

## Cortical time-course of evidence accumulation during semantic processing

Gayane Ghazaryan <sup>1</sup>✉, Marijn van Vliet <sup>1</sup>, Lotta Lammi<sup>1</sup>, Tiina Lindh-Knuutila<sup>1</sup>, Sasa Kivisaari <sup>1</sup>, Annika Hultén<sup>1,2</sup> & Riitta Salmelin <sup>1,2</sup>

Our understanding of the surrounding world and communication with other people are tied to mental representations of concepts. In order for the brain to recognize an object, it must determine which concept to access based on information available from sensory inputs. In this study, we combine magnetoencephalography and machine learning to investigate how concepts are represented and accessed in the brain over time. Using brain responses from a silent picture naming task, we track the dynamics of visual and semantic information processing, and show that the brain gradually accumulates information on different levels before eventually reaching a plateau. The timing of this plateau point varies across individuals and feature models, indicating notable temporal variation in visual object recognition and semantic processing.

<sup>1</sup>Department of Neuroscience and Biomedical Engineering, Aalto University, P.O. Box 12200, FI-00076 Aalto, Finland. <sup>2</sup>Aalto NeuroImaging, Aalto University, P.O. Box 12200, Aalto FI-00076, Finland. ✉email: [gayane.ghazaryan@aalto.fi](mailto:gayane.ghazaryan@aalto.fi)

Concepts are fundamental building blocks of our understanding of the world and communication with others. Brain regions associated with semantic knowledge have been extensively studied<sup>1–4</sup>, yet the mechanisms by which concepts are accessed in the brain are still not well understood. Object recognition is a common task that requires accessing concepts, and it is a prime target for experimental research on semantic processing. Despite how rapidly we can recognize an object, the process likely consists of multiple phases<sup>5</sup>, involving the interplay between visual and semantic properties<sup>6</sup> and the emergence and accumulation of information over time<sup>7,8</sup>. In this work, we track and examine the dynamic progression of semantic processing in the human brain over the course of visual object recognition. In particular, we examine the accumulation of information through time.

Object recognition is essential in everyday functioning and, indeed, it has received great interest in human neuroscience<sup>6,9–11</sup>. The underlying process is thought to progress from a focus on low-level visual features to a focus on complex semantic representations<sup>6</sup>. Semantic knowledge models<sup>12–15</sup> together with machine learning methods<sup>16,17</sup> have been utilized to link brain activation patterns and semantic processing. Applying such methods to time-sensitive neuroimaging data, semantic processing has been shown to follow a coarse-to-fine progression<sup>6</sup>. Coarse semantic categories, but not individual concepts, can be discriminated based on earlier brain response patterns; by around 150 ms, it is possible to decode categories of objects<sup>6</sup>. Individual concepts, however, can be decoded only at later time points, around 300–450 ms after stimulus onset<sup>7,18–20</sup>.

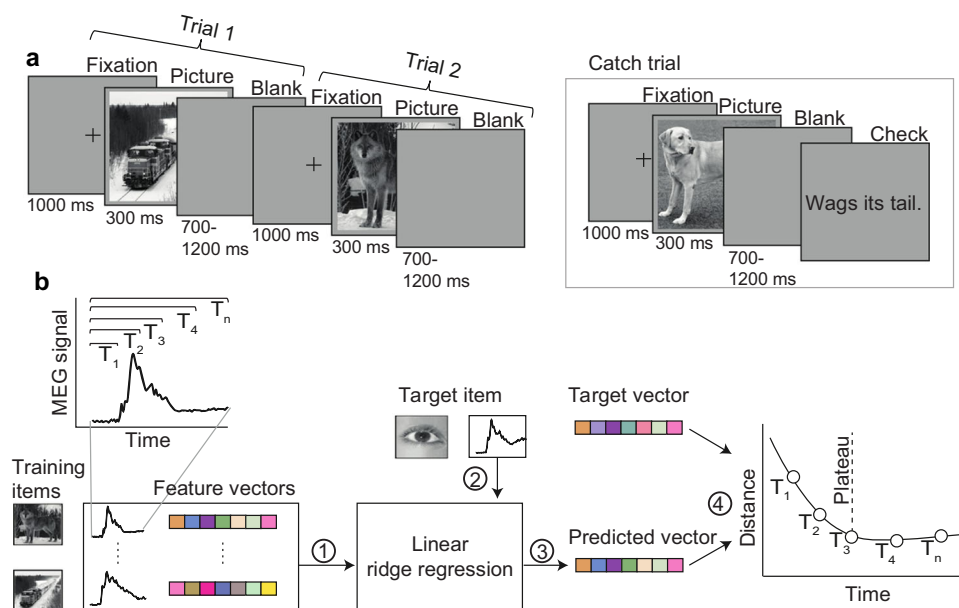
Previous studies, which have focused on decoding at isolated time windows in sequence, have revealed a pattern of increasing decoding accuracy up to a peak, followed by a gradual decrease in accuracy<sup>7,18,19</sup>. Furthermore, cross-temporal decoding (decoding information from one time point with models trained on other time points) has shown that the underlying brain activation patterns evolve rapidly following stimulus onset, with some

generalization of similar encoded information across nearby time points<sup>21,22</sup>. Such generalization indicates that information is maintained or accumulated from overlapping processes<sup>8,23</sup>.

