Spatio-temporally efficient coding assigns functions to hierarchical structures of the visual system

Duho Sihn[1], Sung-Phil Kim[1]*

[1] Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan, 44919, Republic of Korea

* Corresponding author

E-mail: spkim@unist.ac.kr

Abstract

Hierarchical structures constitute a wide array of brain areas, including the visual system. One of the important questions regarding visual hierarchical structures is to identify computational principles for assigning functions that represent the external world to hierarchical structures of the visual system. Given that visual hierarchical structures contain both bottom-up and top-down pathways, the derived principles should encompass these bidirectional pathways. However, existing principles such as predictive coding do not provide an effective principle for bidirectional pathways. Therefore, we propose a novel computational principle for visual hierarchical structures as spatio-temporally efficient coding underscored by the efficient use of given resources in both neural activity space and processing time. This coding principle optimises bidirectional predictions over hierarchical structures by simultaneously minimising temporally differences in neural responses and maximising entropy in neural representations. Simulations demonstrated that the proposed spatio-temporally efficient coding was able to assign the function of appropriate neural representations of natural visual scenes to visual hierarchical structures. Furthermore, spatio-temporally efficient coding was able to predict well-known phenomena, including deviations in neural responses to unfamiliar inputs and bias in preferred orientations. Our proposed spatio-temporally efficient coding may facilitate deeper mechanistic understanding of the computational processes of hierarchical brain structures.

Keywords

Efficient coding, Hierarchical structure, Neural response, Neural representation, Receptive field, Selectivity, Visual system.

1. Introduction

It is well-established that a wide array of brain areas have a hierarchical structure, including the visual system (Felleman and Van Essen, 1991; Mesulam, 1998; Harris et al., 2019; Hilgetag and Goulas, 2020). Studies have identified a link between hierarchical structures and gene expression (Burt et al., 2018; Hansen et al., 2021), suggesting that hierarchical structures are genetically determined *a priori*. Given that one of the major functions of the brain is to represent the external world (deCharms and Zador, 2000; Kriegeskorte and Diedrichsen, 2019), an ensuing question arises: How do *a priori* hierarchical brain structures attain functions to represent the external world? This question can be addressed by identifying a fundamental neural coding principle that assigns representational functions to hierarchical structures.

The traditional view of the neural coding of visual hierarchical structures is bottom-up visual information processing, whereby simple features are processed in a lower visual hierarchy and more complex features created by integrating simple features are processed in a higher visual hierarchy (Hubel and Wiesel, 1962; Hubel and Wiesel, 1968; Riesenhuber and Poggio, 1999; Riesenhuber and Poggio, 2000; Serre et al., 2007; DiCarlo et al., 2012; Yamins et al., 2014). However, this view does not consider the role of top-down pathways that are abundant even in the early visual system, such as the lateral geniculate nucleus (Murphy and Sillito, 1987; Wang et al., 2006) and primary visual cortex (Zhang et al., 2014; Muckli et al., 2015; Huh et al., 2018).

The role of top-down visual processing is especially prominent in predictive coding (Rao and Ballard, 1999; Spratling, 2017). According to predictive coding, a higher hierarchy performs top-down predictions of neural responses in a lower hierarchy. Both inference and learning of predictive coding are based on the minimisation of bottom-up prediction errors. Predictive coding has been used to explain the neural responses corresponding to prediction errors (Friston, 2005) and extends from the explanations of perception to action (Friston, 2010; Clark, 2013). A recent study combined predictive coding with sparse coding (i.e., sparse deep predictive coding) and demonstrate that it could enhance perceptual explanatory power (Boutin et al., 2021).

Nevertheless, predictive coding has several theoretical shortcomings. Since inference in predictive coding aims to minimise prediction errors, the hierarchical structure would require an additional

information processing subsystem to perform this inference. In addition, because bottom-up transmitted information contains only prediction errors, predictive coding requires the presence of error units (biological neurons) in the hierarchical structure to represent this prediction error, yet such error units remain as hypothetical entities and evidence for prediction error responses is limited in some conditions (Solomon et al., 2021).

A possible approach to overcome the shortcomings of both bottom-up processing and predictive coding in visual hierarchical structures is to make bottom-up predictions similarly to top-down predictions across hierarchies, instead of transmitting bottom-up prediction errors. These bidirectional predictions would realise both context-independent bottom-up predictions and context-dependent top-down predictions (Teufel and Fletcher, 2020). Such bidirectional predictions eliminate the necessity for hypothetical error units, while presumably elucidating the neural responses of hierarchical structures underlying bottom-up feature integration and top-down predictive coding. A neural coding principle underlying bidirectional predictions of hierarchical structures can be found in the theory of efficient coding that draws upon the efficient use of given resources (Laughlin, 2001; Bullmore and Sporns, 2012), which crucially include limited time resources related to processing speed (Griffiths et al., 2015; Lieder and Griffiths, 2020). A possible solution to promote the most efficient use of limited time resources by the bidirectional prediction system is to render present neural responses equal to future ones before the occurrence of future neural responses. This can be achieved by minimising the

temporal differences between present and future neural responses. Accordingly, we consider this temporal difference minimisation as our *learning principle*, referred to as *temporally efficient coding*. A minimal temporal difference indicates that bidirectional prediction is accurate. Here, *inference* simply refers to a bidirectional prediction mediated by top-down and bottom-up pathways. Unlike inference in predictive coding, which requires further error minimisation, inference in temporally efficient coding involves simple monosynaptic transmission (i.e., a connection that passes via a single step rather than multiple steps).

Since temporally efficient coding includes a trivial solution, in that neural responses are invariant to changes in external events. we circumvent this issue by adding a complementary neural coding (learning) principle that maximises the informational entropy of neural responses. It maximises the neural response space available to represent the external world under the constraints of both the number of neurons and maximum firing rates. Maximal entropy coding indicates that the system uses spatial resources of neural responses efficiently (Attneave, 1954; Barlow, 1961; Laughlin, 1981), referred to as *spatially efficient coding*. By combining spatially efficient coding and temporally efficient coding, we propose a neural coding principle termed *spatio-temporally efficient coding* (Fig 1).

The learning of spatio-temporally efficient coding renders bidirectional predictions of future neural responses by minimising temporal differences between present and future neural responses. The brain may also need such future predictions in order to overcome delays in information transmission

102 and to represent the external world in real time (Berry et al., 1999). Recent studies have revealed that
103 there is information for future visual inputs in the sensory cortex (Palmer et al., 2015; Chen et al.,
104 2017; Sederberg et al., 2018; Liu et al., 2021). Therefore, these future predictions in the proposed
105 spatio-temporally efficient coding are biologically plausible.

106 It is noteworthy that minimising temporal differences in upper hierarchies may, at first glance, be
107 reminiscent of the slow feature analysis (Wiskott and Sejnowski, 2002; Berkes and Wiskott, 2005;
108 Creutzig and Sprekeler, 2008). However, the upper hierarchies of the proposed spatio-temporally
109 efficient coding need to change dynamically to predict rapidly changing visual inputs and neural
110 responses of the lower hierarchies, whereas the upper hierarchies of the slow feature analysis need
111 to change slowly. Minimising temporal differences in spatio-temporally efficient coding does not aim at
112 extracting slow features, but rather aim at the representation of rapidly changing visual inputs with
113 stabilizing neural responses as quickly as possible.

114 To determine how a visual hierarchical structure with spatio-temporally efficient coding represents
115 external visual events, we repeatedly exposed a visual hierarchical structure to natural scene images,
116 enabling it to learn the parameters of bidirectional prediction networks between top-down and bottom-
117 up hierarchies using spatio-temporally efficient coding (Fig 1). Using simulations, we observed that
118 spatio-temporally efficient coding enabled the visual hierarchical structure to implement desirable
119 neural representations. In addition, the visual hierarchical structure learned to represent natural scene

images by spatio-temporally efficient coding and could predict well-known perceptual responses. First, the visual hierarchical structure exhibited deviant neural responses to unfamiliar visual inputs (i.e., neural responses to unfamiliar inputs were smaller or larger than those to familiar ones), indicating selectivity (Margoliash, 1983; Waydo et al., 2006) and/or familiarity suppression (Huang et al., 2018; Issa et al., 2018) of visual cortical neurons. Second, the visual hierarchical structure demonstrated simple cell-like receptive fields with preference for horizontal and vertical orientations over oblique orientations, resembling the orientation preference of visual cortical neurons (Furmanski and Engel, 2000; Li et al., 2003; Girshick et al., 2011). These simulation results demonstrated that the proposed spatio-temporally efficient coding could assign functions representing the external world to hierarchical structures in the visual system without theoretical complexity and hypothetical entities.

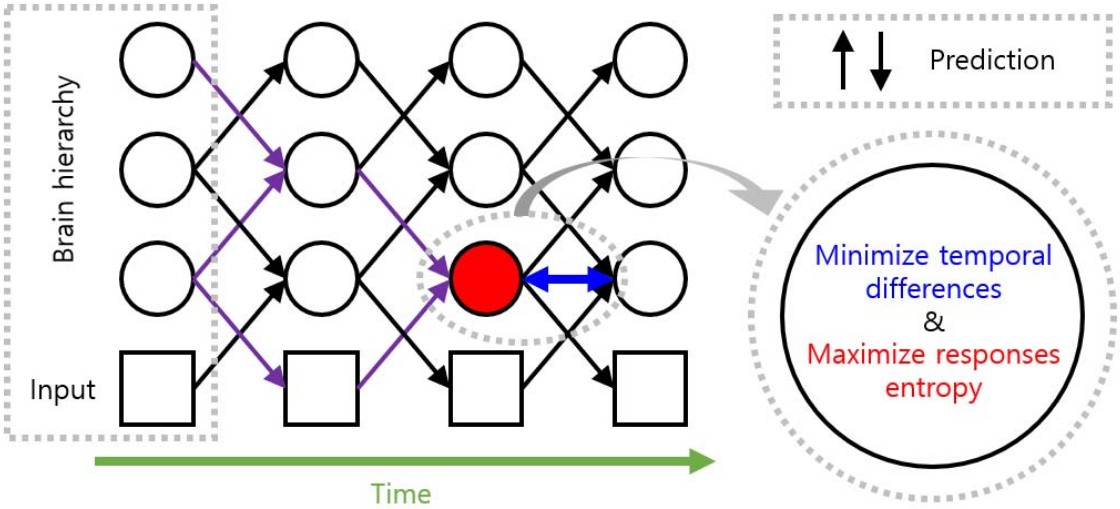


Fig 1. Spatio-temporally efficient coding. This illustration depicts a hierarchical structure of the brain and learning objectives. The hierarchical structure learns to represent input from the external world, which is depicted as black squares (e.g., visual input). Open black circles indicate an ensemble of neuronal units in each hierarchy of the brain. Inference based on spatio-temporally efficient coding is made by neuronal units as bottom-up and top-down predictions over time (black arrows). Learning in spatio-temporally efficient coding consists of two objectives: minimising the temporal differences between present and future neural responses and maximising the informational entropy of neural responses. For example, predictions (purple arrows) are optimised to minimise the temporal differences between present neural responses at the corresponding hierarchy (red filled circle) and future neural responses (circle to the right of the red filed circle) while concurrently maximising the informational entropy of neural responses at the corresponding hierarchy (red filled circle).

