sequences with approximately 6 recurrences of each community) from each of 8 data sets, affording approximately 57,000 representational distances for each of the 3 community pairs. Hence, the number of independent pairwise observations about identity was approximately 225 times larger than the number about community. Accordingly, on purely statistical grounds, the sensitivity of our paradigm for detecting community sensitivity is expected to be approximately 15 times *lower* than for detecting identity selectivity.

As a consequence of this statistical disparity, we were unable to establish the temporal development of "community sensitivity" over the 3 days/sessions (see Supplementary Fig. S8). For "object identity", we could demonstrate temporal developments not only over days/sessions but even over individual runs (Kakaei & Braun, 2024).

2.6.5. Dimensional reduction

To visualize the representational geometry of community structure in two dimensions, we calculated a distance matrix $D_{w,u,r}(i,j) = \langle d_{ij} \rangle$ of response distances corrected for temporal proximity within each run r, for every parcel w and observer u. Averaging over the runs produced matrices $D_{w,u}$ of size 15×15 of the average distances between the 15 recurring objects in the discriminative subspace $\mathbb{S}$.

As we did not expect different observers to exhibit comparable activity patterns and distance matrices, we did not wish to average these matrices directly. To sidestep the difficulty, we permuted the object order of the matrix 10^4 times while maintaining graph structure (adjacency and module membership), to first obtain an ensemble average matrix $\overline{D}_{w,u}$ for each observer, and finally the observer average $\overline{\overline{D}}_w$ of ensemble averages.

Using multidimensional scaling (Matlab function *mdscale*), we converted the observer average matrix $\overline{\overline{D}}_w$ to a two-dimensional map of 15 locations approximating these pairwise distances. These maps reveal the average response distance between objects within and between communities, as well as the average distance between 'linking' and other objects. Note that the three-fold rotational symmetry of these maps is owed to the permutation procedure.

3. RESULTS

3.1. Behavior

Observers readily became familiar with recurring objects, as confirmed by the time course of performance in classifying objects as 'familiar' (recurring) or 'novel' (non-recurring), which exceeded 75% correct after one session and approached 90% performance after two further sessions (Fig. 1C). Typically, the classification of a particular object changes from 'novel' to 'familiar' at a particular point in time ("onset of familiarity"). After the experiment, several observers mentioned having invented linguistic labels for each recurring object ('anchor', 'butterfly', 'hedgehog', etc.). Some observers mentioned noticing that objects repeated in close temporal proximity in the 'structured condition'. However, no observer mentioned noticing that the recurring objects formed three distinct "communities".

In the structured condition, familiarity increased slightly faster and "onsets of familiarity" occurred somewhat sooner. Specifically, performance was slightly but significantly higher during much of the first and third sessions, and comparable in the second session (Fig. 1C). Moreover, after an object became familiar, the next object to do so was significantly more likely than chance to be a 'same adjacent' (SA) object and significantly less likely to be a 'different non-adjacent' (DN) object. Specifically, the frequency of successive onsets of familiarity was elevated by 0.15 ($p < 0.05$) for SA pairs, and reduced by –0.15 ($p < 0.05$) for DN pairs, but did not differ significantly for either DA pairs ("linking objects") or SN pairs.

Average reaction times mirrored the performance results in that they were higher before than after the "onset of familiarity" ($p < 0.01$). In the structured condition, reaction times for linking objects (members of DA pairs) and internal objects (all others) did not differ significantly during either the first, second, or third session ($p < 0.01$). Thus, the behavioral effects of sequence structure did not extend to reaction times. This was consistent with the behavioral results reported previously (Kakaei et al., 2021).

3.2. Representation of temporal community structure

To assess the effects of temporal community structure, we analyzed the multivariate BOLD activity of each brain parcel over 9*s* (or 9*TR*), starting with the onset of object presentation. Specifically, we analyzed linear distances between multivariate responses after reducing the dimensionality of the originally $O(1{,}000)$-dimensional responses to the 14 dimensions of an 'optimal subspace' $\mathcal{S}$. We chose this subspace such as to maximize the discriminability of responses to different recurring objects, using Fisher's linear discriminant analysis (LDA). Unlike principal component spaces, the optimal subspaces disregarded variance that was shared between responses to different objects and emphasized variance that distinguished responses to different objects. The dimensionality of the subspace (14) reflected the number of recurring objects (15) and was large enough to capture the major part of the response variance.

Over all 758 parcels, the first 14 principal components captured $64 \pm 5\%$ of the total response variance following an object presentation. However, about one-third of this variance was shared between presentations of different objects and was thus uninformative about the objects. The 14-dimensional subspaces $\mathbb{S}$ identified by LDA captured the remaining two-thirds ($66 \pm 5\%$) of the PCA variance, which were informative about the objects present. Moreover, the subspaces $\mathbb{S}$ tended to distribute variance more uniformly over dimensions ($3 \pm 3\%$ per dimension) than principal components did ($4 \pm 6\%$ per dimension).

In principle, response distances could have reflected temporal community structure in different ways, as illustrated schematically in Figure 3. For example, responses to objects in the same community could be systematically *closer together* than to objects in different communities, indicating greater representational similarity ('positive sensitivity'; Fig. 3A). Alternatively, responses to objects in the same community could be systematically *further apart* than to objects in different communities, indicating less representational similarity ('negative sensitivity'; Fig. 3B). A third possibility would be no systematic relationship between response distance and community membership (Fig. 3C). Optimal subspaces were chosen such as to maximize distances between objects regardless of temporal community and thus favored neither possibility over another. In fact, optimal subspaces were computed in the same way whether or not temporal communities were present (structured and unstructured conditions).

A difficulty in assessing community sensitivity is that it is confounded by the known temporal auto-correlation of multivariate activity (A. C. Schapiro et al., 2013). As

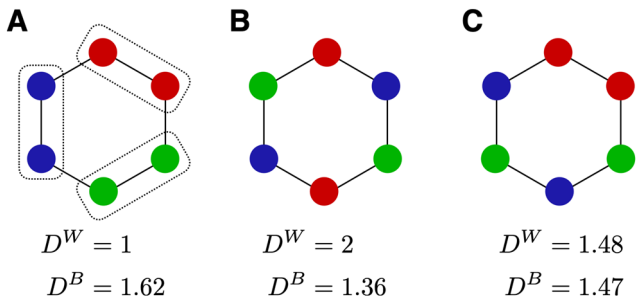

Fig. 3. Possible representations of temporal community structure (highly schematic). Disks represent multivariate responses to 6 objects, and two-dimensional distances represent multivariate distance. Colors represent temporal communities. In each case, the average distance within and between communities is provided (D^W and D^B, respectively). (A) Responses to objects are closer within than between communities (dotted boxes). (B) Responses to objects are further apart within than between communities. (C) Responses to objects are, on average, comparably distant within and between communities.

illustrated in Figure 4A, pairwise distances were computed for all observers, runs, and parcels w, to obtain average pairwise distance D_w^{within} communities, average pairwise distance $D_w^{between}$ between communities, and the average separability $\Delta_w^{BW} = D_w^{between} - D_w^{within}$. For every parcel w, we established "community sensitivity" by assessing whether or not Δ_w^{BW} values differed significantly from zero with t-statistic t^{BW}. Two measures were taken to correct for temporal auto-correlations and to dissociate community sensitivity and temporal proximity (see Section 2.6.1). Firstly, for each raw pairwise distance and its latency, we computed a residual pairwise distance by subtracting the average distance at that latency. Secondly, we compared the t-statistic t^{BW} for subsets of pairwise distances covering different temporal latency ranges ($\tau_{LB} \le \tau \le 30; \tau_{LB} \in \{1,\ldots,30\}$). For all parcels, significance decreased monotonically when lower bound τ_{LB} was raised and shorter latencies were progressively excluded. Thus, the situation was summarized by the *largest* value of τ_{LB} at which t^{BW} statistic was significant, which value was termed τ_{sig}. A high value of τ_{sig} indicated significance over all latency ranges, both including shorter latencies (low values of τ_{LB}) and excluding shorter latencies (high values of τ_{LB}). A low value of τ_{sig} indicated significance only for ranges that included shorter latencies (low values of τ_{LB}).

When raw pairwise distances were used, almost all parcels (613 out of 758 parcels) exhibited significant separability Δ_w^{BW}. When residual pairwise distances were considered, ninety-three parcels retained significant Δ_w^{BW} (left margin of Fig. 4B). In 28 of these 93 parcels, between-community separability was higher ($t^{BW} > 0$) and in the remaining parcels, it was lower ($t^{BW} < 0$).

This disparity between raw and residual distances shows that community structure is confounded by temporal auto-correlation to a considerable degree. This is also evident from strong dependence of t^{BW} on the range of temporal latencies τ_{sig} (Fig. 4B). When only latency ranges including shorter latencies are considered ($\tau_{sig} \le 5$), many more parcels are consistently significant than when ranges excluding shorter latencies are also considered ($\tau_{sig} > 15$). Applying the strictest criterion and considering only parcels with significant Δ_w^{BW} (FDR corrected $p < 0.05$) for all latency bounds $\tau_{LB} \in \{1,\ldots,30\}$ ($\tau_{sig} = 30$), we obtained 27 parcels that we considered *'community sensitive'*. These parcels are listed in Appendix Table A1 and illustrated in Figure 4C to D and in Supplementary Figure S1.

The above analysis yielded interpretable results for strongly-structured presentation sequences, where every object can be objectively assigned to one particular community. When the analysis was repeated for unstructured presentation sequences (by counterfactually assuming a structured sequence and assigning communities accordingly), no systematic results were obtained, as shown in Supplementary Figure S3. Specifically, apparent community sensitivity is observed only when uncorrected distances over low-latency ranges are considered. Correcting for temporal correlations eliminates this spurious sensitivity. The static matrix of average pairwise distances provides an instructive baseline for spurious 'sensitivity' that is entirely due to temporal correlations. Apart from very short latencies, the results from this matrix are comparable to results from unstructured sequences, for both *positively* and *negatively* community-sensitive parcels temporal correlations (Supplementary Fig. S3B, C). Results for structured sequences are dramatically different (both higher and lower), corroborating the validity of our analysis of community sensitivity.

Fourteen community-sensitive parcels with *higher* separability of between-community pairs ($\Delta^{BW} > 0$) were located in bilateral occipital regions and in ventral occipitotemporal regions of the right hemisphere (visual cortex, lateral occipital cortex, fusiform and lingual gyrus, anterior inferior temporal cortex, as well as intraparietal cortex and middle frontal cortex; Fig. 4C; Appendix Table A1). Eleven of these parcels were also identity-selective. Thirteen other parcels exhibited significantly *lower* separability of between-community pairs ($\Delta^{BW} < 0$) and were located in the superior temporal cortex, supramarginal gyrus, insula, operculum, medial frontal cortex, and the frontal pole (Fig. 4C; Appendix Table A1). In this latter group, 12 parcels were not identity-selective.

The respective cortical distributions of the representations of object identity and community membership are compared and illustrated in Figure 5. The criterion for

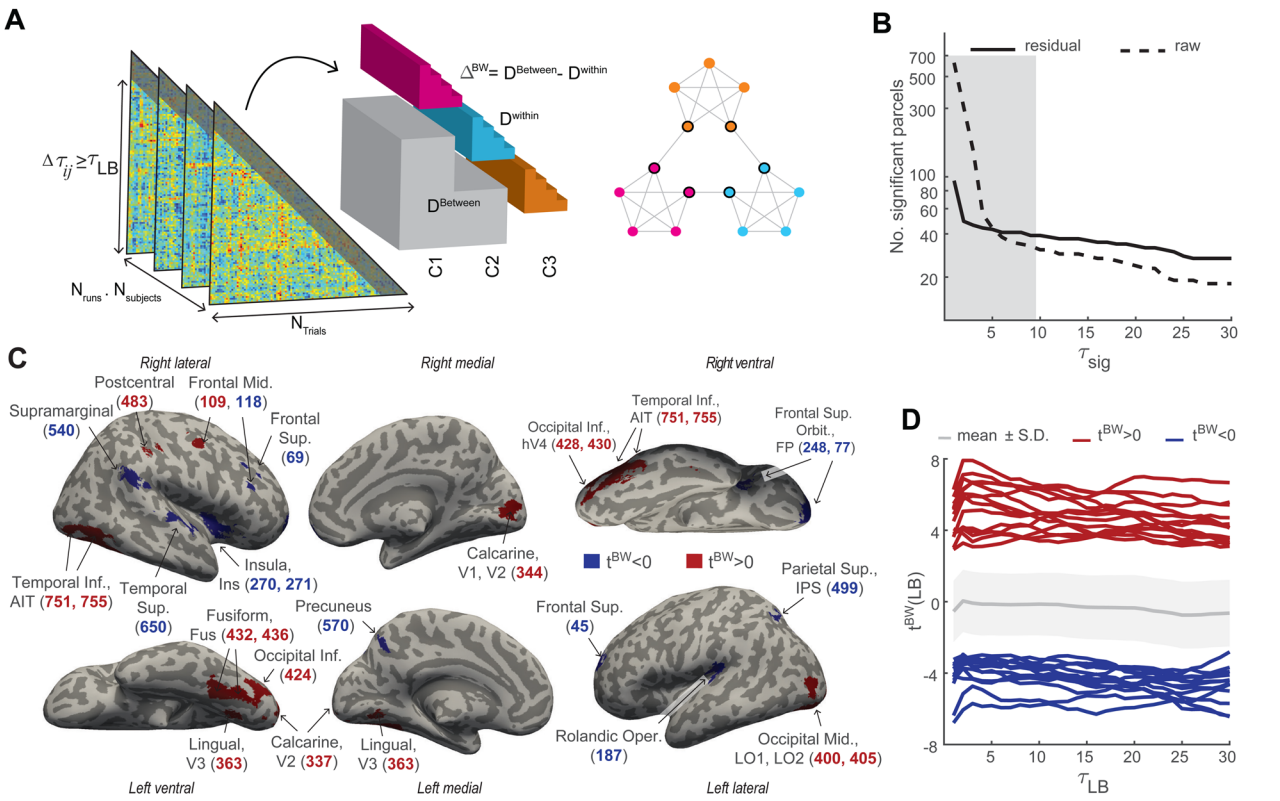

Fig. 4. Distribution of sensitivity to temporal community structure. (A) Pairwise distances between object responses (triangular matrices) were corrected for the average auto-correlation, thresholded by latency τ_{ij} (lower bound τ_{LB}, indicated by shading), and sorted into different subsets—within-community pairs (cyan, magenta, orange) and between-community pairs (grey)—according to object positions on the modular path (right). For the average signed difference Δ^{BW}, statistic t^{BW} was computed. (B) Number of parcels with consistent significance up to τ_{sig} for residual (solid) and raw (dashed) pairwise distances. The average duration of a community visit was 9.4 ± 0.15 (gray shading). (C) Representation of 'temporal communities' by parcels with consistently significant Δ^{BW}. In 14 parcels (red), between-community pairs are significantly more separable ($t^{BW} > 0$, corrected $p < 0.05$) over all latency ranges whereas, in 13 parcels (blue), within-community pairs are more separable ($t^{BW} < 0$) over all ranges. Labeling indicates parcels in visual cortex (V1, V2, V3, hV4), lateral occipital cortex (LO1, LO2), fusiform and lingual gyrus (Fus, Lin), anterior inferiotemporal cortex (AIT), intraparietal sulcus (IPS), superior temporal cortex, supramarginal gyrus, medial frontal cortex, precuneus, insula (Ins), Rolandic operculum, precentral cortex, and frontal pole (FP). (D) Between-community separability t^{BW} for different latency ranges (lower bound t^{BW}), for 'community-sensitive' parcels with positive $t^{BW} > 0$ (red) and negative $t^{BW} < 0$ (blue). The mean and S.D. of separability over all parcels are shown in gray.

community-sensitivity was a significantly positive or negative *t*-score value t^{BW}, whereas the criterion for identity-selectivity was a significantly positive minimum statistic of classification accuracy α_{min} (for details, see Kakaei & Braun, 2024). Figure 5C shows average classification accuracy $\alpha^{identity}$ as well as α_{min}. Coloring indicates whether parcels combined identity-selectivity with *positive* community-sensitivity (11 parcels, orange) or *negative* community-sensitivity (1 parcel, cyan), or whether parcels were either exclusively community-sensitive (3 parcels *positively* in red, 12 parcels *negatively* in blue) or exclusively identity-selective (112 parcels, yellow). Of the 124 identity-selective parcels, 12 parcels (approximately 10%) were additionally community-sensitive. Jointly selective/sensitive parcels were most common in the mid-level visual cortex (ventral occipital cortex, lingual and fusiform gyrus) and somewhat less common in the early visual cortex (V1, V2, V3, hV4). Jointly selective/sensitive parcels were largely absent from high-level visual areas in the parietal and frontal cortex (inferior parietal sulcus, superior parietal lobule, insula, inferior and medial frontal cortex), but were present in the anterior inferior temporal cortex. The one negatively community-sensitive parcel in the intraparietal sulcus appeared to be an exception. In summary, jointly selective/sensitive parcels were present at all levels of the ventral visual pathway.

Note that the comparison of community- and identity-selectivity was skewed by disparate statistical power. The assessment of community sensitivity was based on approximately 225 times fewer observed response distances than the assessment of identity-selectivity (see Section 2), so statistical sensitivity was expected to be

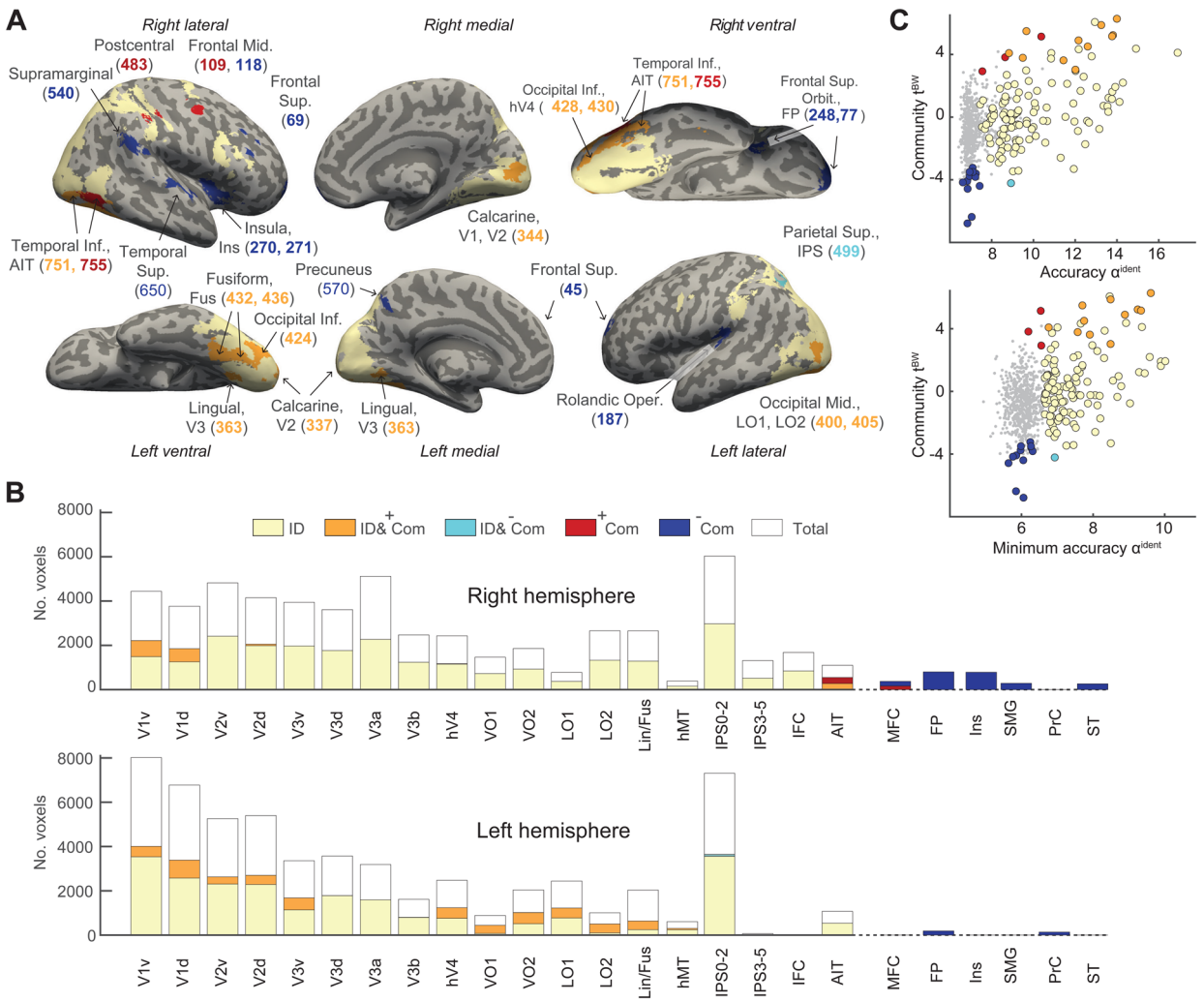

Fig. 5. Comparison of selectivity for object identity and sensitivity to community structure. (A) Anatomical distribution of 142 parcels that are identity-selective, community-sensitive, or both. 11 parcels (orange) are both identity-selective and *positively* community-sensitive, while 1 parcel (cyan) combines identity-selectivity with *negative* community-sensitivity. 15 parcels are exclusively community-sensitive, 3 parcels *positively* (red) and 12 parcels *negatively* (blue). The remaining 112 parcels (yellow) are exclusively identity-selective. (B) Share of identity and community representation in 142 parcels with significant representation, assigned to 29 topographical regions, as defined by L. Wang et al. (2015). In the right hemisphere, two parcels (428 and 430) are missing because they could not be assigned to any topographical regions. Coloring corresponds to (A) and indicates the fraction of voxels from parcels with different selectivity. Visual cortex (V1-hV4), ventral occipital cortex (VO), lateral occipital cortex (LO), lingual and fusiform gyri (LIN/FS), medial temporal areas (MST, hMT), intraparietal sulcus (IPS), superior parietal lobule (SPL), anterior inferior temporal cortex (AIT), insula and supramarginal gyrus (INS/SMG), inferior frontal cortex (IFC), medial frontal cortex (MFC), and frontal pole (FP). (C) Quantitative comparison of selectivity for identity and sensitivity for community over all parcels. Identity-selectivity is quantified either by average classification accuracy $\alpha^{identity}$ (top) or by the minimum statistic of classification accuracy (bottom). Community-sensitivity is measured by positive or negative values of t^{BW}. Significantly sensitive parcels are represented by colored disks, and non-sensitive parcels by grey dots. Coloring corresponds to (A).

approximately 15 times lower. Accordingly, if community-sensitivity was detected in only a fraction of identity-selective parcels, this could, in part, have been due to this disparity in statistical power.

Nominally non-identity-selective parcels with *positive* community-sensitive were located in the anterior inferior temporal cortex and in the medial frontal cortex. As seen in the top panel of Figure 5C, the average classification accuracy $\alpha^{identity}$ of these 3 parcels was comparable to other identity-selective parcels. However, these parcels just missed the minimum statistics criterion for significance, as seen in the bottom panel. It seems possible that community-sensitivity degraded identity-selectivity in these parcels, in the sense that reduced response distances within a community might also have reduced distances between the different objects of this community.

Non-identity-selective parcels with *negative* community-sensitivity were located in the insula, the medial frontal cortex, and at the frontal pole. These 11 parcels exhibited no trace of identity-selectivity in terms of either the observer average or the minimum statistics. Negative sensitivity implies that responses to objects from different communities were more similar than responses to objects from the same community. As discussed below, it seems possible that the responses in these areas placed particular emphasis on 'linking objects', thereby highlighting the 'novelty' or 'surprise' associated with the transition to another community and the appearance of unexpected objects.