2. Results

2.1. Spatio-temporally efficient coding in visual hierarchical structures

In the present study, visual information processing in hierarchical structures was established as biologically inspired temporal processing. Specifically, visual information processing is described as a function f_t for both image X_{image} and neural responses X_h in each hierarchy h such that it maps from X_{image} and X_h at time $t-1$ to those at time t:

$$f_t: X_{image,t-1} \times X_{h=1,t-1} \times \ldots \times X_{h=H,t-1} \rightarrow X_{image,t} \times X_{h=1,t} \times \ldots \times X_{h=H,t} \quad (1)$$

where $X_{h,t-1}$ and $X_{h,t}$ are the neural responses X_h at time $t-1$ and t, respectively, at hierarchy h. The function values of adjacent hierarchies X_h and X_{h+1}, as well as those for X_{image} and $X_{h=1}$, are mutually influenced, representing the hierarchical structure. These influences are expressed as mathematical transformations based on synaptic weights, biases, and sigmoid nonlinearities added to the transformations (see Section 4.2 for details of the transformations). This function f_t makes

inferences using spatio-temporally efficient coding.

Learning in spatio-temporally efficient coding minimises the ensuing objectives of both temporally and spatially efficient coding. The objective of temporally efficient coding is given by:

$$L_{Temporal} = \sum_{h=0}^{H} \sum_{n=1}^{N} \left(f_t \circ f_{t-1}\left(X_{image=n}\right) - f_{t+1} \circ f_t \circ f_{t-1}\left(X_{image=n}\right) \right)^2 \quad (2)$$

where $\circ$ is the function composition, $X_{image=n}$ indicates that the image is fixed to the nth sample throughout the temporal processing, squaring is operated component-wise, and the outer summation $\sum_{h=0}^{H}(\cdot)$ is operated on the range of f where $h = 0$ denotes $X_{h=0} := X_{image}$. The minimisation of the given objective minimises the temporal differences between present and future neural responses. The objective of spatially efficient coding is given by

$$L_{Spatial} = \sum_{h=1}^{H} \sum_{n=1}^{N} \log P\left(f_t\left(X_{image=n}\right)|_{X_h} \right) \quad (3)$$

where $f_t(\cdot)|_{X_h}$ indicates a restriction of the range of the function value $f_t(\cdot)$ to X_h, and $P(\cdot)$ is a probability. Because the term within the summations is an estimation of negative informational entropy, minimising the objective maximises the informational entropy of f_t. Finally, the objective of spatio-temporally efficient coding is a linear combination of those two objectives:

$$L = L_{Temporal} + \lambda L_{Spatial} \quad (4)$$

where λ is a regularisation parameter. A smaller λ indicates a greater emphasis on the temporally efficient coding objective, whereas a larger λ indicates the opposite.

In the present study, the depth H of hierarchies was set to 2, the minimum depth to realise both bottom-up and top-down pathways in the same hierarchy (Fig 2A). For a given image, the duration of temporal processing of f_t was given as 5 (i.e., $t \in [1,5]$) in both learning and inference, because four time steps are required for the image information to reach the top hierarchy and return over $H = 2$ hierarchies, in addition to one time step to obtain future neural responses. The minibatch size N was set to 40 (see Section 4.2. for more details).

In our simulations, we repeatedly exposed the hierarchical structure to natural scene images, which enabled it to learn the bidirectional predictions between top-down and bottom-up hierarchies using spatio-temporally efficient coding with a range of the balancing parameter λ. Successful learning was confirmed by minimising or stabilising L during learning. Further, we verified that the learned hierarchical structure could successfully reconstruct an input image, as shown in Fig 2C.

2.2. Balancing between temporally and spatially efficient coding

The temporally efficient coding objective that minimises the temporal difference between present and future neural responses not only renders a neural representation suitable for the objective but also renders the visual information process temporally stable. In contrast, the spatially efficient coding objective that maximises the informational entropy of neural responses renders the process noisy.

Based on our simulations, we confirmed this phenomenon by observing that neural responses were temporally stable with smaller λ (i.e., weighing toward temporally efficient coding), whereas they became noisy with larger λ (i.e., weighing toward spatially efficient coding) (Fig 2C). In addition, when the λ values were too small or too large (i.e., strongly biased to either temporally or spatially efficient coding), the hierarchical structure failed to represent visual images appropriately (Fig 2D), highlighting the need for an appropriate balance between spatially efficient coding and temporally efficient coding in the visual hierarchical structure.

A change in balance between the two objectives also altered the relative strengths between bottom-up and top-down synaptic connections. Simulation results demonstrated that top-down synaptic strengths from hierarchies 2 to 1 were increased compared to bottom-up synaptic strengths from images to hierarchy 1 as λ increased, with an emphasis on spatially efficient coding (Fig 2B, see Section 4.2 for details on the computation of synaptic strength). This finding suggests that the balance between spatially and temporally efficient coding could be related to the balance between bottom-up and top-down synaptic strengths in the brain.

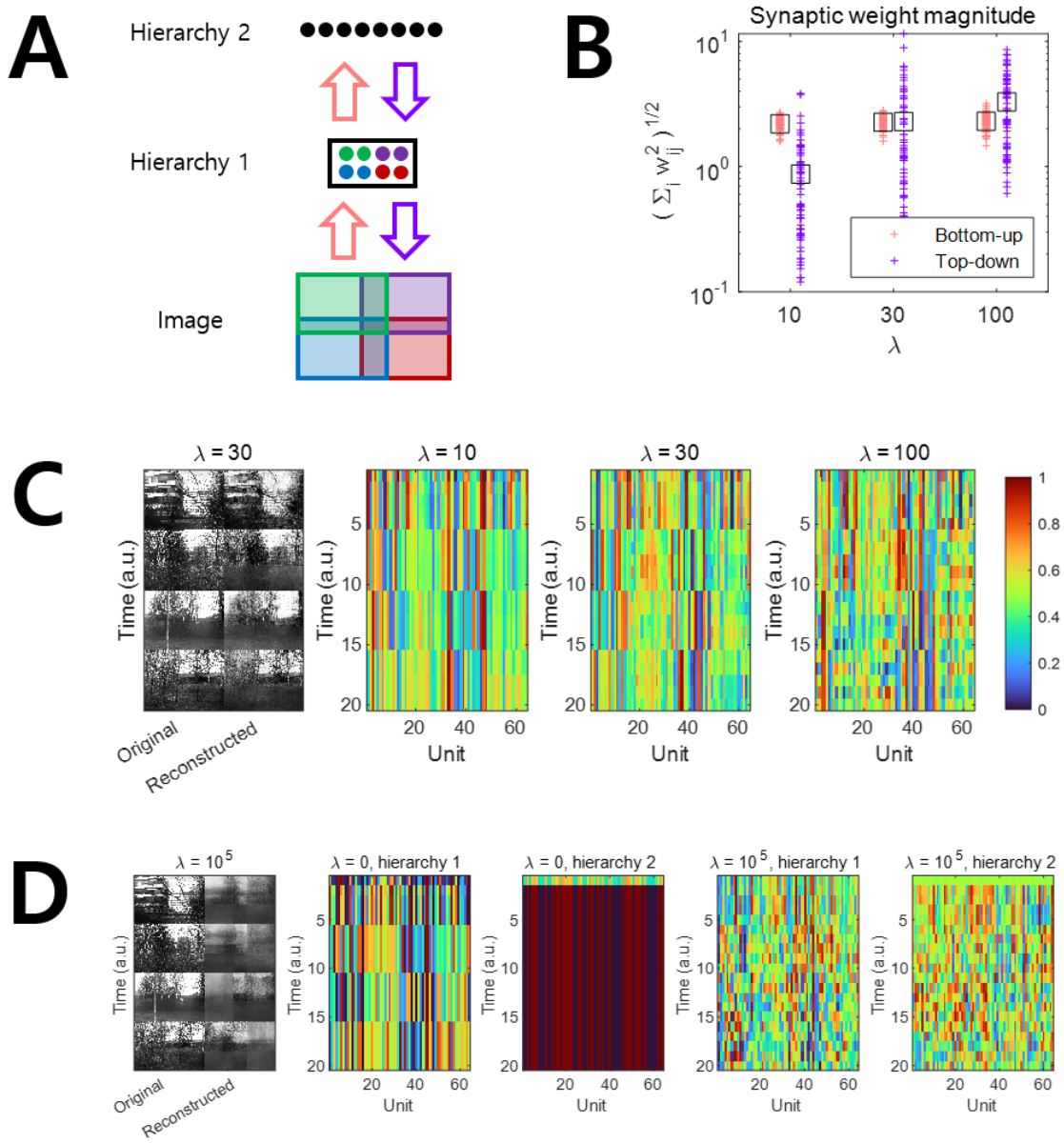


Fig 2. A simulation of spatio-temporally efficient coding. (A) Architecture of the simulation model. An input image is divided into four subspaces with overlaps, denoted by different colours. A subset of neuronal units in hierarchy 1 receive corresponding parts of the image (matching colours). There is no spatial correspondence between hierarchies 1 and 2. Pink and purple arrows indicate bottom-up and

top-down predictions, respectively. (B) The magnitude of synaptic weights on hierarchy 1 units is compared for different λ s, where λ is a balancing parameter in the objective function, $L = L_{Temporal} + \lambda L_{Spatial}$ (see Eq. 4). Each cross indicates the L2 norm of synaptic weights for each unit. Black squares indicate the mean. (C) Left: Four different original input images and their reconstructions from the neuronal responses of hierarchy 1. Right: Representative neural responses of the hierarchy 1 units for different λs. The neural response is the output of the sigmoid function and is therefore normalised to a range between 0 and 1. (D) Similar to (C), but with examples of λs that are too small or too large to balance between temporally and spatially efficient coding.

2.3. Appropriate neural representations of the external world

As a major function of the brain is to represent the external world (deCharms and Zador, 2000; Kriegeskorte and Diedrichsen, 2019), we investigated whether spatio-temporally efficient coding creates appropriate neural representations of the external world. We quantified the appropriateness of neural representations by examining the correlations between a natural scene image space and a neural response space. Specifically, we first measured the Euclidean distances from a natural scene image (i.e., reference image) to other natural scene images in the image space and the distances from the neural responses for the reference image to those for other compared images in the neural

response space (Fig 3A). When calculating the distances between neural responses, we used the neural responses at five time steps after receiving a given image based on the assumption that the neural responses became temporally steady after five steps of temporal processing (see Fig 1). We then measured the Pearson's linear correlation between the distances in the image space and those in the neural response space. A high correlation indicates that neural responses tend to be similar when the visual system perceives natural scenes close to each other (in the sense of Euclidean distance). We set each image sample as a reference and repeated the calculations for the correlation across all image samples. We defined global similarity as the average correlation obtained by using the distances from one image to all other images, and local similarity as the average correlation obtained by using the distances from one image to its neighbouring images only (1% of all images). For illustration, we visualised the distance from the reference image in Fig 3A to others by projecting the images onto a 2D space using t-distributed stochastic neighbour embedding (t-SNE) (van der Maaten and Hinton, 2008). We performed the same visualisation of the corresponding neural responses in each hierarchy. The distances in the 2D image space were largely preserved in the 2D neural response space (Fig 3B).

From the simulations, we identified high global similarities (> 0.6) between the natural scene image space and neural response space (Fig 3C). Global similarities were higher for the lower hierarchy than for the higher hierarchy, and increased with smaller λ for the lower hierarchy (Fig 3C). Local

similarities were close to 1 for all λs and hierarchies, demonstrating that the spatio-temporally efficient coding produced locally well-suited neural representations of natural scene images (Fig 3D).