3.3. Representation of object pairs

Structured presentation sequences consist of different types of object pairs, such as adjacent and non-adjacent pairs, or 'linking' pairs (between different communities) and 'internal' pairs (within the same community). Thus, it was natural to wonder whether different types of object pairs might have contributed differentially to our average measure, t^{BW}, for "community sensitivity"?

To address this question, we compared the statistical significance of the signed difference in response distances between and within communities for all object pairs, t^{BW}, and for specific types of object pairs: non-adjacent objects in different communities (DN), non-adjacent objects in the same community (SN), adjacent objects in different communities (DA), and adjacent objects in the same community (SA). The results are shown in Figure 6. The separability measure t^{SA} was negatively correlated with t^{BW} ($\rho = -0.91$, $p < 0.01$), whereas the measure t^{DN} was positively correlated with t^{BW} ($\rho = 0.93$, $p \ll 0.01$). The separability measures t^{DA} and t^{SN} were also negative correlated with t^{BW}, though much less strongly ($\rho = -0.15$, $p < 0.01$ and $\rho = -0.19$, $p < 0.01$; respectively). These results were robust and held for all lower temporal bounds $\tau_{LB} \leq 30$, except for the correlation between t^{BW} and t^{DA}, which held only for $\tau_{LB} \leq 28$.

These results show that the representation of community structure (indexed by t^{BW}) includes a reduced separation of *SA* pairs (indexed by t^{SA}), as well as an increased separation of *DN* pairs (indexed by t^{DN}). Recall that *SA* (and *DA*) pairs occur in presentation sequences (with probability 1/60), whereas *SN* (and *DN*) pairs never occur. The selective modulation of representational distance for one of the two adjacent (and therefore occurring) pairs appears to be a correlate of temporal community structure. The same can be said for the selective modulation of representational distance for one of the two non-adjacent (and therefore non-occurring) pairs. Furthermore, the correlation between community representation and separation of *SA* and *DN* pairs is evident not only in the few parcels meeting the statistical threshold for community sensitivity (red and blue dots in Fig. 6), but also in all other parcels as well (grey dots in Fig. 6). Thus, reduced separation of *SA* pairs and increased separation of *DN* pairs appear to be a general feature of the cortical representation of community structure.

The results described above depend critically on the correction for temporal correlations (Supplementary Fig. S4). Without this correction, the t^{BW} measure for between-community separation is dominated by the influence of short-latency pairs. When shorter latencies are excluded and $\tau_{LB} \geq 5$, the correction ceases to make a difference. This underlines again that correcting for average temporal correlations is key to establishing representations of community structure.

3.4. Representational space

A previous study with structured sequences (A. C. Schapiro et al., 2013) reported that within-community

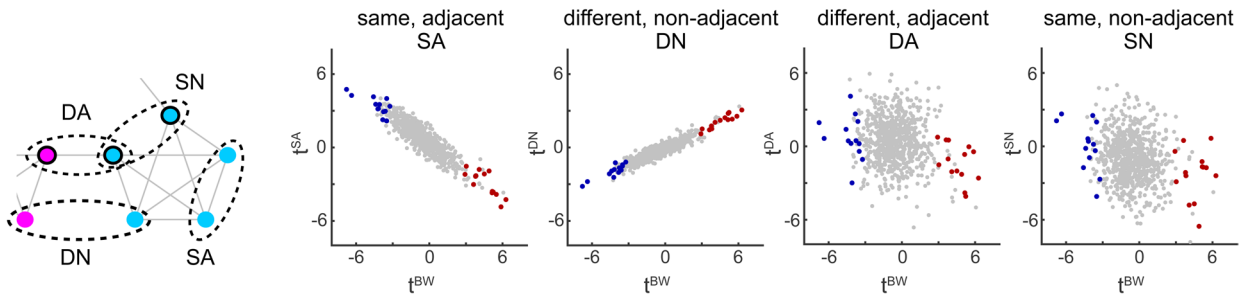

Fig. 6. Neural representation of different types of object pairs. Pairs of recurring objects may be in the same (*S*) or different (*D*) communities, and may occupy adjacent (*A*) or non-adjacent (*N*) positions on the path. Differential separability of between- and within-community pairs (measured by t-score value t^{BW}) is compared to separability t^{SA} of adjacent objects in the same community (left), t^{DN} of non-adjacent objects in different communities (middle left), t^{DA} of adjacent objects in different communities (middle right), and t^{SN} of non-adjacent objects in the same community (right), for all 758 parcels. Community-sensitive parcels are shown in red or blue (as in Fig. 4C).

distances are typically smaller than between-community distances and illustrated this finding with multidimensional scaling. We sought to replicate this by visualizing the relative proximity of different types of object pairs. To obtain interpretable results, we employed a permutation procedure that allowed us to average proximity matrices over observers (see Section 2 for details). The resulting arrangements exhibited a three-fold rotational symmetry that was owed to this permutation procedure and therefore was artificial.

For the 14 *positively* community-sensitive parcels, the relative proximity of different object pairs is illustrated in Figure 7. In all cases except one, objects were clustered by community (i.e., spaced more closely within than between communities), with Temporal-Inf-R-10 providing the most extreme example. Additionally, 'linking' objects tended to be positioned differently than internal objects, in all but two cases closer to each other (and to the center) (Calcarine-L-9, Calcarine-R-5, Lingual-L-1, Occiptal-Mid-L-4, Occiptial-Mid-L-9, Occipital-Inf-L-2, Fusiform-L-2, Fusiform-L-6, Postcentral-R-11, Temporal-Inf-R-10). Exceptions were Frontal-Mid-R-7, where only internal objects clustered by community, and Occipital-Inf-R2/4, where linking objects were more distant from each other. As these illustrations show only relative distances, Supplementary Figure S6A provides absolute response distances in terms of the average and standard error over parcels, separately for internal objects and linking objects, as well as within and between communities. Response distances of internal objects within the same community correspond to the grand average over all object pairs, whereas distances between different communities were significantly larger. Additionally, distances between linking and internal (or linking) objects within the same community were significantly smaller. Thus, both clustering by community and relative proximity of linking objects was statistically significant. On average, this corresponded to the possibility shown schematically in Figure 3A.

Results for the 13 *negatively* community-sensitive parcels are shown in Figure 8. The clustering of internal objects (Frontal-Sup-L-12, Frontal-Sup-Orb-R-3, Frontal-Mid-R-16, Frontal-Med-Orb-R-3, Parietal-Sup-L-8, and Temporal-Sup-R-6) was variable but, when averaged over parcels, internal objects were more distant within than between communities (Supplementary Fig. S6A). Specifically, within the same community, response distances of internal objects to other internal objects (or linking objects) were significantly larger than the grand average over all object pairs, whereas between communities response distances were smaller. On average, this corresponded to the possibility shown schematically in Figure 3B. In six parcels, all linking objects were distant from each other (and from

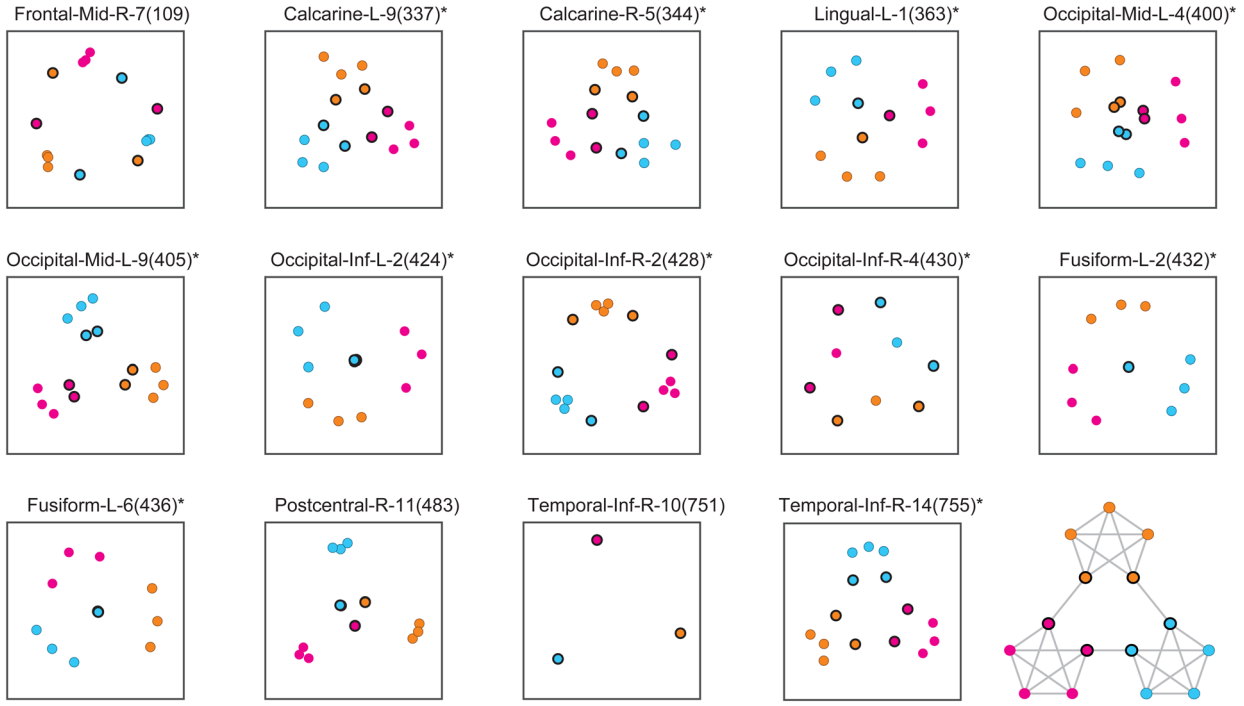

Fig. 7. Representation of temporal community structure in positively sensitive parcels. Multidimensional reduction of the pairwise distance matrix averaged over path permutations and over observers. Communities are distinguished by color and linking objects by a black outline, as indicated by the path diagram (inset). Fourteen parcels exhibited higher separability between communities than within communities ($t^{BW} > 0$). Identity-selective parcels are marked with ⋆.