Neural responses exhibited certain noise with spatio-temporally efficient coding, especially when λ was high (Fig 2C). We investigated whether this noise interfered with the neural representations of the visual images. To assess the interference by noise, we assumed that the neural responses converged to represent a given image after certain time steps of the bidirectional predictions. Here, we set the time step to 10 and measured the variability in neural responses from time step 1 (i.e., when receiving an input image) to time step 9 until convergence. Temporal variations were computed using the Euclidean distance from the neural responses at each time step (1, …, 9) to neural responses at time step 10. We also computed the Euclidean distance from the neural responses at time step 10 and those for an image that was the nearest to the given image in the image space. We compared the temporal variations above with this distance to the nearest neighbouring image by calculating the ratio of the temporal variations over the distance to the nearest neighbouring image. A ratio exceeding 1 indicates that the noise could significantly interfere with neural representations to be confused with nearby images. Within the range of our simulation parameters ($\lambda \in [10, 100]$), the ratio was kept below 1 on average for the majority of time steps at each hierarchy, even with higher values of λ (Fig 3E), indicating that the noise caused by maximising entropy with spatially efficient coding did not significantly interfere with neural presentations over time.

18

266

267

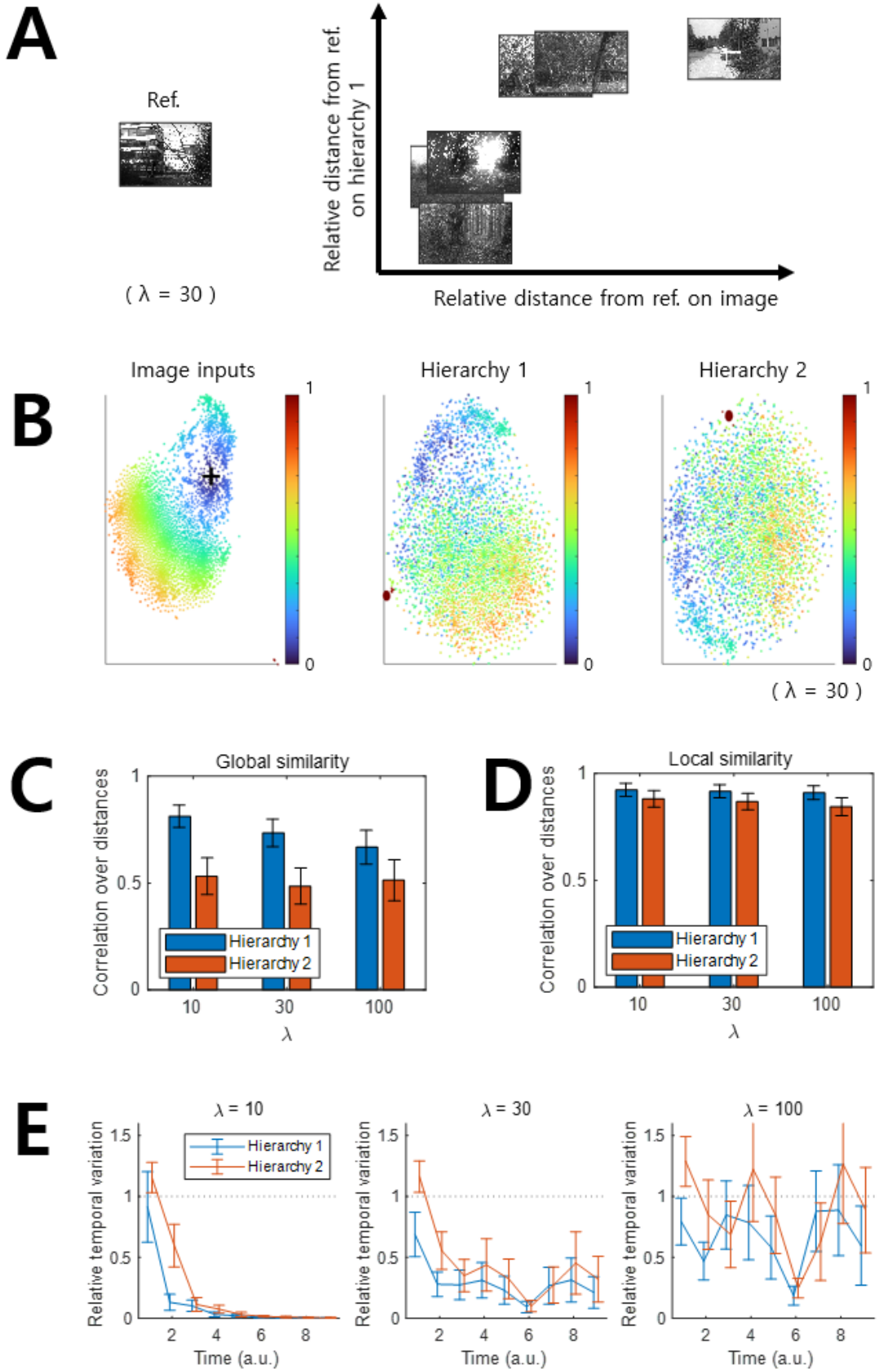

A
Ref.
(λ = 30)
Relative distance from ref. on hierarchy 1
Relative distance from ref. on image
B
Image inputs
Hierarchy 1
Hierarchy 2
1
0
(λ = 30)
C
Global similarity
Correlation over distances
Hierarchy 1
Hierarchy 2
10
30
100
λ
D
Local similarity
Correlation over distances
Hierarchy 1
Hierarchy 2
10
30
100
λ
E
λ = 10
λ = 30
λ = 100
Relative temporal variation
Hierarchy 1
Hierarchy 2
Time (a.u.)

Fig 3. Neural representations of visual images by spatio-temporally efficient coding. (A) Examples of relative Euclidean distances from the reference image (Ref. left) to other images in the natural scene image space are depicted on the horizontal axis. Relative Euclidean distances from the neural representations of the reference image at hierarchy 1 to those of corresponding other images are depicted on the vertical axis. (B) Left: Visualisation using t-distributed stochastic neighbour embedding (t-SNE). The colour code indicates normalised distances from the reference image in (A) to all other images in the natural scene image set, measured in the t-SNE-generated natural scene image space. The black cross indicates the reference image in (A). Middle: Visualisation of neural responses of images at hierarchy 1 using t-SNE. The colours of each neural response represent the normalised distance of the corresponding image measured in the left plot. Similar patterns of colour codes indicate that the similarity between images is well preserved in neural representations. Right: The same visualisation of neural responses at hierarchy 2. (C) Global similarity based on correlations of Euclidean distances between overall natural scene images (N = 4,212) with those of corresponding neural responses. Error bars indicate standard deviations. (D) Local similarity based on correlations of Euclidean distances between overall natural scene images (N = 42) with those of corresponding neural responses. (E) The ratio of the temporal variations of neural responses relative to the Euclidean distance from the nearest image (see text for details on the computation of temporal variations). The dotted lines denote the ratio of 1. Error bars indicate standard deviations.

2.4. Deviant neural responses to unfamiliar inputs

The visual system often responds selectively to sensory inputs (Margoliash, 1983; Waydo et al., 2006). Even for the type of sensory inputs to which the visual system is responsive, unfamiliar inputs induce larger neural responses compared to familiar inputs (Huang et al., 2018; Issa et al., 2018). These large neural responses to unfamiliar inputs are thought to be due to prediction errors (Issa et al., 2018). Accordingly, we investigated whether spatio-temporally efficient coding could predict the phenomenon of large neural responses to unfamiliar inputs without the introduction of prediction error responses mediated by error units.

In the simulations, the visual hierarchical structure learned to be familiar with natural scene images, and novel handwritten digit images were used as unfamiliar visual inputs (Fig 4A). The simulation results revealed that neural responses were distributed over middle values for familiar images and over smaller or larger values for unfamiliar inputs (Fig 4B and 4C), suggesting that spatio-temporally efficient coding could predict the phenomenon of deviant neural responses to unfamiliar inputs.

Further, we investigated the neural responses to locally unfamiliar and globally familiar images. We composed new images such that a portion of only one quadrant of the image was unfamiliar, and the rest was familiar (Fig 5A). According to global precedence (Navon, 1977; Rezvani et al., 2020), such

partially unfamiliar images are preferentially perceived as familiar ones. The simulation results demonstrated that deviant neural responses were predominantly localised to the hierarchy 1 units connected to unfamiliar portions of the images (Fig 5B). In addition, deviant neural responses rarely propagated to hierarchy 2 units (Fig 5C) and thus did not perturb hierarchy 1 units connected to familiar inputs in return (Fig 5D and 5E), consistent with the predictions of global precedence.

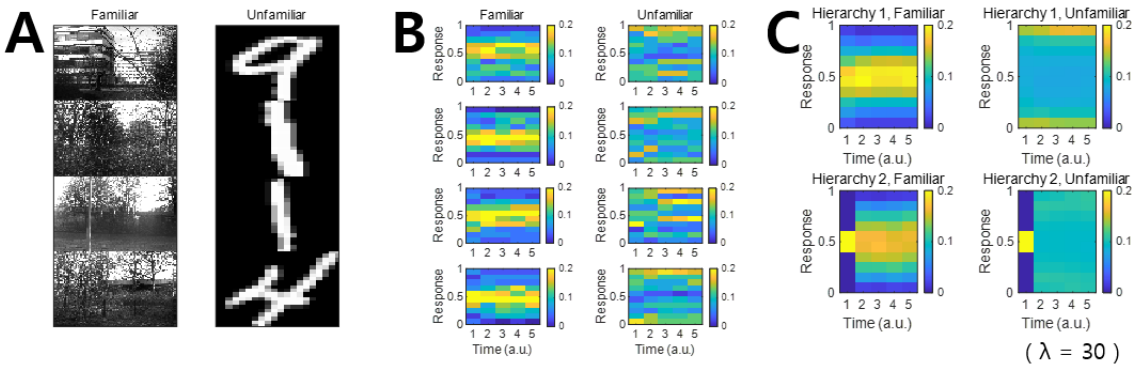


Fig 4. Neural response distributions for familiar and unfamiliar inputs. (A) Natural scene images are used as familiar inputs (for learning in the visual hierarchical structure) and handwritten digit images are used as unfamiliar inputs (not used in learning). (B) Examples of neural response distributions of hierarchy 1 units for each of the five time steps of bidirectional predictions, corresponding to images in (A). The colour scale indicates the proportion. (C) Overall neural response distributions in response to all input images (either familiar or unfamiliar) at hierarchies 1 and 2, respectively.