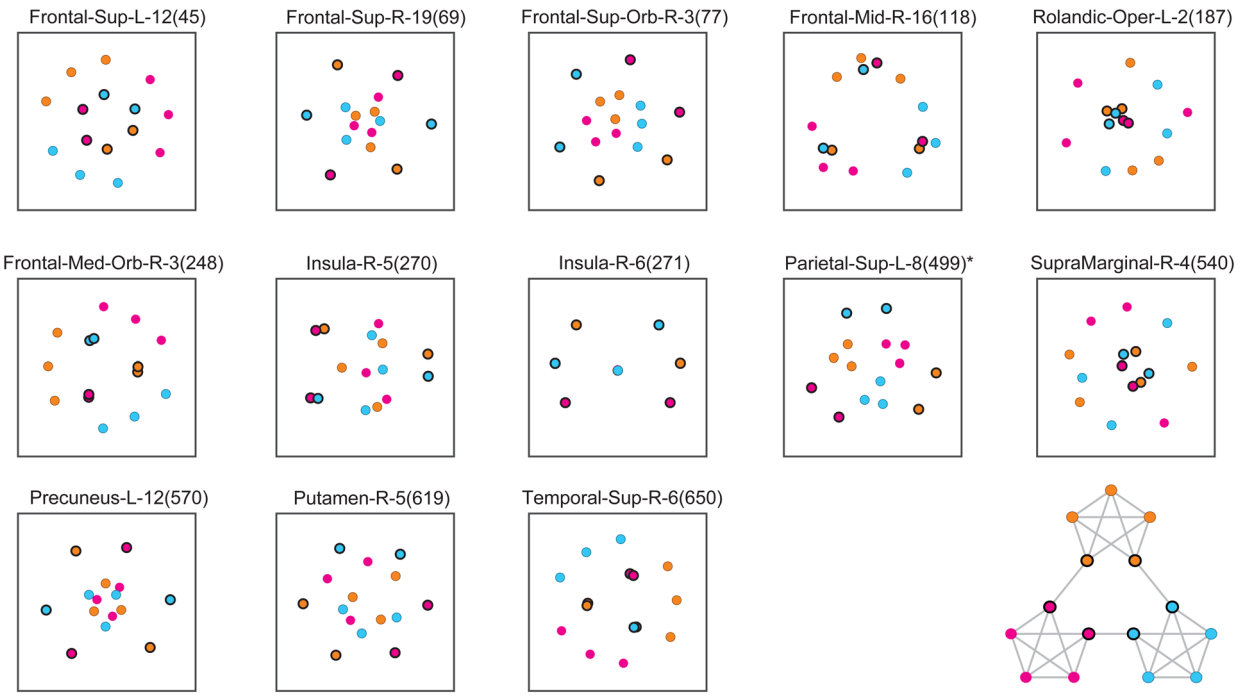

Fig. 8. Representation of temporal community structure in negatively sensitive parcels. Multidimensional reduction of the pair-wise distance matrix averaged over path permutations and observers. Communities are distinguished by color and linking objects by a black outline, as indicated by the path diagram (inset). Thirteen parcels exhibited lower separability between communities than within communities ($t^{BW} < 0$). Identity-selective parcels are marked with $\star$.

the center), suggesting that the representation in these parcels individuated different transitions between communities (Frontal-Sup-R-19, Frontal-Sup-Orb-R-3, Insula-R-6, Parietal-Sup-L-8, Precuneus-L-12, Putamen-R-5). However, in seven other parcels, linking objects were separated less well than internal objects (Frontal-Sup-L-12, Frontal-Mid-R-16, Rolandic-Oper-L-2, Frontal-Med-Orb-R-3, Insula-R-5, SupraMarginal-R-4, Temporal-Sup-R-6), suggesting that the representation conflated different transitions between communities.

It is instructive to also compare parcels that were not classified as either identity-selective or community-sensitive. Results for 15 randomly chosen 'non-selective' parcels are shown in Supplementary Figure S5. Perhaps not surprisingly, the results were quite heterogeneous and few differences reached statistical significance when averaged over parcels (Supplementary Fig. S6A). However, in several individual parcels, clustering by communities and/or prominent representation of linking objects was evident.

As a final control, we analyzed average pairwise distances in the responses obtained with unstructured sequences. Here, we failed to observe significant deviations from the grand average distance, either for internal and linking objects or within and between communities (Supplementary Fig. S6B). This corroborates that the results obtained with structured sequences were due to the presence of temporal structure and/or temporal communities.

4. DISCUSSION

We investigated incidental and automatic learning of regularities and dependencies without explicit behavioral task (Aslin, 2017; Fiser & Lengyel, 2022; Perruchet, 2019; Perruchet & Pacton, 2006; Saffran & Kirkham, 2018; A. Schapiro & Turk-Browne, 2015). Our aim was to compare the cortical basis of concurrent learning of statistical structure with two timescales, namely, explicit learning to recognize complex objects presented for ~3 s (Cox et al., 2005; Tian & Grill-Spector, 2015; Wallis & Bülthoff, 2001) and implicit learning of task-irrelevant contingencies in the sequence of object presentations ("temporal communities" lasting ~30 s) (Fiser & Aslin, 2002; Kakaei et al., 2021; Miyashita, 1988; Säringer et al., 2022; Turk-Browne et al., 2005, 2009). Our results show that cortical representations of both object identity and temporal community structure coexist in large parts of the ventral occipitotemporal cortex.

Previous studies have localized view-invariant object representations in inferior temporal cortex (IT) and lateral occipital complex (LOC) (Grill-Spector et al., 2001; Sáry et al., 1993; Van Meel & Op de Beeck, 2020). Single-unit responses in IT of non-human primates reflect the intrinsic

contingencies of an invariant representation and correlate closely with recognition performance (Jia et al., 2021; Li & DiCarlo, 2008, 2010, 2012). Human fMRI show differential adaptation in IT for congruent and incongruent shapes (Van Meel & Op de Beeck, 2018). In addition, evidence for view-invariant representations has been reported in primary-visual cortex (Eger et al., 2008), at more anterior sites such as fusiform gyrus, and ventral occipito-temporal cortex (Brants et al., 2016; Visconti di Oleggio Castello et al., 2021), as well as in several areas of the dorsal pathway (Freud et al., 2017; Jeong & Xu, 2016; Konen & Kastner, 2008; Poirier et al., 2006; Visconti di Oleggio Castello et al., 2021).

Our results confirm and extend these previous findings on cortical regions with view-invariant object representations, as described in our companion study (Kakaei & Braun, 2024). In brief, we established cross-validated multivariate representations of object identity for smallish 'functional parcels' (~ $1.7cm^3$ cortex volume) defined previously by a functional parcellation (MD758; Dornas & Braun, 2018). Parcels in which significant identity information was prevalent (Allefeld et al., 2016) were located in both the ventral and dorsal visual pathways, beginning with early visual areas (V1-hV4), extending to more anterior parts of ventral occipitotemporal cortex into anterior inferior temporal cortex, as well as to anterior inferior frontal cortex (Kakaei & Braun, 2024).

Our motivation to compare cortical representations of object shape and temporal object sequence derived from classical studies of object recognition in non-human primates (Erickson & Desimone, 1999; Miyashita, 1988). These studies had shown that the responsiveness of single neurons in the inferiotemporal cortex developed selectivity not only for the identity but also for the presentation order of objects, provided that animals had consistently viewed these visual objects in the same sequential order. As this order was irrelevant to the animal's task, the development of a neural representation for sequential order constituted a prototypical instance of incidental or implicit learning.

In extensive subsequent work with "paired-associate tasks", the sequential order of objects was made task-relevant so that learning of temporal associations became explicit. Over the course of training, the prevalence of pair-encoding neurons was found to increase in anterior parts of inferiotemporal cortex IT (Hirabayashi & Miyashita, 2014; Messinger et al., 2001; Naya et al., 2001, 2003). Additionally, neurons in IT were found to encode "object-general semantic value" in the sense of identifying whether a particular object was "familiar" or "novel" (Tamura et al., 2017). Here, we investigated the possibility that such "object-general" information could extend to membership in a "temporal community" of objects.

Previous behavioral studies have shown that humans can implicitly learn spatiotemporal associations between objects and use these regularities to enhance their cognitive performance. Observers can automatically capture spatial (Fiser & Aslin, 2001) and temporal (Fiser & Aslin, 2002; Turk-Browne et al., 2008) regularities as both joint and conditional probabilities of stimuli co-occurrence. This surpasses simple object-object associations and extends to higher-order association probabilities, over multiple objects. Even when the conditional probability between object pairs is uniform and thus uninformative of the underlying association between objects, humans are sensitive to higher-order regularities (higher-moments of conditional probability distribution) (Kahn et al., 2018; Kakaei et al., 2021; Karuza, Kahn, et al., 2017; A. C. Schapiro et al., 2013). This capability for incidental learning of complex regularities can facilitate performance in various domains, including language (e.g., Saffran et al., 1996), motor (e.g., Hunt & Aslin, 2001), spatial attention (e.g., Chun & Jiang, 1998; Jiang & Wagner, 2004), and object recognition learning (e.g., Kakaei et al., 2021).

The literature on implicit or explicit learning of temporal associations shows that both domain-specific and domain-general brain regions can be involved (for reviews, see Batterink et al., 2019; Fiser & Lengyel, 2022). Neural correlates of statistical learning are evident in early domain-specific sensory areas where spatiotemporal regularities are first extracted, to mid-level sensory areas where these representations are supposedly integrated. In the visual domain, spatiotemporal regularities emerge in lateral and ventral occipito-temporal and parieto-occipital regions in humans (Henin et al., 2021; Karuza, Emberson, et al., 2017; Rosenthal et al., 2016; Turk-Browne et al., 2009) and are observed in inferiotemporal and anterior inferiotemporal regions in non-human primates (Kaposvari et al., 2018; Meyer et al., 2014; Miyashita, 1988; Sakai & Miyashita, 1991). More abstract and generalized representations of temporal associations have been reported in more downstream, domain-general areas such as medial temporal lobe, striatum and frontal regions. Moreover, the majority of studies point to an essential role of the medial temporal lobe (MTL), particularly the hippocampus, in statistical learning (Hindy et al., 2016; Hsieh et al., 2014; A. Schapiro & Turk-Browne, 2015; A. C. Schapiro et al., 2012, 2013, 2016; Schendan et al., 2003; Turk-Browne et al., 2009, 2010). This is particularly true when sequences are repeated and when ordinal knowledge is of particular interest to observers (for reviews, see Davachi & DuBrow, 2015; Eichenbaum et al., 2016). MTL seems to be engaged in statistical learning that occurs early in the learning process but seems to disengage as learning progresses, particularly after consolidation. Concurrently, the encoding of statistical knowledge

seems to transfer from MTL to the striatal-frontal network (Batterink et al., 2019; Durrant et al., 2013). Higher cortical regions in insular cortex and prefrontal cortex (PFC), including inferior frontal gyrus (IFG) and medial prefrontal cortex (mPFC), also show sensitivity to statistical regularities, particularly when the complexity increases (Giorgio et al., 2018; Henin et al., 2021; Karlaftis et al., 2019; Kourtzi & Welchman, 2019; A. C. Schapiro et al., 2013; R. Wang et al., 2017).

Here, we adapted the paradigm of A. C. Schapiro et al. (2013) and used object sequences with higher-order "temporal community structure". In such sequences, pair probabilities are uniform in that every object is succeeded by one of four other objects with equal probability. This avoids the novelty/surprise effects that would arise if some object transitions were more common or rare than others. We term sequences with temporal communities "strongly structured", to distinguish them from "unstructured" pseudo-random sequences where every object can be succeeded by any other object (Kakaei et al., 2021).

We studied cortical representations with a "representational similarity analysis" (RSA) approach, which relies on comparing pairwise distances between multivariate BOLD responses to different objects. A difficulty with this approach is that multivariate BOLD patterns are known to be significantly autocorrelated over 10 s of seconds (Alink et al., 2015; Henriksson et al., 2015), in part due to hemodynamic effects (Friston et al., 1994; Zarahn et al., 1997). Accordingly, it was essential to distinguish between response similarity due to genuine "temporal community" effects and response similarity due to mere temporal proximity (i.e., systematically shorter latencies between objects in the same community) (Cai et al., 2019; Gilron et al., 2016).

We took two measures to control for this confound and to distinguish between community and latency effects. First, we computed and analyzed 'residual distances' by subtracting from each observed distance at a certain latency the *average* distance at that latency (see Section 2.6.1). Second, we assessed consistency by analyzing and comparing distances in different latency ranges, for example including or excluding short latencies. These measures turned out to be essential, as nearly the entire brain would have spuriously appeared to be 'community-sensitivity' without them. They also proved effective, as they revealed 'community-sensitivity' only in multivariate BOLD responses to "strongly-structured" sequences and not in responses to "unstructured" sequences. Accordingly, we are confident that these measures identify genuine cortical representations of "temporal community".

Our analysis of multivariate BOLD responses in 758 'functional parcels' revealed two functionally and anatomically distinct kinds of 'community-sensitivity' (see also Fig. 3). The first kind—termed *positively-sensitive*—showed *greater* similarity of responses within communities than between communities and was observed mostly in domain-specific, visual brain regions. The second kind—termed *negatively-sensitive*—exhibited *lesser* similarity of responses within communities and was observed mostly in domain-general areas. We now discuss these two groups in more detail.