23

318

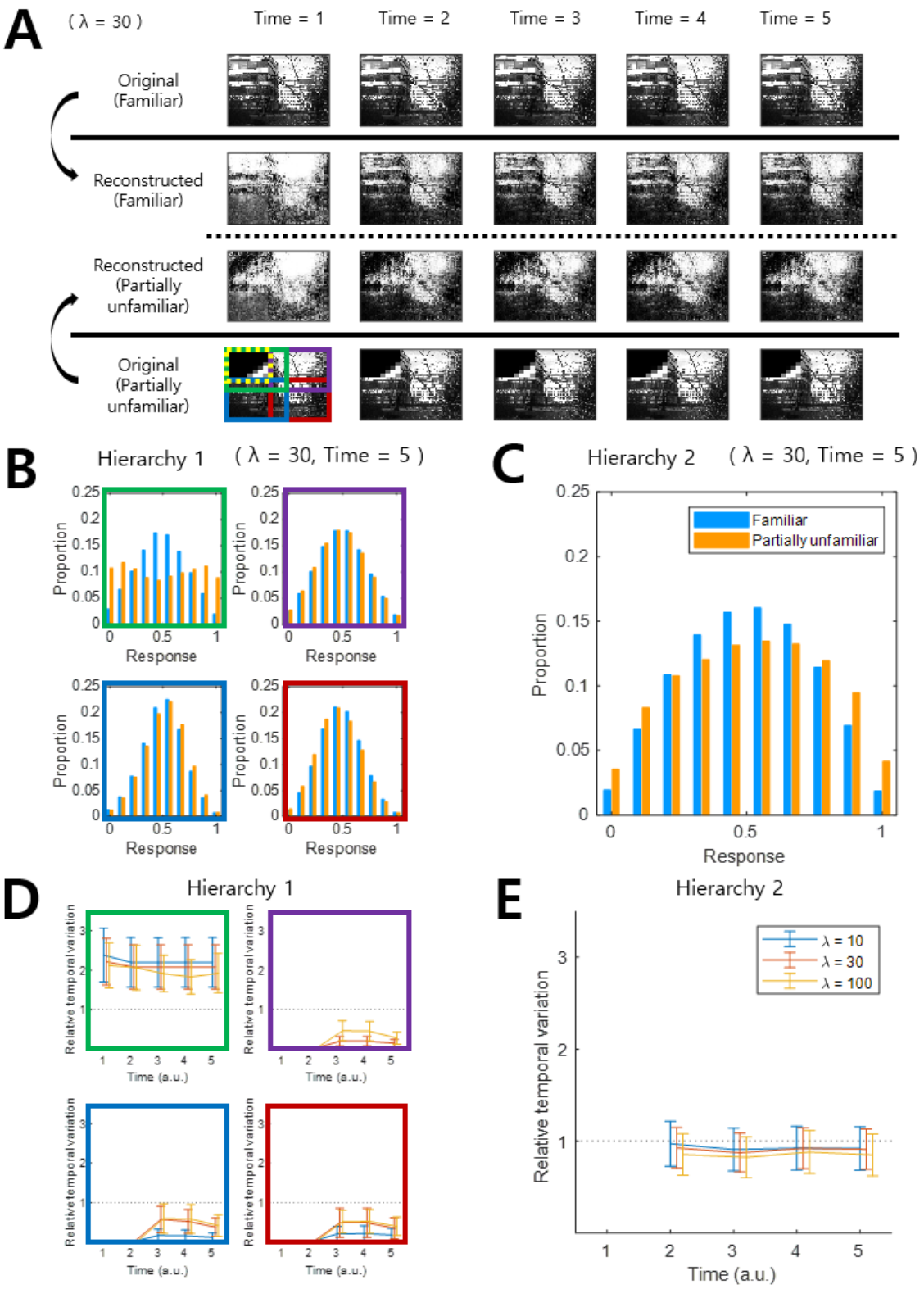

A
(λ = 30)
Time = 1
Time = 2
Time = 3
Time = 4
Time = 5
Original (Familiar)
Reconstructed (Familiar)
Reconstructed (Partially unfamiliar)
Original (Partially unfamiliar)
B
Hierarchy 1
(λ = 30, Time = 5)
Proportion
Response
C
Hierarchy 2
(λ = 30, Time = 5)
Familiar
Partially unfamiliar
Proportion
Response
D
Hierarchy 1
Relative temporal variation
Time (a.u.)
E
Hierarchy 2
λ = 10
λ = 30
λ = 100
Relative temporal variation
Time (a.u.)

Fig 5. Neural response distributions for partially unfamiliar inputs. (A) Image reconstruction from neural responses over time. Using spatio-temporally efficient coding of learned natural scene images, the visual hierarchical structure reconstructs a familiar image well from time steps 2 to 5 (top). The same hierarchical structure also reconstructs a partially familiar image well, with the exception of a part of the image, marked by a yellow dotted line (bottom). (B) Neural response distributions of hierarchy 1 units corresponding to each of the four parts of the image. Box colours match with the image parts in (A). The green box contains unfamiliar input. (C) Neural response distributions of hierarchy 2 units. Note that there is no spatial segregation of hierarchy 2 units corresponding to the parts of the image (see Fig 2A). (D, E) The ratios of the distance (relative temporal variation) between neural responses for familiar and partially unfamiliar inputs relative to the average distance from an image to neighbouring images (N = 42) at hierarchies 1 (D) and 2 (E), respectively. The scale bars indicate the standard deviation. The dotted horizontal line indicates a ratio of 1.

2.5. Preferred orientation biases of receptive fields

Neurons in the visual system prefer horizontal and vertical orientations over oblique orientations (Furmanski and Engel, 2000; Li et al., 2003). Indeed, orientation discrimination is more sensitive to horizontal and vertical orientations than to oblique orientations (Girshick et al., 2011). These biases

are due to the environmental statistics of natural scenes (Girshick et al., 2011). We investigated whether hierarchy 1 units in the visual hierarchical structure that learned by spatio-temporally efficient coding of natural scene images exhibited such biases. We presented a moving bar oriented in one of eight angles that moved in the direction perpendicular to the orientation angle (Fig 6A) and defined the response of each unit to that orientation by the largest response during presentation. We then defined the preferred orientation of each unit as the orientation that elicited the largest response. Simulation results revealed that hierarchy 1 units preferred horizontal and vertical orientations over oblique orientations in all λ conditions (Fig 6B), consistent with the orientation bias of visual cortical neurons and in accordance with context-independent bottom-up prediction (Teufel and Fletcher, 2020).

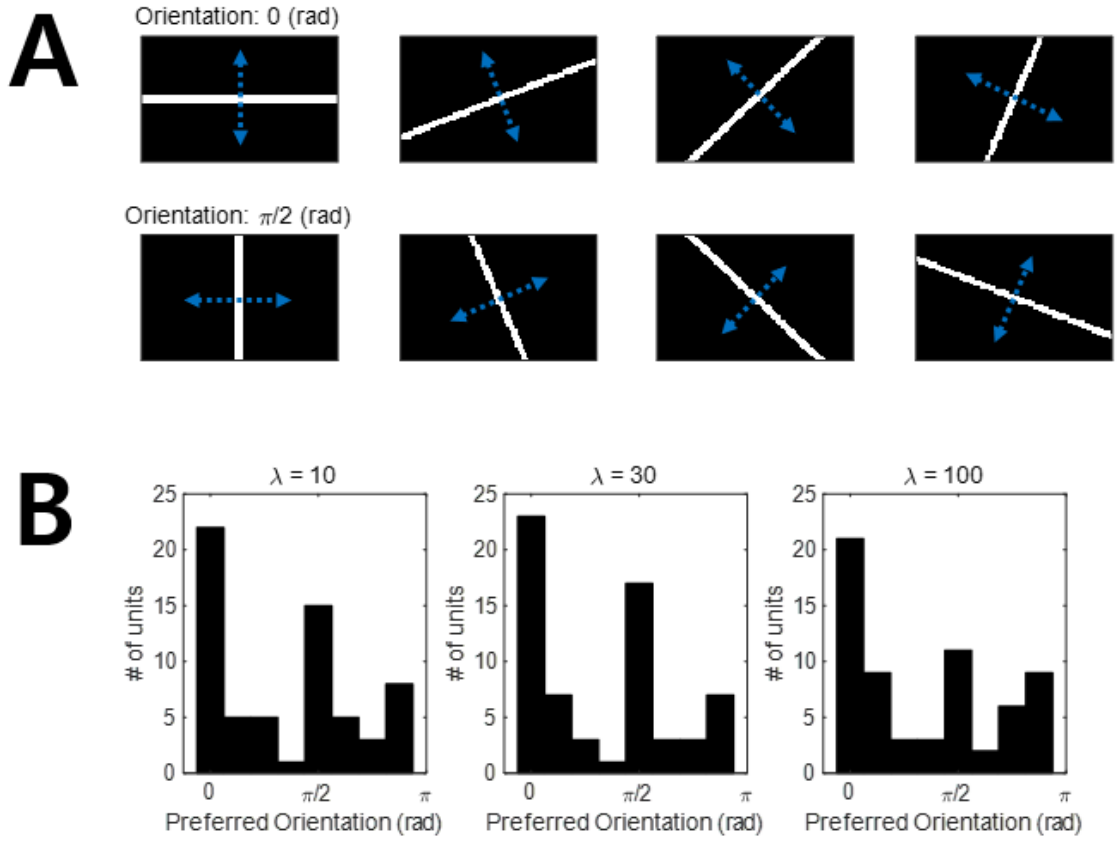


Fig 6. Orientation preference of hierarchy 1 units. (A) Orientation images used to test the visual orientation preference of neuronal units at hierarchy 1 in the visual hierarchical structure that learned by the spatio-temporally efficient coding of natural scene images. White bars in each of the eight orientations are moved in the perpendicular directions denoted by blue dotted arrows. (B) Histograms of the orientation preference of hierarchy 1 units at time step 5 for different values of λ.

3. Discussion

The present study aimed to find computational principles that enables visual hierarchical structures to attain the function to represent external visual information. To address the lack of neural coding principles to encompass both bottom-up and top-down pathways, we propose spatio-temporally efficient coding as a novel computational model. As a principled way of efficiently using given resources in both neural activity space and processing time, this coding principle optimises bidirectional predictions over hierarchical structures by simultaneously minimising temporally differences in neural responses and maximising entropy in neural representations. Simulation results showed that the proposed spatio-temporally efficient coding assigned the function of appropriate neural representations of natural scenes to a visual hierarchical structure and that it could predict deviations in neural responses to unfamiliar inputs and a bias in preferred orientations, which are well known characteristics of the visual system.

Cortical Hebbian-like spike-timing-dependent plasticity (STDP) (Feldman, 2000; Froemke and Dan, 2002) is a potential mechanism underlying the temporally efficient coding proposed in this study. If bidirectional predictions of the hierarchical structure increase future postsynaptic responses, STDP will strengthen these synaptic connections, which may increase future postsynaptic responses. Indeed, this phenomenon is underscored by a minimisation of the temporal difference in neural responses, rendering near future neural responses closer to distant future neural responses. As such, the concept of temporally efficient coding proposed herein is consistent with cortical Hebbian-like

STDP.

Since its initial proposal (Attneave, 1954; Barlow, 1961), spatially efficient coding has been validated experimentally (Laughlin, 1981). However, observed correlations between neurons, which maximise entropy to a lesser extent compared to mere spatially efficient coding assuming no inter-neuronal correlations, have yet to be incorporated into the principle of spatially efficient coding. Empirically observed neuronal correlations may drive computational processes of the brain away from strict spatially efficient coding. Recent studies suggest that biological visual systems are intermediate between strict spatially efficient coding and correlated neural responses (Stringer et al., 2019). Therefore, to create biologically plausible computational models, it is necessary to mitigate the spatially efficient coding objective by combining firing-rate-dependent correlations (de la Rocha et al., 2007). This enables more accurate predictions of visual perception mediated by visual hierarchical structures. As we focused on integrating spatially efficient coding with temporally efficient coding for computation in hierarchical structures, this study did not incorporate the correlations between neurons in spatially efficient coding, which will be pursued in follow-up studies.

Based on our simulations, we observed that the learning of bidirectional prediction networks with spatio-temporally efficient coding was hindered when the balancing parameter λ was too small or too large (Fig 2D). Therefore, it was necessary to confine λ within a certain range, in which the magnitude of λ affected neural responses such that a larger λ rendered responses more variable

(Fig 2C). Such increased variability is likely to originate from recurrent responses via higher hierarchies rather than the noise of neural responses. This was confirmed by the observation that top-down synaptic weights become larger than bottom-up synaptic weights when λ increased (Fig 2B). Although a large λ attenuated the appropriateness of neural representations (Fig 3), it rendered stronger top-down synaptic connections in hierarchy 1 (Fig 2B), which is consistent with the previous finding that top-down synaptic connections are stronger than bottom-up connections in the lateral geniculate nucleus (Sillito et al., 2006). As to why top-down synaptic weights increase with a larger λ value (Fig 2B), we speculate that learning via spatio-temporally efficient coding may increase the range of neural responses to maximise entropy through top-down pathways. While the bottom-up pathways originating from external inputs are invariant during learning, the top-down pathways originating from higher hierarchy neural responses are more flexible to adjustment to maximise entropy during learning.

The deviant neural responses to unfamiliar inputs observed in this study (Fig 4 and 5) arise from compact, well-organised neural representations for familiar inputs (distributed over the middle value). As such, neural representations for familiar inputs extrude neural responses to unfamiliar inputs into a range of deviant neural responses. We conjecture that the visual system may generate deviant neural responses via a similar mechanism. Simulations in the present study demonstrated that neural responses were distributed around smaller or larger extremes for unfamiliar inputs and around

411 intermediate values for familiar inputs (Fig 4B and 4C). In a separate analysis, we allowed the neural
412 responses to familiar inputs be distributed only around lower values, similar to sparse coding
413 (Olshausen and Field, 1996; Olshausen and Field, 1997), and observed that neural responses to
414 unfamiliar inputs only exhibited higher values.

415 Spatio-temporally efficient coding predicted a bias in preferred orientations (Fig 6). In this regard,
416 spatially efficient coding alone has been reported to predict bias in preferred orientations (Ganguli and
417 Simoncelli, 2014). Notably, spatio-temporally efficient coding was able to predict this bias well, even
418 when λ was low, that is, when the spatially efficient coding objective was less weighted (Fig 6B).
419 Therefore, this bias prediction should be viewed as a result of spatio-temporally efficient coding, not
420 as a result of spatially efficient coding alone.

421 The present study has several limitations. First, for simplicity, our simulation model contained only two
422 hierarchies. However, it is necessary to explore how spatio-temporally efficient coding operates in
423 models with more hierarchies. We also modelled 64 neuronal units at each hierarchy, as we assumed
424 that this would be sufficient to represent the natural scene images used in this study. Nevertheless,
425 the interactions between the number of neuronal units, levels of hierarchy, and spatio-temporally
426 efficient coding require further investigation. Second, we demonstrated that the visual hierarchical
427 structure could learn to represent static natural scene images with spatio-temporally efficient coding,
428 but future follow-up studies will investigate whether the visual hierarchical structure learns to

429 represent moving scenes using the same coding principle. Finally, the scope of the present study was
430 limited to the visual system given that its hierarchical structure is well documented, but spatio-
431 temporally efficient coding may be applied to other systems (e.g., somatosensory system) or to
432 movements and planning.

433

434 3.1. Conclusions

435 In the present study, we proposed spatio-temporally efficient coding, inspired by the efficient use of
436 given resources in neural systems, as a neural coding mechanism to assign representational
437 functions to the hierarchical structures of the visual system. Simulations demonstrated that the visual
438 hierarchical structure could represent the external world (i.e., natural scenes) appropriately using
439 bidirectional predictions (Fig 3). Furthermore, spatio-temporally efficient coding predicted the well-
440 known properties of visual cortical neurons, including deviations in neural responses to unfamiliar
441 images (Fig 4 and 5) and bias in preferred orientations (Fig 6). Our proposed spatio-temporally
442 efficient coding may facilitate deeper mechanistic understanding of the computational processes of
443 hierarchical brain structures.

444

4. Materials and Methods

4.1. Datasets

For the simulations, van Hateren's natural scene image dataset (van Hateren and van der Schaaf, 1998) was used. The dataset was downloaded from http://bethgelab.org/datasets/vanhateren/. The images were downsized to 64×96 pixels. For the comparison tests, the MNIST handwritten digit dataset (Lecun et al., 1998) was used. The dataset was downloaded from http://yann.lecun.com/exdb/mnist/. The images were resized to 64×96 pixels to fit the images used in the simulations. All image data were rescaled between 0 and 1.

4.2. Detailed descriptions of spatio-temporally efficient coding

For convenience, $X_{h=0}$ denotes X_{image}. The details of the visual information processing f_t are as follows: If $h > 0$, then

$$f_t|_{X_{h,t}} = \sigma\left(W_{h+1,h}^T X_{h+1,t-1} + W_{h-1,h}^T X_{h-1,t-1} + b_h\right) \quad (5)$$

where $X_{h,t}$ is an X_h value vector at time t, $f_t(\cdot)|_{X_{h,t}}$ indicates restricting the range of the function

value $f_t(\cdot)$ to $X_{h,t}$, $W_{h+1,h}$ is a synaptic weight matrix from hierarchy $h+1$ to h, T is the transpose of a matrix, b_h is a bias vector at hierarchy h, and $\sigma(\cdot)$ is a sigmoid function. In the case of $h+1>H$, the $W_{h+1,h}^T X_{h+1,t-1}$ term should be omitted. If $h=0$, then

$$f_t|_{X_{0,t}} = F\left(X_{h=1,t-1}\right) \tag{6}$$

where $F(\cdot)$ is a neural network with two hidden layers to enable detailed reconstruction of images. The activation function of the last layer of $F(\cdot)$ is a sigmoid function, as in the case of $h>0$, in the hidden layer, it is a rectified linear unit.

To minimise the objective of spatially efficient coding, it is necessary to calculate the probability $P\left(f_t|_{X_h}\right)$. Instead of calculating the exact probabilities, we obtained pre-normalised densities in the sense of probabilities without a partition function. As the value of the partition function is fixed, it does not affect the minimisation process. Kernel density estimation was used to obtain $P\left(f_t|_{X_h}\right)$. Using a Gaussian kernel with width $0.1(\dim X_h)^{1/2}$, the neural response density $Q\left(f_t|_{X_h}\right)$ and compensation density $Q'\left(f_t|_{X_h}\right)$ at $f_t|_{X_h} \in X_h$ were obtained. Then, the pre-normalised density of interest is

$$P\left(f_t|_{X_h}\right) = \frac{Q\left(f_t|_{X_h}\right)}{Q'\left(f_t|_{X_h}\right)}. \tag{7}$$

The compensation density $Q'\left(f_t|_{X_h}\right)$ is obtained using pseudo-uniformly generated samples on $X_h = [0,1]^{\dim X_h}$ instead of the neural responses. The compensation density is necessary to compensate for the non-uniform intrinsic expectation of $Q(\cdot)$ resulting from the fact that X_h is

bounded.

In the present study, the depth of the hierarchies was 2, and images with 64×96 size were divided into four overlapping 38×58 patches. Each patch was connected to 16 of 64 hierarchy 1 units. Further, 64 hierarchy 1 units were fully connected to 64 hierarchy 2 units (Fig 1A). Four neural networks $F(\cdot)$ were connected from hierarchy 1 units to four image patches. Each neural network $F(\cdot)$ had two hidden layers with 128 and 1024 units, respectively.

Because $L = L_{Temporal} + \lambda L_{Spatial}$ is differentiable, the minimisation of the objective in spatio-temporally efficient coding was performed with a gradient descent. The Adam optimiser (Kingma and Ba, 2015) was used to perform the stochastic gradient descent with momentum. The parameters of the Adam optimiser used in this study were $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The optimisation lasted 10^4 iterations for each repetition and restarted with five repetitions. For each iteration, the duration of temporal processing f_t was five (i.e., $t \in [1,5]$), and the minibatch size was 40.

Simulation codes for spatio-temporally efficient coding are available from https://github.com/DuhoSihn/Spatio-temporally-efficient-coding (Sihn, 2021).

Data availability statement

All simulation and analysis codes are available at: https://github.com/DuhoSihn/Spatio-temporally-efficient-coding and on Zenodo (doi: 10.5281/zenodo.5298182)

CRediT authorship contribution statement

Duho Sihn: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration. **Sung-Phil Kim**: Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Acknowledgments

References

1. Attneave F. Some informational aspects of visual perception. Psychological Review 1954; 61(3):183–193. doi: 10.1037/h0054663

2. Barlow HB. Possible principles underlying the transformations of sensory messages. In: Rosenblith WA editor. Sensory Communication, MIT Press; 1961. p 217-234. doi: 10.7551/mitpress/9780262518420.003.0013

3. Berkes P, Wiskott L. Slow feature analysis yields a rich repertoire of complex cell properties. Journal of Vision 2005 Jul 20; 5(6):579-602. doi: 10.1167/5.6.9

4. Berry MJ, Brivanlou IH, Jordan TA, Meister M. Anticipation of moving stimuli by the retina. Nature 1999 Mar 01; 398(6725):334-338. doi: 10.1038/18678

5. Boutin V, Franciosini A, Chavane F, Ruffier F, Perrinet L, Sparse deep predictive coding captures contour integration capabilities of the early visual system. PLoS Comput Biol 2021 Jan 26; 17(1): e1008629. doi: 10.1371/journal.pcbi.1008629

6. Bullmore E, Sporns O. The economy of brain network organization. Nat Rev Neurosci 2012

May 01; 13(5)336–349. doi: 10.1038/nrn3214

7. Burt JB, Demirtaş M, Eckner WJ, Navejar NM, Ji JL, Martin WJ, Bernacchia A, Anticevic A, Murray JD. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. Nat Neurosci 2018 Sep 01; 21(9):1251–1259. doi: 10.1038/s41593-018-0195-0

8. Chen KS, Chen C-C, Chan CK. Characterization of predictive behavior of a retina by mutual information. Front. Comput. Neurosci., 2017 Jul 20; 11:66. doi: 10.3389/fncom.2017.00066

9. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences 2013 May 10; 36(3):181-204. doi: 10.1017/S0140525X12000477

10. Creutzig F, Sprekeler H. Predictive coding and the slowness principle: An information-theoretic approach. Neural Computation 2008 Apr; 20(4):1026-1041. doi: 10.1162/neco.2008.01-07-455

11. de la Rocha J, Doiron B, Shea-Brown E, Josic K, Reyes A. Correlation between neural spike trains increases with firing rate. Nature 2007 Aug 01; 448(7155):802–806. doi: 10.1038/nature06028

12. deCharms RC, Zador A. Neural representation and the cortical code. Annu. Rev. Neurosci.