Positively community-sensitive parcels—where response distances were *smaller* for objects within than between communities—were located almost exclusively in ventral occipitotemporal cortex, with seven parcels in the left hemisphere (Calcarine-337, Occipital-Inf-424, Occipital-Mid-400 and -405, Fusiform-432 and -436, Lingual-363) and five parcels in the right hemisphere (Calcarine-344, Occipital-Inf-428 and -430, Temporal-Inf-751 and -755). Thus, community-sensitive parcels spanned the range of ventral occipitotemporal cortex that also contained parcels selective for object identity. Almost all positively community-sensitive parcels also exhibited significant identity-selectivity. Although the pattern of relative response distances was somewhat heterogeneous (Fig. 7), some significant trends emerged: response distances between 'internal' objects were above average between communities, whereas distances between 'linking' objects were below average both within and between communities (Supplementary Fig. S6).

While positively community-sensitive parcels comprised only a small fraction of identity-selective parcels (11 of 124 parcels), this disparity may exaggerate the true situation. As our paradigm was considerably less sensitive for community than for identity, a number of 'false negatives' was only to be expected. If the respective statistical sensitivities had been comparable, the overlap between the two groups might well have been larger.

These results are consistent with earlier findings that early and mid-level visual areas are sensitive to temporal regularities and can flexibly alter their activity pattern to represent the temporal context (Henin et al., 2021; Karuza, Emberson, et al., 2017; Rosenthal et al., 2016; Turk-Browne et al., 2009). These are also consistent with the classical observation that representations of temporal association develop conjointly with representations of object identity (Erickson & Desimone, 1999; Miyashita, 1988).

In contrast to numerous earlier studies (Hindy et al., 2016; Hsieh et al., 2014; A. Schapiro & Turk-Browne, 2015; A. C. Schapiro et al., 2012, 2013, 2016; Schendan et al., 2003; Turk-Browne et al., 2009, 2010), we failed to observe positive community-sensitivity in the medial temporal lobe (MTL). We do not consider this a contradiction, as our analysis did not focus on MTL and our parcellation included only six parcels in this region

(2 × Hippocampus, 2 × Perirhinal, 2 × Amygdala). Moreover, MTL is thought to engage early in the learning process and the memory engram is thought be transferred to the striatum after consolidation. As our observations spanned multiple days, memory consolidation could have occurred already after the first session, which could also have explained our failure to observe any community-sensitivity in the MTL. Interestingly, we did observe such sensitivity in one parcel of the putamen.

Negatively community-sensitive parcels were located in domain-general cortex, including the temporal cortex (Temporal-Sup-R-650), parietal cortex (Parietal-Sup-L-499, Supramarginal-R-540, Precuneus-L570), superior frontal cortex (Sup-Frontal-L-45 and -R-69, Sup-Frontal-Orbit-77, Sup-Frontal-Med-R-248), and middle and inferior frontal cortex (Mid-Frontal-R-118, Insula-R-270 and -271, and Rolandic-Oper-L-187). Apart from Parietal-Sup-L499, none of these parcels exhibited a significant representation of object identity, further strengthening the dissociation between *negative* and *positive* community representations.

These findings are consistent with previous reports that implicit learning paradigms can engage parieto-frontal, fronto-striatal, and/or ventral attention networks (Batterink et al., 2019). More generally, prefrontal cortex (PFC) is thought to reflect higher-order statistics of event (Henin et al., 2021) and decision strategies adopted by observers (Giorgio et al., 2018; Karlaftis et al., 2019; Kourtzi & Welchman, 2019; R. Wang et al., 2017). Orbitofrontal cortex (OFC) is thought to be engaged when more abstract representations or 'cognitive maps' are required (Behrens et al., 2018; Christophel et al., 2017; Knudsen & Wallis, 2022; Rusu & Pennartz, 2020; Schuck et al., 2016; Wilson et al., 2014). Insula and inferior frontal gyrus are thought to be engaged by working memory tasks, especially under conditions of high load (Rottschy et al., 2012), and to contribute to goal-directed behavior by interacting with the medial temporal lobe hippocampus (Rusu & Pennartz, 2020). Moreover, when objects are viewed in temporally structured sequences, responses in insula and inferior frontal gyrus are suppressed for expected objects (Ferrari et al., 2022). Interestingly, this 'expectation suppression' arises earlier than in the occipitotemporal visual areas (see also Weilnhammer et al., 2021).

The *negative* community-sensitivity observed both here and in previous studies (A. C. Schapiro et al., 2013) is consistent with "context-specific maps" that individuate objects in a given community, without necessarily identifying either the community or objects in other communities. When the context changes, such a map could be reused to individuate objects in the new community. This would be similar to the invariant response patterns in different environments exhibited by grid-cells (Constantinescu et al., 2016; Doeller et al., 2010; Fyhn et al., 2007).

In summary, our results demonstrate incidental learning of temporal associations at all levels of the ventral visual pathway—from the primary visual cortex to the anterior inferior temporal cortex—at the time-scales of both object presentations (seconds) and of temporal contingencies in the object sequence (tens of seconds). This functional overlap suggests that the visual hierarchy develops *convergent representations* (Grill-Spector & Weiner, 2014) that integrate information from a range of time-scales. It seems likely that such convergent representations contribute to context-dependent enhancement of recognition performance. Our findings confirm the classical observation of a conjoint development of representations of object identity and temporal association (Erickson & Desimone, 1999; Miyashita, 1988).

In the domain-general cortex—superior temporal, parietal, frontal, and insular—representations of higher-order temporal context were also evident, but without any stable representations of object identity. Particularly the 'linking objects' that separated different temporal communities in structured presentation sequences tended to be represented distinctly. Thus, our finding suggests that both the ventral occipitotemporal cortex and/or domain-general cortex could be in a position to contribute to "structural learning" (Kemp & Tenenbaum, 2008; Tenenbaum et al., 2011) and the development of causal insight and understanding (Lake et al., 2017; Shafto et al., 2011).

DATA AND CODE AVAILABILITY

Direct linear discriminant analysis and prevalence inference is available on github.com/cognitive-biology/DLDA. MR data will be made available upon request.

AUTHORS CONTRIBUTION

Ehsan Kakaei: Conceptualization, data curation, formal analysis, visualization, and writing of original draft. Jochen Braun: Conceptualization, linear algebra, formal analysis, supervision, and reviewing & editing.

DECLARATION OF COMPETING INTEREST

The authors are not aware of any competing interest.

ACKNOWLEDGMENTS

We thank Claus Tempelmann, Martin Kanowski, and Denise Scheermann at the Magnetic Resonance Imaging Laboratory of the Department of Neurology of

Otto-von-Guericke University, Magdeburg. We are grateful to Oliver Speck for providing essential support and balanced perspective. We also thank Stepan Aleshin for helpful discussions and constructive comments. This study was funded by the federal state Saxony-Anhalt and the European Structural and Investment Funds (ESF, 2014-2020), project number ZS/2016/08/80645, as part of doctoral program ABINEP (Analysis, Imaging and Modelling of Neuronal Processes).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available with the online version here: https://doi.org/10.1162/imag_a_00278

REFERENCES

Albers, K. J., Ambrosen, K. S., Liptrot, M. G., Dyrby, T. B., Schmidt, M. N., & Mørup, M. (2021). Using connectomics for predictive assessment of brain parcellations. *Neuroimage*, *238*, 118170. https://doi.org/10.1016/j.neuroimage.2021.118170

Alink, A., Walther, A., Krugliak, A., van den Bosch, J. J., & Kriegeskorte, N. (2015). Mind the drift-improving sensitivity to fMRI pattern information by accounting for temporal pattern drift. *BioRxiv*, 032391. https://doi.org/10.1101/032391

Allefeld, C., Görgen, K., & Haynes, J.-D. (2016). Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *Neuroimage*, *141*, 378–392. https://doi.org/10.1016/j.neuroimage.2016.07.040

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *Cogn Sci*, *8*(1–2), e1373. https://doi.org/10.1002/wcs.1373

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. *Top Cogn Sci*, *11*(3), 482–503. https://doi.org/10.1111/tops.12420

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging*, *23*(2), 137–152. https://doi.org/10.1109/TMI.2003.822821

Behrens, T. E., Muller, T. H., Whittington, J. C., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Roy Statist Soc Ser B*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bi, Y., Wang, X., & Caramazza, A. (2016). Object domain and modality in the ventral visual pathway. *Trends Cognit Sci*, *20*(4), 282–290. https://doi.org/10.1016/j.tics.2016.02.002

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brants, M., Bulthé, J., Daniels, N., Wagemans, J., & Op de Beeck, H. P. (2016). How learning might strengthen existing visual object representations in human object-selective cortex. *Neuroimage*, *127*, 74–85. https://doi.org/10.1016/j.neuroimage.2015.11.063

Cai, M. B., Schuck, N. W., Pillow, J. W., & Niv, Y. (2019). Representational structure or task structure? Bias in neural representational similarity analysis and a Bayesian method for reducing bias. *PLoS Comput Biol*, *15*(5), e1006299. https://doi.org/10.1371/journal.pcbi.1006299

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The distributed nature of working memory. *Trends Cognit Sci*, *21*(2), 111–124. https://doi.org/10.1016/j.tics.2016.12.007

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cogn Psychol*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464–1468. https://doi.org/10.1126/science.aaf0941

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J Exp Psy Learn Mem Cognit*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Cox, D. D., Meier, P., Oertelt, N., & DiCarlo, J. J. (2005). 'Breaking' position-invariant object recognition. *Nat Neurosci*, *8*(9), 1145–1147. https://doi.org/10.1038/nn1519

Davachi, L., & DuBrow, S. (2015). How the hippocampus preserves order: The role of prediction and context. *Trends Cogn Sci*, *19*(2), 92–99. https://doi.org/10.1016/j.tics.2014.12.004

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, *463*(7281), 657–661. https://doi.org/10.1038/nature08704

Dornas, J. V., & Braun, J. (2018). Finer parcellation reveals detailed correlational structure of resting-state fMRI signals. *J Neurosci Meth*, *294*, 15–33. https://doi.org/10.1016/j.jneumeth.2017.10.020

Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2013). Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cerebral Cortex*, *23*(10), 2467–2478. https://doi.org/10.1093/cercor/bhs244

Eger, E., Ashburner, J., Haynes, J.-D., Dolan, R. J., & Rees, G. (2008). fMRI activity patterns in human loc carry information about object exemplars within category. *J Cogn Neurosci*, *20*(2), 356–370. https://doi.org/10.1162/jocn.2008.20019

Eichenbaum, H., Amaral, D. G., Buffalo, E. A., Buzsáki, G., Cohen, N., Davachi, L., Frank, L., Heckers, S., Morris, R. G., Moser, E. I., Nadel, L., O'Keefe, J., Preston, A., Ranganath, C., Silva, A., Witter, M. (2016). Hippocampus at 25. *Hippocampus*, *26*(10), 1238–1249. https://doi.org/10.1002/hipo.22616

Erickson, C. A., & Desimone, R. (1999). Responses of macaque perirhinal neurons during and after visual stimulus association learning. *J Neurosci*, *19*(23), 10404–10416. https://doi.org/10.1523/JNEUROSCI.19-23-10404.1999

Ferrari, A., Richter, D., & de Lange, F. P. (2022). Updating contextual sensory expectations for adaptive behavior. *J Neurosci*, *42*(47), 8855–8869. https://doi.org/10.1523/JNEUROSCI.1107-22.2022

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *J Exp Psychol Learn Mem Cognit*, *28*(3), 458–467. https://doi.org/10.1037/0278-7393.28.3.458

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *J Exp Psychol Gen*, *134*(4), 521–537. https://doi.org/10.1037/0096-3445.134.4.521

Fiser, J., & Lengyel, G. (2022). Statistical learning in vision. *Annu Rev Vis Sci*, *8*, 265–290. https://doi.org/10.1146/annurev-vision-100720-103343

Freud, E., Culham, J. C., Plaut, D. C., & Behrmann, M. (2017). The large-scale organization of shape processing in the ventral and dorsal pathways. *eLife*, *6*, e27576. https://doi.org/10.7554/eLife.34464