541 2000 May 01; 23(1):613-647. doi: 10.1146/annurev.neuro.23.1.613

542 13. DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition?
543 Neuron 2012 Feb 09; 73(3):415-434. doi: 10.1016/j.neuron.2012.01.010

544 14. Feldman DE. Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat
545 barrel cortex. Neuron 2000 Jul 01; 27(1):45-56. doi: 10.1016/S0896-6273(00)00008-8

546 15. Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral
547 cortex. Cereb Cortex, 1991 Jan 01; 1(1):1-47. doi: 10.1093/cercor/1.1.1

548 16. Friston K. A theory of cortical responses. Phil. Trans. R. Soc. B 2005 Apr 29; 360(1456):815-
549 836. doi: 10.1098/rstb.2005.1622

550 17. Friston K. The free-energy principle: a unified brain theory? Nat Rev Neurosci 2010 Feb 01;
551 11(2):127–138. doi: 10.1038/nrn2787

552 18. Furmanski CS, Engel SA. An oblique effect in human primary visual cortex. Nat Neurosci
553 2000 Jun 01; 3(6):535–536. doi: 10.1038/75702

554 19. Froemke RC, Dan Y. Spike-timing-dependent synaptic modification induced by natural spike
555 trains. Nature 2002 Mar 01; 416(6879):433–438. doi: 10.1038/416433a

556 20. Ganguli D, Simoncelli EP. Efficient sensory encoding and bayesian inference with
557 heterogeneous neural populations. Neural Comput 2014 Oct 01; 26(10):2103–2134. doi:

558 10.1162/NECO_a_00638

559 21. Girshick AR, Landy MS, Simoncelli EP. Cardinal rules: visual orientation perception reflects
560 knowledge of environmental statistics. Nat Neurosci 2011 Jul 01; 14(7):926–932. doi:
561 10.1038/nn.2831

562 22. Griffiths TL, Lieder F, Goodman ND. Rational use of cognitive resources: Levels of analysis
563 between the computational and the algorithmic. Top Cogn Sci 2015 Apr 01; 7(2):217-229.
564 doi: 10.1111/tops.12142

565 23. Hansen JY, Markello RD, Vogel JW, Seidlitz J, Bzdok D, Misic B. Mapping gene transcription
566 and neurocognition across human neocortex. Nat Hum Behav 2021 Mar 25. doi:
567 10.1038/s41562-021-01082-z

568 24. Harris JA, Mihalas S, Hirokawa KE, Whitesell JD, Choi H, Bernard A, Bohn P, Caldejon S,
569 Casal L, Cho A, Feiner A, Feng D, Gaudreault N, Gerfen CR, Graddis N, Groblewski PA,
570 Henry AM, Ho A, Howard R, Knox JE, Kuan L, Kuang X, Lecoq J, Lesnar P, Li Y, Luviano J,
571 McConoughey S, Mortrud MT, Naeemi M, Ng L, Oh SW, Ouellette B, Shen E, Sorensen SA,
572 Wakeman W, Wang Q, Wang Y, Williford A, Phillips JW, Jones AR, Koch C, Zeng H.
573 Hierarchical organization of cortical and thalamic connectivity. Nature 2019 Nov 01;
574 575(7781):195–202. doi: 10.1038/s41586-019-1716-z

575 25. Hilgetag CC, Goulas A. ‘Hierarchy’ in the organization of brain networks. Phil. Trans. R. Soc.
576 B 2020 Apr 13; 375(1796):20190319. doi: 10.1098/rstb.2019.0319

577 26. Huang G, Ramachandran S, Lee TS, Olson CR. Neural correlate of visual familiarity in
578 macaque area V2. J. Neurosci. 2018 Oct 17; 38(42):8967-8975. doi:
579 10.1523/JNEUROSCI.0664-18.2018

580 27. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in
581 the cat's visual cortex. The Journal of Physiology 1962 Jan 01; 160(1):106-154. doi:
582 10.1113/jphysiol.1962.sp006837

583 28. Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex.
584 The Journal of Physiology 1968 Mar 01; 195(1):215-243. doi:
585 10.1113/jphysiol.1968.sp008455

586 29. Huh CYL, Peach JP, Bennett C, Vega RM, Hestrin S. Feature-Specific Organization of
587 Feedback Pathways in Mouse Visual Cortex. Current Biology 2018 Jan 08; 28(1):114-120.e5.
588 doi: 10.1016/j.cub.2017.11.056

589 30. Issa EB, Cadieu CF, DiCarlo JJ. Neural dynamics at successive stages of the ventral visual
590 stream are consistent with hierarchical error signals. eLife 2018 Nov 28; 7:e42870. doi:
591 10.7554/eLife.42870

31. Kingma DP, Ba J. Adam: A method for stochastic optimization. Proceedings of the 3th International Conference on Learning Representations; 2015 May 7-9; San Diego, California, United States.

32. Kriegeskorte N, Diedrichsen J. Peeling the onion of brain representations. Annu. Rev. Neurosci. 2019 Jul 08; 42(1):407-432. doi: 10.1146/annurev-neuro-080317-061906

33. Laughlin SB. A simple coding procedure enhances a neuron's information capacity. Z. Naturforsch. 1981; 36(9-10):910-912 doi: 10.1515/znc-1981-9-1040

34. Laughlin SB. Energy as a constraint on the coding and processing of sensory information. Current Opinion in Neurobiology 2001 Aug 01; 11(4):475-480. doi: 10.1016/S0959-4388(00)00237-3

35. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE 1998 Nov; 86(11):2278-2324. doi: 10.1109/5.726791

36. Li B, Peterson MR, Freeman RD. Oblique Effect: A neural basis in the visual cortex. J. Neurophysiol. 2003 Jul 01; 90(1):204-217. doi: 10.1152/jn.00954.2002

37. Lieder F, Griffiths TL. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. Behavioral and Brain Sciences 2020; 43:E1. doi: 10.1017/S0140525X1900061X

609 38. Liu B, Hong A, Rieke F, Manookin MB. Predictive encoding of motion begins in the primate
610 retina. Nature Neuroscience 2021 Aug 02. doi: 10.1038/s41593-021-00899-1

611 39. Margoliash D. Acoustic parameters underlying the responses of song-specific neurons in the
612 white-crowned sparrow. J. Neurosci. 1983 May 01; 3(5):1039-1057. doi:
613 10.1523/JNEUROSCI.03-05-01039.1983

614 40. Mesulam MM, From sensation to cognition. Brain, 1998 Jun 01; 121(6):1013–1052. doi:
615 10.1093/brain/121.6.1013

616 41. Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, Yacoub E.
617 Contextual feedback to superficial layers of V1. Current Biology 2015 Oct 19, 25(20):2690-
618 2695. doi: 10.1016/j.cub.2015.08.057

619 42. Murphy PC, Sillito AM. Corticofugal feedback influences the generation of length tuning in
620 the visual pathway. Nature 1987 Oct 01; 392(6141):727-729. doi: 10.1038/329727a0

621 43. Navon D. Forest before trees: The precedence of global features in visual perception.
622 Cognitive Psychology 1977 Jul 01; 9(3):353-383. doi: 10.1016/0010-0285(77)90012-3

623 44. Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a
624 sparse code for natural images. Nature 1996 Jun 01; 381(6583):607–609. doi:
625 10.1038/381607a0

626 45. Palmer SE, Marre O, Berry MJ, Bialek W. Predictive information in a sensory population.
627 Proc Natl Acad Sci USA 2015 Jun 02; 112(22):6908-6913. doi: 10.1073/pnas.1506855112

628 46. Rao RPN, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of
629 some extra-classical receptive-field effects. Nat Neurosci 1999 Jan 01; 2(1):79-87. doi:
630 10.1038/4580

631 47. Rezvani Z, Katanforoush A, Pouretemad H. Global precedence changes by environment: A
632 systematic review and meta-analysis on effect of perceptual field variables on global-local
633 visual processing. Atten Percept Psychophys. 2020 Jul; 82(5):2348-2359. doi:
634 10.3758/s13414-020-01997-1

635 48. Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nat Neurosci
636 1999 Nov 01; 2(11):1019-1025. doi: 10.1038/14819

637 49. Riesenhuber M, Poggio T. Models of object recognition. Nat Neurosci 2000 Nov 01;
638 3(11):1199-1204. doi: 10.1038/81479

639 50. Sederberg AJ, MacLean JN, Palmer SE. Learning to make external sensory stimulus
640 predictions using internal correlations in populations of neurons. Proc Natl Acad Sci USA
641 2018 Jan 30; 115(5):1105-1110. doi: 10.1073/pnas.1710779115

642 51. Serre T, Oliva A, Poggio T. A feedforward architecture accounts for rapid categorization.

Proc Natl Acad Sci USA 2007 Apr 10; 104(15):6424-6429. doi: 10.1073/pnas.0700622104

52. Sihn D. Simulation codes for spatio-temporally efficient coding. 2021. Available from: https://github.com/DuhoSihn/Spatio-temporally-efficient-coding doi: 10.5281/zenodo.5298182

53. Sillito AM, Cudeiro J, J HE. Always returning: feedback and sensory processing in visual cortex and thalamus. Trends Neurosci. 2006 Jun 01; 29(6):307-316. doi: 10.1016/j.tins.2006.05.001

54. Solomon SS, Tang H, Sussman E, Kohn A. Limited evidence for sensory prediction error responses in visual cortex of macaques and humans. Cereb Cortex 2021 Jun 01; 31(6):3136-3152. doi: 10.1093/cercor/bhab014

55. Spratling MW. A review of predictive coding algorithms. Brain and Cognition 2017 Mar 01; 112:92-97. doi: 10.1016/j.bandc.2015.11.003

56. Stringer C, Pachitariu M, Steinmetz N, Carandini M, Harris KD. High-dimensional geometry of population responses in visual cortex. Nature 2019 Jul 01; 571(7765)361–365. doi: 10.1038/s41586-019-1346-5

57. Teufel C, Fletcher PC. Forms of prediction in the nervous system. Nat Rev Neurosci 2020 Apr 01; 21(4):231–242. doi: 10.1038/s41583-020-0275-5

660 58. Olshausen BA, Field DJ. Sparse coding with an overcomplete basis set: A strategy
661 employed by V1? Vision Research 1997 Dec 01; 37(23):3311-3325. doi: 10.1016/S0042-
662 6989(97)00169-7

663 59. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008 Nov;
664 9(86):2579−2605.

665 60. van Hateren JH, van der Schaaf A. Independent component filters of natural images
666 compared with simple cells in primary visual cortex. Proc. Royal Soc. B 1998 Mar 07;
667 265(1394):359-366. doi: 10.1098/rspb.1998.0303

668 61. Wang W, Jones HE, Andolina IM, Salt TE, Sillito AM. Functional alignment of feedback
669 effects from visual cortex to thalamus. Nat Neurosci 2006 Oct 01; 9(10):1330-1336. doi:
670 10.1038/nn1768

671 62. Waydo S, Kraskov A, Quian Quiroga R, Fried I, Koch C. Sparse representation in the human
672 medial temporal lobe. J. Neurosci. 2006 Oct 04; 26(40):10232-10234. doi:
673 10.1523/JNEUROSCI.2101-06.2006