Friston, K. J., Jezzard, P., & Turner, R. (1994). Analysis of functional MRI time-series. *Hum Brain Mapp*, *1*(2), 153–171. https://doi.org/10.1002/hbm.460010207

Fyhn, M., Hafting, T., Treves, A., Moser, M.-B., & Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, *446*(7132), 190–194. https://doi.org/10.1038/nature05601

Gauthier, I., & Tarr, M. J. (2016). Visual object recognition: Do we (finally) know more now than we did? *Annu Rev Vis Sci*, *2*, 377–396. https://doi.org/10.1146/annurev-vision-111815-114621

Gheysen, F., Van Opstal, F., Roggeman, C., Van Waelvelde, H., & Fias, W. (2011). The neural basis of implicit perceptual sequence learning. *Front Hum Neurosci*, *5*, 137. https://doi.org/10.3389/fnhum.2011.00137

Gilron, R., Rosenblatt, J. D., & Mukamel, R. (2016). Addressing the "problem" of temporal correlations in MVPA analysis. In *2016 International Workshop on Pattern Recognition in Neuroimaging (PRNI), Trento, Italy* (pp. 1–4). IEEE. https://doi.org/10.1109/PRNI.2016.7552348

Giorgio, J., Karlaftis, V. M., Wang, R., Shen, Y., Tino, P., Welchman, A., & Kourtzi, Z. (2018). Functional brain networks for learning predictive statistics. *Cortex*, *107*, 204–219. https://doi.org/10.1016/j.cortex.2017.08.014

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Res*, *41*(10–11), 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nat Rev Neurosci*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *Neuroimage*, *62*(2), 852–855. https://doi.org/10.1016/j.neuroimage.2012.03.016

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., Doyle, W., Devinsky, O., & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Sci Adv*, *7*(8), eabc4530. https://doi.org/10.1126/sciadv.abc4530

Henriksson, L., Khaligh-Razavi, S.-M., Kay, K., & Kriegeskorte, N. (2015). Visual representations are dominated by intrinsic fluctuations correlated between areas. *Neuroimage*, *114*, 275–286. https://doi.org/10.1016/j.neuroimage.2015.04.026

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nat Neurosci*, *19*(5), 665–667. https://doi.org/10.1038/nn.4284

Hirabayashi, T., & Miyashita, Y. (2014). Computational principles of microcircuits for visual object processing in the macaque temporal cortex. *Trends Neurosci*, *37*(3), 178–187. https://doi.org/10.1016/j.tins.2014.01.002

Hsieh, L.-T., Gruber, M. J., Jenkins, L. J., & Ranganath, C. (2014). Hippocampal activity patterns carry information about objects in temporal context. *Neuron*, *81*(5), 1165–1178. https://doi.org/10.1016/j.neuron.2014.01.015

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *J Exp Psychol Gen*, *130*(4), 658–680. https://doi.org/10.1037/0096-3445.130.4.658

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med Image Anal*, *5*(2), 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

Jeong, S. K., & Xu, Y. (2016). Behaviorally relevant abstract object identity representation in the human parietal cortex. *J Neurosci*, *36*(5), 1607–1619. https://doi.org/10.1523/JNEUROSCI.1016-15.2016

Jia, X., Hong, H., & DiCarlo, J. J. (2021). Unsupervised changes in core object recognition behavior are predicted by neural plasticity in inferior temporal cortex. *eLife*, *10*, e60830. https://doi.org/10.7554/eLife.60830

Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cuing—Configuration or individual locations? *Percept Psychophys*, *66*, 454–463. https://doi.org/10.3758/BF03194893

Kahn, A. E., Karuza, E. A., Vettel, J. M., & Bassett, D. S. (2018). Network constraints on learnability of probabilistic motor sequences. *Nat Hum Behav*, *2*(12), 936–947. https://doi.org/10.1038/s41562-018-0463-8

Kakaei, E., Aleshin, S., & Braun, J. (2021). Visual object recognition is facilitated by temporal community structure. *Learn Mem*, *28*(5), 148–152. https://doi.org/10.1101/lm.053306.120

Kakaei, E., & Braun, J. (2024). Gradual change of cortical representations with growing visual expertise for synthetic shapes. *Imaging Neurosci*, *2*, 1–28. https://doi.org/10.1162/imag_a_00255

Kaposvari, P., Kumar, S., & Vogels, R. (2018). Statistical learning signals in macaque inferior temporal cortex. *Cerebral Cortex*, *28*(1), 250–266. https://doi.org/10.1093/cercor/bhw374

Karlaftis, V. M., Giorgio, J., Vértes, P. E., Wang, R., Shen, Y., Tino, P., Welchman, A. E., & Kourtzi, Z. (2019). Multimodal imaging of brain connectivity reveals predictors of individual decision strategy in statistical learning. *Nat Hum Behav*, *3*(3), 297–307. https://doi.org/10.1038/s41562-018-0503-4

Karuza, E. A., Emberson, L. L., Roser, M. E., Cole, D., Aslin, R. N., & Fiser, J. (2017). Neural signatures of spatial statistical learning: Characterizing the extraction of structure from complex visual scenes. *J Cogn Neurosci*, *29*(12), 1963–1976. https://doi.org/10.1162/jocn_a_01182

Karuza, E. A., Kahn, A. E., Thompson-Schill, S. L., & Bassett, D. S. (2017). Process reveals structure: How a network is traversed mediates expectations about its architecture. *Sci Rep*, *7*(1), 1–9. https://doi.org/10.1038/s41598-017-12876-5

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proc Natl Acad Sci USA*, *105*(31), 10687–10692. https://doi.org/10.1073/pnas.0802631105
Knudsen, E. B., & Wallis, J. D. (2022). Taking stock of value in the orbitofrontal cortex. *Nat Rev Neurosci*, *23*(7), 428–438. https://doi.org/10.1038/s41583-022-00589-2
Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nat Neurosci*, *11*(2), 224–231. https://doi.org/10.1038/nn2036
Kourtzi, Z., & Welchman, A. E. (2019). Learning predictive structure without a teacher: Decision strategies and brain routes. *Curr Opin Neurobiol*, *58*, 130–134. https://doi.org/10.1016/j.conb.2019.09.014
Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends Cognit Sci*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011
Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the onion of brain representations. *Annu Rev Neurosci*, *42*, 407–432. https://doi.org/10.1146/annurev-neuro-080317-061906
Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Front Syst Neurosci*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008
Kumar, M., Anderson, M. J., Antony, J. W., Baldassano, C., Brooks, P. P., Cai, M. B., Chen, P.-H. C., Ellis, C. T., Henselman-Petrusek, G., Huberdeau, D., Hutchinson, J. B., Li, Y. P., Lu, Q., Manning, J. R., Mennen, A. C., Nastase, S. A., Richard, H., Schapiro, A. C., Schuck, N. W., … Norman, K. A. (2022). BrainIAK: The brain imaging analysis kit. *Apert Neuro*, *1*(4), 1–19. https://doi.org/10.52294/31bb5b68-2184-411b-8c00-a1dacb61e1da
Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behav Brain Sci*, *40*, e253. https://doi.org/10.1017/S0140525X16001837
Lengyel, G., Žalalytė, G., Pantelides, A., Ingram, J. N., Fiser, J., Lengyel, M., & Wolpert, D. M. (2019). Unimodal statistical learning produces multimodal object-like representations. *eLife*, *8*, e43942. https://doi.org/10.7554/eLife.43942
Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*, *321*(5895), 1502–1507. https://doi.org/10.1126/science.1160028
Li, N., & DiCarlo, J. J. (2010). Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex. *Neuron*, *67*(6), 1062–1075. https://doi.org/10.1016/j.neuron.2010.08.029
Li, N., & DiCarlo, J. J. (2012). Neuronal learning of invariant object representation in the ventral visual stream is not dependent on reward. *J Neurosci*, *32*(19), 6611–6620. https://doi.org/10.1523/JNEUROSCI.3786-11.2012
Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annu Rev Neurosci*, *19*(1), 577–621. https://doi.org/10.1146/annurev.ne.19.030196.003045
Messinger, A., Squire, L. R., Zola, S. M., & Albright, T. D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proc Natl Acad Sci USA*, *98*(21), 12239–12244. https://doi.org/10.1073/pnas.211431098
Meyer, T., Ramachandran, S., & Olson, C. R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *J Neurosci*, *34*(28), 9332–9337. https://doi.org/10.1523/JNEUROSCI.1215-14.2014
Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, *335*(6193), 817–820. https://doi.org/10.1038/335817a0
Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Soc Cogn Affect Neurosci*, *14*(6), 667–685. https://doi.org/10.1093/scan/nsz037
Naya, Y., Yoshida, M., & Miyashita, Y. (2001). Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science*, *291*(5504), 661–664. https://doi.org/10.1126/science.291.5504.661
Naya, Y., Yoshida, M., Takeda, M., Fujimichi, R., & Miyashita, Y. (2003). Delay-period activities in two subdivisions of monkey inferotemporal cortex during pair association memory task. *Eur J Neurosci*, *18*(10), 2915–2918. https://doi.org/10.1111/j.1460-9568.2003.03020.x
Op de Beeck, H. P., & Baker, C. I. (2010). The neural basis of visual object learning. *Trends Cogn Sci*, *14*(1), 22–30. https://doi.org/10.1016/j.tics.2009.11.002
Patel, A. X., Kundu, P., Rubinov, M., Jones, P. S., Vértes, P. E., Ersche, K. D., Suckling, J., & Bullmore, E. T. (2014). A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *Neuroimage*, *95*, 287–304. https://doi.org/10.1016/j.neuroimage.2014.03.012
Perruchet, P. (2019). Dual nature of anticipatory classically conditioned reactions. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 179–198). Psychology Press. https://doi.org/10.4324/9781315792385-9
Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends Cognit Sci*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006
Poirier, C. C., De Volder, A. G., Tranduy, D., & Scheiber, C. (2006). Neural changes in the ventral and dorsal visual streams during pattern recognition learning. *Neurobiol Learn Mem*, *85*(1), 36–43. https://doi.org/10.1016/j.nlm.2005.08.006
Rosenthal, C. R., Andrews, S. K., Antoniades, C. A., Kennard, C., & Soto, D. (2016). Learning and recognition of a non-conscious sequence of events in human primary visual cortex. *Curr Biol*, *26*(6), 834–841. https://doi.org/10.1016/j.cub.2016.01.040
Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T., & Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *Neuroimage*, *60*(1), 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050
Rusu, S. I., & Pennartz, C. M. (2020). Learning, memory and consolidation mechanisms for behavioral control in hierarchically organized cortico-basal ganglia systems. *Hippocampus*, *30*(1), 73–98. https://doi.org/10.1002/hipo.23167
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926
Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annu Rev Psychol*, *69*, 181. https://doi.org/10.1146/annurev-psych-122216-011805
Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature*, *354*(6349), 152–155. https://doi.org/10.1038/354152a0

Sáringer, S., Fehér, Á., Sáry, G., & Kaposvári, P. (2022). Online measurement of learning temporal statistical structure in categorization tasks. *Mem Cogn*, *50*(7), 1530–1545. https://doi.org/10.3758/s13421-022-01302-5

Sáry, G., Vogels, R., & Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science*, *260*(5110), 995–997. https://doi.org/10.1126/science.8493538

Schapiro, A., & Turk-Browne, N. (2015). Statistical learning. *Brain Mapp*, *3*, 501–506. https://doi.org/10.1016/B978-0-12-397025-1.00276-1

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr Biol*, *22*(17), 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nat Neurosci*, *16*(4), 486–492. https://doi.org/10.1038/nn.3331

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, *26*(1), 3–8. https://doi.org/10.1002/hipo.22523

Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An fMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, *37*(6), 1013–1025. https://doi.org/10.1016/s0896-6273(03)00123-5

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human orbitofrontal cortex represents a cognitive map of state space. *Neuron*, *91*(6), 1402–1412. https://doi.org/10.1016/j.neuron.2016.08.019

Shafto, P., Kemp, C., Mansinghka, V., & Tenenbaum, J. B. (2011). A probabilistic model of cross-categorization. *Cognition*, *120*(1), 1–25. https://doi.org/10.1016/j.cognition.2011.02.010

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum Brain Mapp*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

Smith, S. M., & Brady, J. M. (1997). Susan—a new approach to low level image processing. *Int J Comput Vis*, *23*(1), 45–78. https://doi.org/10.1023/A:1007963824710

Tamura, K., Takeda, M., Setsuie, R., Tsubota, T., Hirabayashi, T., Miyamoto, K., & Miyashita, Y. (2017). Conversion of object identity to object-general semantic value in the primate temporal cortex. *Science*, *357*(6352), 687–692. https://doi.org/10.1126/science.aan4800

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Tian, M., & Grill-Spector, K. (2015). Spatiotemporal information during unsupervised learning enhances viewpoint invariant object recognition. *J Vis*, *15*(6), 7. https://doi.org/10.1167/15.6.7

Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *J Exp Psychol Learn Mem Cogn*, *34*(2), 399–407. https://doi.org/10.1037/0278-7393.34.2.399

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *J Exp Psychol Gen*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *J Cogn Neurosci*, *21*(10), 1934–1945. https://doi.org/10.1162/jocn.2009.21131

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *J Neurosci*, *30*(33), 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni MRI single-subject brain. *Neuroimage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Van Meel, C., & Op de Beeck, H. P. (2018). Temporal contiguity training influences behavioral and neural measures of viewpoint tolerance. *Front Hum Neurosci*, *12*, 13. https://doi.org/10.3389/fnhum.2018.00013

Van Meel, C., & Op de Beeck, H. P. (2020). An investigation of the effect of temporal contiguity training on size-tolerant representations in object-selective cortex. *Neuroimage*, *217*, 116881. https://doi.org/10.1016/j.neuroimage.2020.116881

Visconti di Oleggio Castello, M., Haxby, J. V., & Gobbini, M. I. (2021). Shared neural codes for visual and semantic information about familiar faces in a common representational space. *Proc Natl Acad Sci USA*, *118*(45), e2110474118. https://doi.org/10.1073/pnas.2110474118

Wallis, G., Backus, B. T., Langer, M., Huebner, G., & Bülthoff, H. (2009). Learning illumination-and orientation-invariant representations of objects through temporal association. *J Vis*, *9*(7), 6. https://doi.org/10.1167/9.7.6

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proc Natl Acad Sci USA*, *98*(8), 4800–4804. https://doi.org/10.1073/pnas.071028598

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, *25*(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277

Wang, R., Shen, Y., Tino, P., Welchman, A. E., & Kourtzi, Z. (2017). Learning predictive statistics: Strategies and brain mechanisms. *J Neurosci*, *37*(35), 8412–8427. https://doi.org/10.1523/JNEUROSCI.0144-17.2017

Weilnhammer, V., Fritsch, M., Chikermane, M., Eckert, A.-L., Kanthak, K., Stuke, H., Kaminski, J., & Sterzer, P. (2021). An active role of inferior frontal cortex in conscious experience. *Curr Biol*, *31*(13), 2868.e8–2880.e8. https://doi.org/10.1016/j.cub.2021.04.043

Weiner, K. S., & Zilles, K. (2016). The anatomical and functional specialization of the fusiform gyrus. *Neuropsychologia*, *83*, 48–62. https://doi.org/10.1016/j.neuropsychologia.2015.06.033

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005

Ye, J., Xiong, T., & Madigan, D. (2006). Computational and theoretical analysis of null space and orthogonal linear discriminant analysis. *J Mach Learn Res*, *7*(43), 1183–1204. http://jmlr.org/papers/v7/ye06a.html

Yu, H., & Yang, J. (2001). A direct LDA algorithm for high-dimensional data—With application to face recognition. *Pattern Recogn*, *34*(10), 2067–2070. https://doi.org/10.1016/S0031-3203(00)00162-X

Zarahn, E., Aguirre, G. K., & D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics. *Neuroimage*, *5*(3), 179–197. https://doi.org/10.1006/nimg.1997.0263

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imag*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

APPENDIX

Appendix Table A1. List of community-selective parcels and their anatomical region.

Region	Parcel	MNI			Community	Identity	Topog.
	No.	X	y	z	t_{BW}	α(%)	assign.
Middle frontal	109	48	-3	56	2.9	-	-
Calcarine	337	-12	-99	-5	5.2	13.8	V1d
	344	9	-85	6	5.2	13.8	V1v
Lingual	363	-12	-65	-5	3.8	9.5	V3v
Occipital	400	-38	-86	4	3	12.0	LO2
(middle)	405	-16	-100	1	6.3	14.0	V2d
Occipital	424	-31	-83	-8	4.5	12.6	-
(inferior)	428	36	-85	-7	5.9	13.2	hV4
	430	42	-73	-9	3.6	11.4	hV4
Fusiform	432	-27	-71	-11	4.9	12.2	VO2
	436	-31	-53	-13	4.1	8.8	PHC1
Postcentral	483	51	-26	47	3.8	-	-
Temporal	751	51	-67	-8	5.1	-	AIT
(inferior)	755	46	-53	-11	5.5	9.7	AIT
Superior frontal	45	-23	58	23	-4.1	-	-
	69	21	63	4	-6.4	-	-
Superior frontal (orbital)	77	23	61	-5	-6.8	-	-
Middle frontal	118	44	35	34	-3.8	-	-
Rolandic operculum	187	-42	-25	18	-3.2	-	-
Superior frontal (medial orbital)	248	9	65	-9	-4.6	-	-
Insula	270	39	20	-4	-3.6	-	-
	271	41	7	4	-3.8	-	-
Parietal (superior)	499	-28	-69	50	-4.2	8.9	-
Supramarginal	540	56	-44	29	-4.4	-	-
Precuneus	570	-3	-62	46	-3.5	-	-
Putamen	619	26	11	6	-4.2	-	-
Temporal (superior)	650	63	-8	5	-3.5	-	-

Numerical parcel ID, geometrical centroid x/y/z in MNI, between-community separability t^{BW}, identity classification α, and topographical assignment, if any.

References

M. J. Anderson. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1):32–46, 2001. doi: 10.1111/j.1442-9993.2001.01070.pp.x.

L. J. Batterink, K. A. Paller, and P. J. Reber. Understanding the neural bases of implicit and statistical learning. *Topics in Cognitive Science*, 11(3):482–503, 2019. doi: 10.1111/tops.12420.

M. Booth and E. T. Rolls. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral cortex (New York, NY: 1991)*, 8(6):510–523, 1998. doi: 10.1093/cercor/8.6.510.

M. M. Chun and Y. Jiang. Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive psychology*, 36(1):28–71, 1998.

D. D. Cox, P. Meier, N. Oertelt, and J. J. DiCarlo. 'breaking'position-invariant object recognition. *Nature neuroscience*, 8(9):1145–1147, 2005. doi: 10.1038/nn1519.

A. Derrington and P. Lennie. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *The Journal of physiology*, 357(1):219–240, 1984.

J. V. Dornas and J. Braun. Finer parcellation reveals detailed correlational structure of resting-state fmri signals. *Journal of Neuroscience Methods*, 294:15–33, 2018.

R. Epstein and N. Kanwisher. A cortical representation of the local visual environment. *Nature*, 392(6676):598–601, 1998.

C. A. Erickson and R. Desimone. Responses of macaque perirhinal neurons during and after visual stimulus association learning. *Journal of Neuroscience*, 19(23): 10404–10416, 1999. doi: 10.1523/JNEUROSCI.19-23-10404.1999.

D. J. Felleman, Y. Xiao, and E. McClendon. Modular organization of occipito-temporal pathways: cortical connections between visual area 4 and visual area 2 and posterior inferotemporal ventral area in macaque monkeys. *Journal of Neuroscience*, 17(9):3185–3200, 1997.

J. Fiser and R. N. Aslin. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12(6):499–504, 2001. doi: 10.1111/1467-9280.00392.

J. Fiser and R. N. Aslin. Statistical learning of higher-order temporal structure from visual shape sequences. *J. Exp. Psychol. Learn. Mem. Cognit.*, 28(3):458, 2002. doi: 10.1037/0278-7393.28.3.458.

J. Fiser and G. Lengyel. Statistical learning in vision. *Annu. Rev. Vis. Sci.*, 8: 265–290, 2022. doi: 10.1146/annurev-vision-100720-103343.

F. Gheysen, F. Van Opstal, C. Roggeman, H. Van Waelvelde, and W. Fias. The neural basis of implicit perceptual sequence learning. *Front. Hum. Neurosci.*, 5: 137, 2011. doi: 10.3389/fnhum.2011.00137.

K. Grill-Spector and K. S. Weiner. The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, 15(8): 536–548, 2014. doi: 10.1038/nrn3747.

K. Grill-Spector, T. Kushnir, S. Edelman, G. Avidan, Y. Itzchak, and R. Malach. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, 24(1):187–203, 1999. doi: 10.1016/S0896-6273(00)80832-6.

K. Grill-Spector, Z. Kourtzi, and N. Kanwisher. The lateral occipital complex and its role in object recognition. *Vision research*, 41(10-11):1409–1422, 2001.

K. Grill-Spector, R. Henson, and A. Martin. Repetition and the brain: neural models of stimulus-specific effects. *Trends in cognitive sciences*, 10(1):14–23, 2006. doi: 10.1016/j.tics.2005.11.006.

J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, 2001.

N. C. Hindy, F. Y. Ng, and N. B. Turk-Browne. Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nat. Neurosci.*, 19(5):665–667, 2016. doi: 10.1038/nn.4284.

L.-T. Hsieh, M. J. Gruber, L. J. Jenkins, and C. Ranganath. Hippocampal activity patterns carry information about objects in temporal context. *Neuron*, 81(5):1165–1178, 2014. doi: 10.1016/j.neuron.2014.01.015.

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106, 1962.

R. H. Hunt and R. N. Aslin. Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, 130(4):658, 2001.

A. E. Kahn, E. A. Karuza, J. M. Vettel, and D. S. Bassett. Network constraints on learnability of probabilistic motor sequences. *Nature human behaviour*, 2(12):936–947, 2018. doi: 10.1038/s41562-018-0463-8.

E. Kakaei and J. Braun. Gradual change of cortical representations with growing visual expertise for synthetic shapes. *Imaging Neuroscience*, 2024a. doi: 10.1162/imag_a_00255.

E. Kakaei and J. Braun. Incidental learning of predictive temporal context within cortical representations of visual shape. *Imaging Neuroscience*, 2:1–23, 2024b.

E. Kakaei, S. Aleshin, and J. Braun. Visual object recognition is facilitated by temporal community structure. *Learning & Memory*, 28(5):148–152, 2021.

N. Kanwisher, R. P. Woods, M. Iacoboni, and J. C. Mazziotta. A locus in human extrastriate cortex for visual shape analysis. *Journal of Cognitive Neuroscience*, 9 (1):133–142, 1997.

N. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. 2002.

E. A. Karuza, A. E. Kahn, S. L. Thompson-Schill, and D. S. Bassett. Process reveals structure: How a network is traversed mediates expectations about its architecture. *Scientific reports*, 7(1):12733, 2017. doi: 10.1038/s41598-017-12876-5.

R. Kiani, H. Esteky, K. Mirpour, and K. Tanaka. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of neurophysiology*, 97(6):4296–4309, 2007.

E. Kobatake and K. Tanaka. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of neurophysiology*, 71(3):856–867, 1994.

D. J. Kravitz, K. S. Saleem, C. I. Baker, L. G. Ungerleider, and M. Mishkin. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, 17(1):26–49, 2013. doi: 10.1016/j.tics.2012.10.011.

N. Kriegeskorte, M. Mur, and P. A. Bandettini. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:249, 2008a. doi: 10.3389/neuro.06.004.2008.

N. Kriegeskorte, M. Mur, D. A. Ruff, R. Kiani, J. Bodurka, H. Esteky, K. Tanaka, and P. A. Bandettini. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6):1126–1141, 2008b.