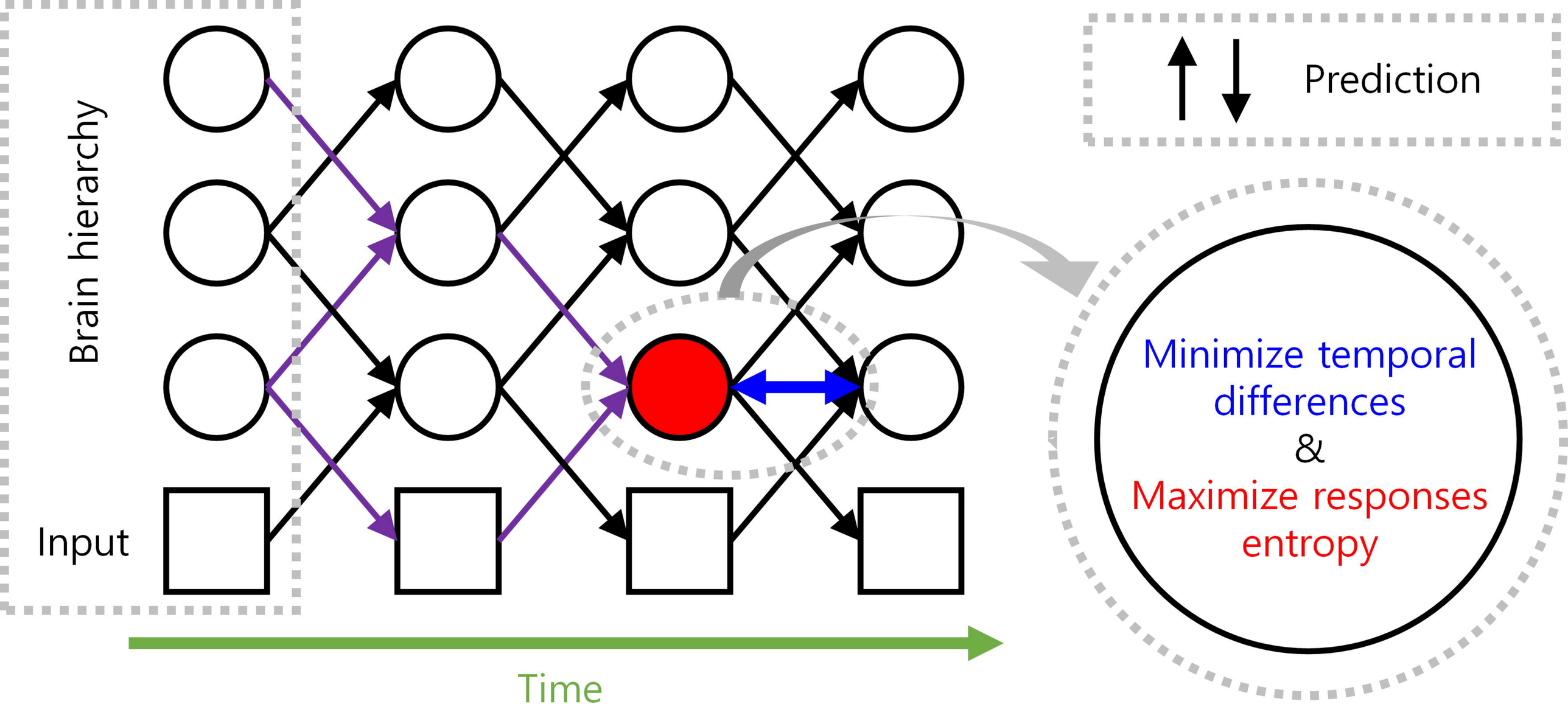
674 63. Wiskott L, Sejnowski TJ. Slow feature analysis: Unsupervised learning of invariances. Neural
675 Computation 2002 Apr 01; 14(4):715-770. doi: 10.1162/089976602317318938

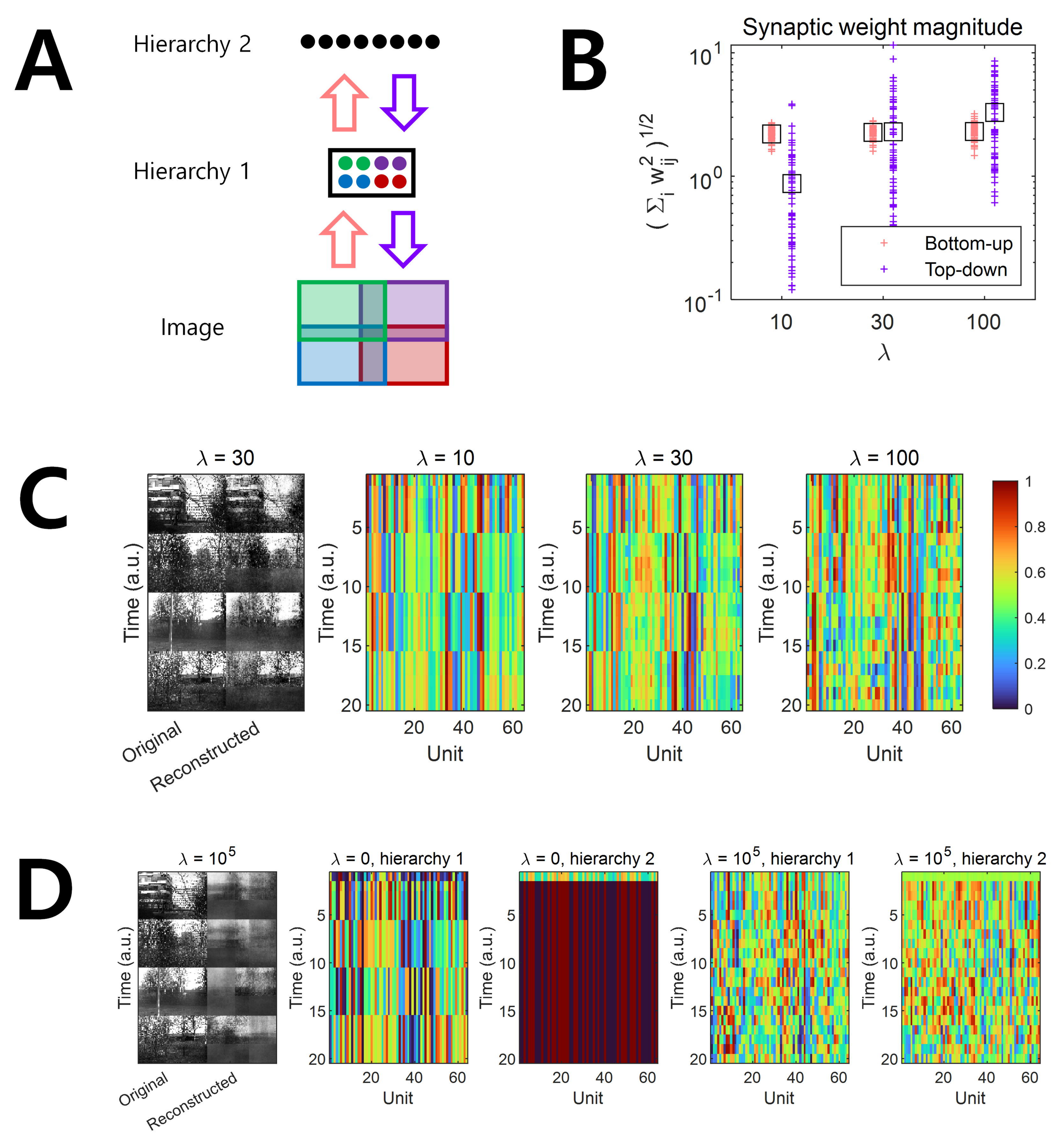
676 64. Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. Performance-

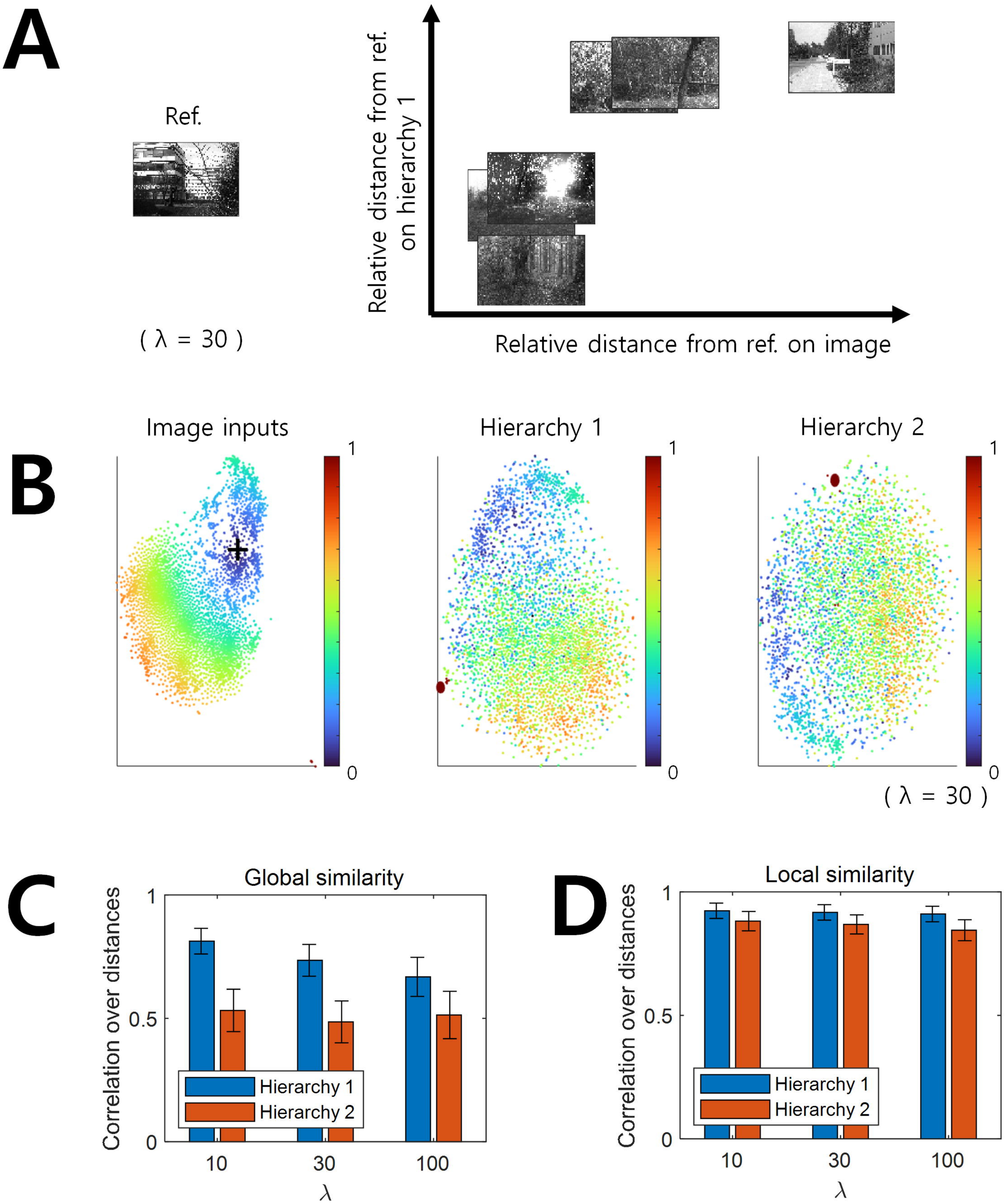
677 optimized hierarchical models predict neural responses in higher visual cortex. Proc Natl
678 Acad Sci USA 2014 Jun 10; 111(23):8619-8624. doi: 10.1073/pnas.1403112111