N. Li and J. J. DiCarlo. Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *science*, 321(5895):1502–1507, 2008. doi: 10.1126/science.1160028.

N. Li and J. J. DiCarlo. Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex. *Neuron*, 67(6): 1062–1075, 2010. doi: 10.1016/j.neuron.2010.08.029.

N. Li and J. J. DiCarlo. Neuronal learning of invariant object representation in the ventral visual stream is not dependent on reward. *Journal of Neuroscience*, 32 (19):6611–6620, 2012. doi: 10.1523/JNEUROSCI.3786-11.2012.

D. B. McMahon and D. A. Leopold. Stimulus timing-dependent plasticity in high-level vision. *Current biology*, 22(4):332–337, 2012. doi: 10.1016/j.cub.2012.01.003.

T. Meyer and C. R. Olson. Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences*, 108(48): 19401–19406, 2011. doi: 10.1073/pnas.1112895108.

T. Meyer, S. Ramachandran, and C. R. Olson. Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *Journal of Neuroscience*, 34(28):9332–9337, 2014. doi: 10.1523/JNEUROSCI.1215-14.2014.

G. Mirabella, G. Bertini, I. Samengo, B. E. Kilavik, D. Frilli, C. Della Libera, and L. Chelazzi. Neurons in area v4 of the macaque translate attended visual features into behaviorally relevant categories. *Neuron*, 54(2):303–318, 2007.

Y. Miyashita. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, 335(6193):817–820, 1988. doi: 10.1038/335817a0.

K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby. Beyond mind-reading: multi-voxel pattern analysis of fmri data. *Trends in cognitive sciences*, 10(9):424–430, 2006. doi: 10.1016/j.tics.2006.07.005.

T. J. Palmeri and M. Tarr. Visual object perception and long-term memory. *Visual memory*, pages 163–207, 2008. doi: 10.1093/acprof:oso/9780195305487.003.0006.

A. W. Roe, L. Chelazzi, C. E. Connor, B. R. Conway, I. Fujita, J. L. Gallant, H. Lu, and W. Vanduffel. Toward a unified theory of visual area v4. *Neuron*, 74(1):12–29, 2012.

N. C. Rust and J. J. DiCarlo. Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area v4 to it. *Journal of Neuroscience*, 30(39):12978–12995, 2010.

J. R. Saffran, R. N. Aslin, and E. L. Newport. Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–1928, 1996a. doi: 10.1126/science.274.5294.1926.

J. R. Saffran, E. L. Newport, and R. N. Aslin. Word segmentation: The role of distributional cues. *Journal of memory and language*, 35(4):606–621, 1996b. doi: 10.1006/jmla.1996.0032.

K. Sakai and Y. Miyashita. Neural organization for the long-term memory of paired associates. *Nature*, 354(6349):152–155, 1991. doi: 10.1038/354152a0.

A. C. Schapiro, L. V. Kustner, and N. B. Turk-Browne. Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, 22(17):1622–1627, 2012. doi: 10.1016/j.cub.2012.06.056.

A. C. Schapiro, T. T. Rogers, N. I. Cordova, N. B. Turk-Browne, and M. M. Botvinick. Neural representations of events arise from temporal community structure. *Nat. Neurosci.*, 16(4):486–492, 2013. doi: 10.1038/nn.3331.

A. C. Schapiro, N. B. Turk-Browne, K. A. Norman, and M. M. Botvinick. Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, 26 (1):3–8, 2016. doi: 10.1002/hipo.22523.

P. H. Schiller and N. K. Logothetis. The color-opponent and broad-band channels of the primate visual system. *Trends in neurosciences*, 13(10):392–398, 1990.

P. H. Schiller, B. L. Finlay, and S. F. Volman. Quantitative studies of single-cell properties in monkey striate cortex. ii. orientation specificity and ocular dominance. *Journal of neurophysiology*, 39(6):1320–1333, 1976.

N. B. Turk-Browne, B. J. Scholl, M. M. Chun, and M. K. Johnson. Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of cognitive neuroscience*, 21(10):1934–1945, 2009. doi: 10.1162/jocn.2009.21131.

N. B. Turk-Browne, B. J. Scholl, M. K. Johnson, and M. M. Chun. Implicit perceptual anticipation triggered by statistical learning. *J. Neurosci.*, 30(33):11177–11187, 2010. doi: 10.1523/JNEUROSCI.0858-10.2010.

N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

D. C. Van Essen and J. L. Gallant. Neural mechanisms of form and motion processing in the primate visual system. *Neuron*, 13(1):1–10, 1994.

C. Van Meel and H. P. Op de Beeck. Temporal contiguity training influences behavioral and neural measures of viewpoint tolerance. *Frontiers in Human Neuroscience*, 12:13, 2018. doi: 10.3389/fnhum.2018.00013.

G. Wallis and H. H. Bülthoff. Effects of temporal association on recognition memory. *Proc. Natl. Acad. Sci. USA*, 98(8):4800–4804, 2001. doi: 10.1073/pnas.071028598.

G. Wallis, B. T. Backus, M. Langer, G. Huebner, and H. Bülthoff. Learning illumination-and orientation-invariant representations of objects through temporal association. *J. Vis.*, 9(7):6–6, 2009. doi: 10.1167/9.7.6.

J. Ye, T. Xiong, and D. Madigan. Computational and theoretical analysis of null space and orthogonal linear discriminant analysis. *Journal of Machine Learning Research*, 7(7), 2006. URL `http://jmlr.org/papers/v7/ye06a.html`.

H. Yu and J. Yang. A direct lda algorithm for high-dimensional data—with application to face recognition. *Pattern Recognition*, 34(10):2067–2070, 2001. doi: 10.1016/S0031-3203(00)00162-X.

Appendix A

Mathematical methods

A.1 Cross-validation measures

A data point k in an N_d dimensional space is defined as a vector $\mathbf{x_k}$:

$$\mathbf{x_k} = [x_{ik} | i \in \{1...N_d\}] \tag{A.1}$$

The **Euclidean distance** d_{kl} and the **normalized distance** $\hat{d}_{kl}$ between two data points k and l are calculated by:

$$\begin{aligned} d_{kl} &= \sqrt{\sum_{i}^{N_d} (x_{ik} - x_{il})^2} \\ \hat{d}_{kl} &= \frac{d_{kl}}{\sqrt{N_d}} \end{aligned} \tag{A.2}$$

A class L of size N_L, and its corresponding **train-set** L^{train} and **test-set** L^{test},

are defined as:

$$\begin{aligned} X_L &= \{\mathbf{x_k} | \forall k \in L\} \\ X_L^{train} &= \{\mathbf{x_k} | \forall k \in L^{train}\} \\ X_L^{test} &= \{\mathbf{x_k} | \forall k \in L^{test}\} \end{aligned} \tag{A.3}$$

For each class, the **centroids** $\mathbf{c_L^{tn}}$ of the train-set X_L^{train} and $\mathbf{c_L^{ts}}$ of the test-sets X_L^{test} are defined as vectors:

$$\begin{aligned} \mathbf{c_L^{tn}} = \langle X_L^{train} \rangle_k = \frac{1}{N_L^{train}} \sum_{\forall k \in L^{train}} x_{ik} = [c_{ik}^{tn} | i \in \{1...N_d\}] \\ \mathbf{c_L^{ts}} = \langle X_L^{test} \rangle_k = \frac{1}{N_L^{test}} \sum_{\forall k \in L^{test}} x_{ik} = [c_{ik}^{ts} | i \in \{1...N_d\}] \end{aligned} \tag{A.4}$$

A.1.1 Classification accuracy

In order to calculate the accuracy of a classifier in a classification problem, first we calculated the distance $D_{LL'} = D(X_L^{test}, \mathbf{c_{L'}^{tn}})$ of all test data of class L from centroid of the training class L'. Then, we defined the **accuracy** α as the probability of finding the minimum distance between a test-set and its corresponding centroid of the train-set:

$$\begin{aligned} \Delta D_{LL'} &= D_{LL} - D_{LL'} \\ \alpha &= P(\Delta D_{LL'} < 0 | L \neq L') \end{aligned} \tag{A.5}$$

A.1.2 Class-pair discriminability

For two classes L and L', first we projected their test-data on the line $\hat{\mathbf{e}}_{LL'}$ connecting their train centroids:

$$\begin{aligned} \mathbf{e}_{LL'} &= \mathbf{c}_{\mathbf{L}}^{\mathbf{tn}} - \mathbf{c}_{\mathbf{L'}}^{\mathbf{tn}} \\ \hat{\mathbf{e}}_{LL'} &= \frac{\mathbf{e}_{LL'}}{\|\mathbf{e}_{LL'}\|} \end{aligned} \tag{A.6}$$

Then, we calculated the mean and the standard deviation of the projected data:

$$\begin{aligned} \mu_L &= \frac{1}{N_L^{test}} \sum_{\forall k \in L^{test}} \mathbf{x_k} \cdot \hat{\mathbf{e}}_{LL'} \\ \sigma_L^2 &= \frac{1}{N_L^{test} - 1} \sum_{\forall k \in L^{test}} |\mathbf{x_k} \cdot \hat{\mathbf{e}}_{LL'} - \mu_L|^2 \end{aligned} \tag{A.7}$$

Finally, we defined the **pair-wise discriminability** $\delta_{LL'}$ and average discriminability δ as:

$$\begin{aligned} \delta_{LL'} &= \frac{|\mu_L - \mu_{L'}|}{\sqrt{0.5(\sigma_L^2 + \sigma_{L'}^2)}} \\ \delta &= \frac{1}{\kappa(\kappa - 1)} \sum_{L=1}^{\kappa} \sum_{L'=1}^{\kappa} (\delta_{LL'} | L' \neq L) \end{aligned} \tag{A.8}$$

A.1.3 F-ratio

We used the non-parametric approach of Anderson (2001) to calculate an overall discriminability measure between all classes. First, we calculate the within-class variance SS_W as the average squared distances between all the test-data within a

class and their corresponding test centroids:

$$\begin{aligned} SS_{WL} &= \sum_{\forall k \in L^{test}} D(X_L^{test}, \mathbf{c_L^{ts}})^2 \\ SS_W &= \frac{1}{N^{test}} \sum_{L=1}^{\kappa} SS_{WL} \end{aligned} \tag{A.9}$$

where $N^{test} = \sum_{L=1}^{\kappa} N_L^{test}$.

Then, we calculated the between-class variance SS_B and the total variance SS_T as:

$$\begin{aligned} \mathbf{c^{ts}} &= \langle X^{test} \rangle_k \\ SS_B &= \frac{1}{N^{test}} \sum_{L=1}^{\kappa} N_L^{test} D(\mathbf{c^{ts}}, \mathbf{c_L^{ts}})^2 \\ SS_T &= \frac{1}{N^{test}} \sum_{L=1}^{\kappa} \sum_{\forall k \in L^{test}} D(X_L^{test}, \mathbf{c_L^{ts}})^2 \end{aligned} \tag{A.10}$$

Finally, we calculated the ***F*-ratio** F as:

$$F = \frac{SS_B(N^{test} - \kappa)}{SS_W(\kappa - 1)} \tag{A.11}$$

Declaration of Honour

“I hereby declare that I prepared this thesis without the impermissible help of third parties and that none other that the aids indicated have been used; all sources of information are clearly marked, including my own publications.

In particular I have not consciously:

- fabricated data or rejected undesirable results,
- misused statistical methods with the aim of drawing other conclusions than those warranted by the available data,
- plagiarized external data or publications,
- presented the results of other researchers in a distorted way.

I am aware that violations of copyright may lead to injunction and damage claims by the author and also to prosecution by law enforcement authorities.

I hereby agree that the thesis may be electronically reviewed with the aim of identifying plagiarism.

This work has not been submitted as a doctoral thesis in the same or a similar form in Germany, nor in any other country. It has not yet been published as a whole.”

Tübingen, 28.01.2025

Ehsan Kakaei