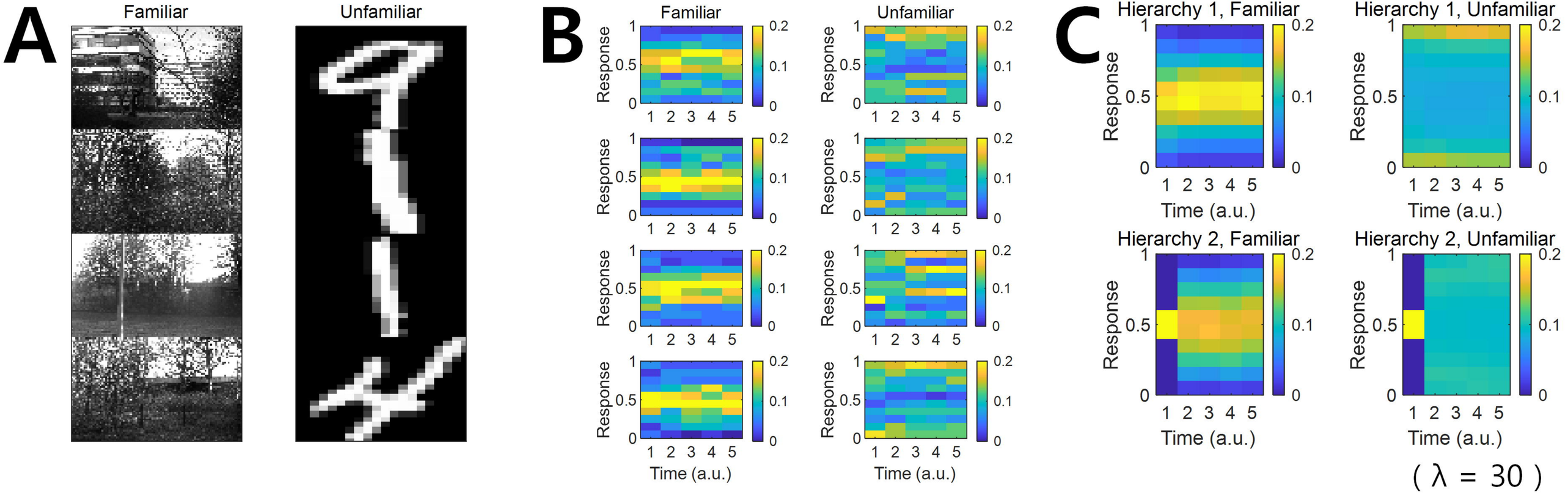
679 65. Zhang S, Xu M, Kamigaki T, Hoang Do JP, Chang W, Jenvay S, Miyamichi K, Luo L, Dan Y.
680 Long-range and local circuits for top-down modulation of visual cortex processing. Science
681 2014 Aug 08; 345(6197):660-665. doi: 10.1126/science.1254126

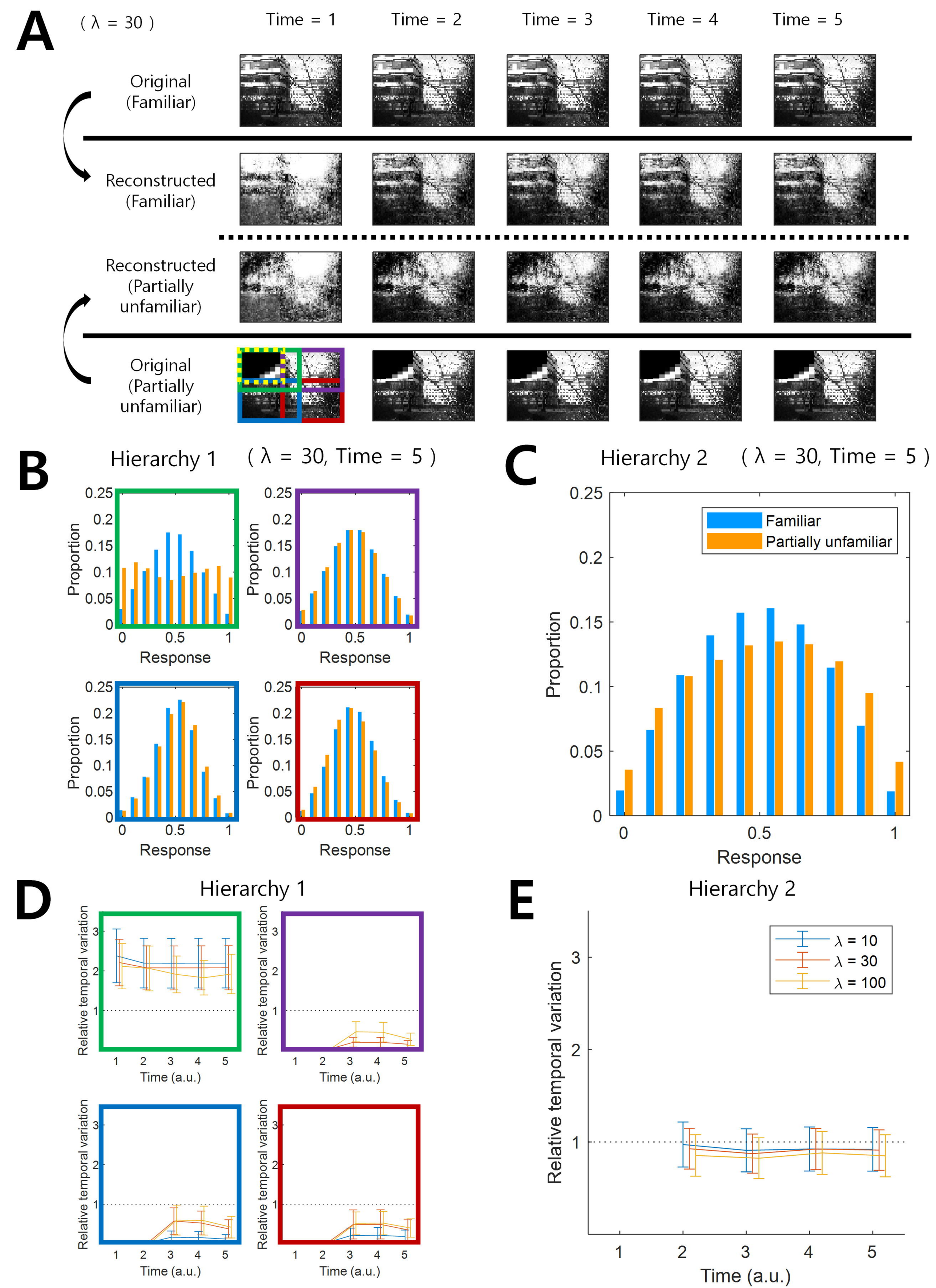
682

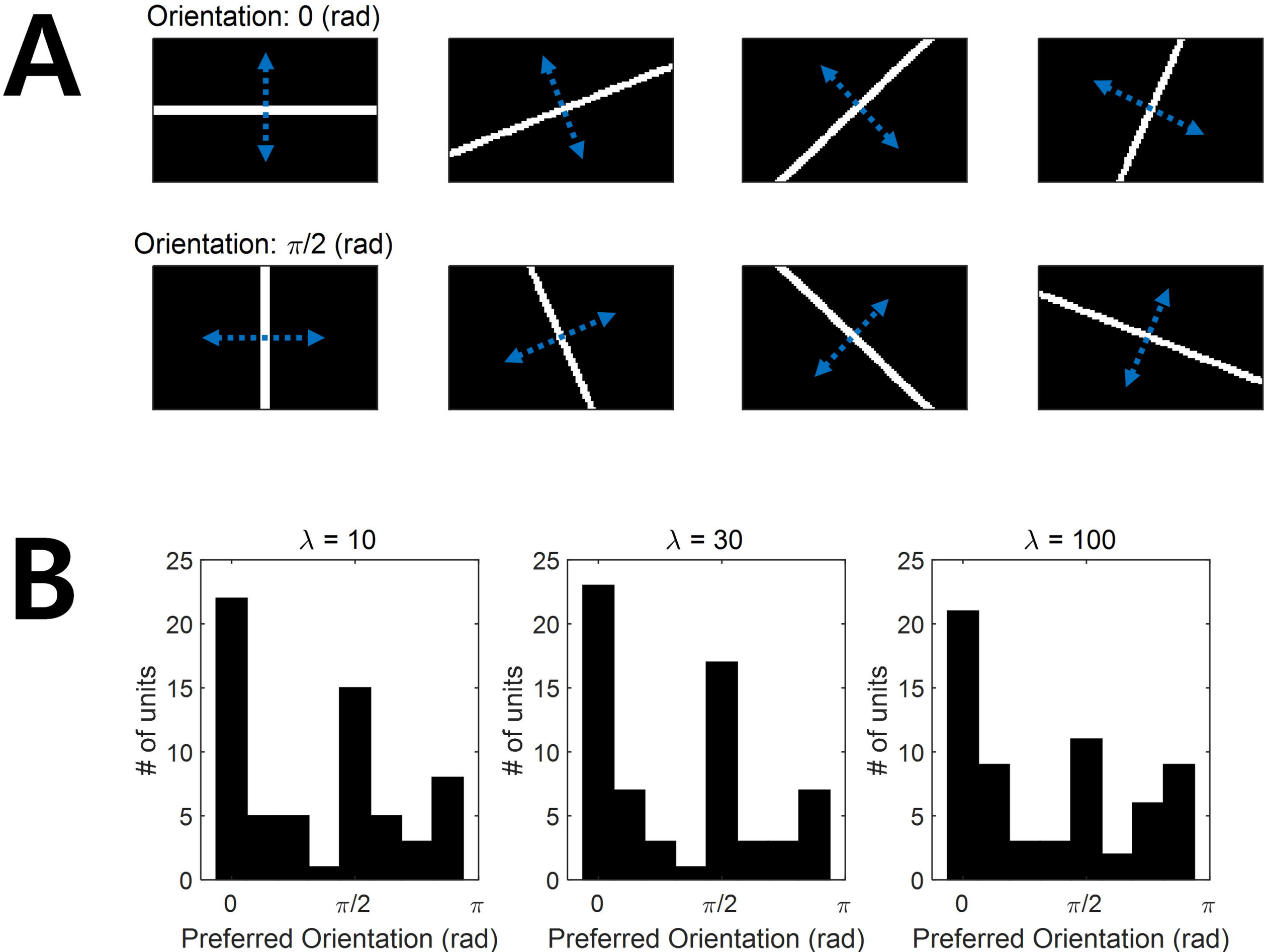

Brain hierarchy
Input
Time
Prediction
Minimize temporal differences
&
Maximize responses entropy


A
Hierarchy 2
Hierarchy 1
Image
B
Synaptic weight magnitude
$(\Sigma_i w_{ij}^2)^{1/2}$
10^1
10^0
10^{-1}
10
30
100
λ
Bottom-up
Top-down
C
λ = 30
λ = 10
λ = 30
λ = 100
Time (a.u.)
Unit
Original
Reconstructed
D
$\lambda = 10^5$
λ = 0, hierarchy 1
λ = 0, hierarchy 2
$\lambda = 10^5$, hierarchy 1
$\lambda = 10^5$, hierarchy 2
Time (a.u.)
Unit
Original
Reconstructed


A
Ref.
(λ = 30)
Relative distance from ref. on hierarchy 1
Relative distance from ref. on image
B
Image inputs
Hierarchy 1
Hierarchy 2
1
0
(λ = 30)
C
Global similarity
Correlation over distances
Hierarchy 1
Hierarchy 2
10
30
100
λ
D
Local similarity
Correlation over distances
Hierarchy 1
Hierarchy 2
10
30
100
λ
E
λ = 10
λ = 30
λ = 100
Relative temporal variation
Hierarchy 1
Hierarchy 2
Time (a.u.)


A
Familiar
Unfamiliar
B
Familiar
Unfamiliar
Response
Time (a.u.)
C
Hierarchy 1, Familiar
Hierarchy 1, Unfamiliar
Hierarchy 2, Familiar
Hierarchy 2, Unfamiliar
(λ = 30)


A
(λ = 30)
Time = 1
Time = 2
Time = 3
Time = 4
Time = 5
Original (Familiar)
Reconstructed (Familiar)
Reconstructed (Partially unfamiliar)
Original (Partially unfamiliar)
B
Hierarchy 1
(λ = 30, Time = 5)
Proportion
Response
C
Hierarchy 2
(λ = 30, Time = 5)
Familiar
Partially unfamiliar
Proportion
Response
D
Hierarchy 1
Relative temporal variation
Time (a.u.)
E
Hierarchy 2
λ = 10
λ = 30
λ = 100
Relative temporal variation
Time (a.u.)


A
Orientation: 0 (rad)
Orientation: π/2 (rad)
B
λ = 10
λ = 30
λ = 100
# of units
0
5
10
15
20
25
0
π/2
π
Preferred Orientation (rad)