To investigate how the brain processes information and accesses a concept, we used MEG (magnetoencephalography) brain response data from a picture viewing experiment, in which participants were shown pictures of objects and asked to silently identify them (Fig. 1a). This task focuses on object perception through to concept access<sup>5</sup>, and excludes later processes involving phonological forms and speech production. We contrast two approaches for decoding semantic representations: the traditionally used sliding approach taking one time point at a time, and a cumulative modeling approach (Fig. 1b) that widens the window at each time step. We demonstrate that, indeed, the brain gradually accumulates semantic information with eventual stabilization of accumulated information and access to a fully enriched object identity. For brain-level object recognition, it seems essential to take into account all information gathered up to a certain time point, instead of limiting the view to a sequence of single snapshots, analyzed in isolation.

## Results

Participants performed a silent visual concept identification task while MEG responses were recorded (Fig. 1a, see “Methods” for further details). We represented each concept as a semantic feature vector derived from a large text corpus using word2vec<sup>24,25</sup>, and trained models to predict these feature vectors from the MEG responses. To emphasize semantic properties of concepts over low-level visual features of specific stimulus images, we analyzed average brain responses to multiple different exemplars of the same concept. We performed analyses on both the grand average MEG responses, and separately for each participant. The MEG responses, spanning a period of 1000 ms after the stimulus onset, were downsampled and binned into 20 ms time points, resulting in time series of 50 points for each of the 60 concepts. Each time



**Fig. 1 Overview of experiment and methodology.** **a** Experimental paradigm. In each trial, participants were presented with a picture of an object and asked to identify it; randomly occurring catch trials ensured compliance. **b** Overview of the brain-to-semantic mapping method. Linear regression models were (1) trained on a set of brain response-semantic vector pairs and (2) tested on the brain response to one left-out concept, (3) yielding a predicted vector. (4) The distance between the predicted vector and the target vector was calculated. This procedure was repeated for all concepts. The same procedure was performed for each instance of the cumulative window. The dashed line corresponds to the point after which prediction-target distance no longer decreased (plateau point). Semantic vectors were created using the word2vec algorithm on a large Finnish text corpus.

point contained MEG responses from 204 sensors. Following a neural decoding approach, we employed multivariate linear ridge regression to predict feature vectors from MEG responses (see “Methods” for further details).

In order to discriminate between conceptual and perceptual processes (known to be intertwined<sup>26</sup>), we compared the semantic feature model (word2vec) to a visual feature model that aims to describe visual object recognition in a manner that is similar to the primate visual cortex (CORnet)<sup>27</sup>. Visual feature vectors were derived by inputting the same graphical images that were shown to participants into the CORnet-S model that consists of four layers, corresponding to the cortical regions V1, V2, V4, and IT. Thus, we considered five different feature vector models: the word2vec model derived from a large text corpus; and four levels of the CORnet-S visual processing model. To evaluate predictive performance, we used a leave-one-concept-out zero-shot approach, in which trained models were tested on concepts that were excluded from the training sets. We used the prediction-target distance as a metric and investigated how this varied as a function of time.

**Grand Average.** Using grand average data, we first sought to examine the degree of shared encoded information between different time points. To do this, we employed a cross-temporal decoding approach in which models are trained on data from one time point and tested on another time point. All feature models indicated significant generalization of information encoded in the brain response ( $p < 0.05$ , based on a permutation test with 1000 permutations, FDR corrected) (Fig. 2a, b). The generalization was most pronounced for consecutive time points, with further away time points showing less generalization. The start of this window, defined as the point after which there was significant generalization between consecutive time points in the cross-temporal decoding (Fig. 2b), varied from 250 (V1, V2) to 270 ms (word2vec) (Fig. 3).

We examined the underlying reason for this generalization by considering two alternatives: either the encoded information is maintained, or accumulation of information continues throughout this period. For each feature model, we compared two different types of models with (1) sliding and (2) cumulative approaches. All sliding models showed a decrease in prediction-target distance over time until a trough, followed by an increase in distance. In contrast, the cumulative models showed decreasing prediction-target distance, with an eventual plateau (Fig. 2c). Both approaches yielded mean prediction-target distances significantly lower than chance at some time points ( $p < 0.05$ , based on a permutation test with 1000 permutations, FDR corrected). For the sliding models, time points of significantly lower than chance distance were between 140 to 760 ms for word2vec, 100 ms to 780 ms for V1, 80 to 760 ms for V2, 80 to 780 ms for V4, and 80 to 680 ms for IT. For the cumulative models, corresponding time points were 100 ms onwards for word2vec, 100 ms onwards for V1, 80 ms onwards for V2, 80 ms onwards for V4, and 120 ms onwards for IT. The cumulative model eventually yielded significantly lower distance compared to sliding models for all feature models ( $p < 0.05$ , based on a permutation test with 1000 permutations) from 320 ms onwards for word2vec, and from 340 ms/320 ms/240 ms/320 ms onwards for the visual feature models V1/V2/V4/IT. Results are shown in Fig. 2c.

As the cumulative model performed better than the sliding model, the next target of interest was the time when the cumulative model plateaued, as this would indicate whether or not information was accumulated during the generalization window. We reasoned that if the cumulative models plateaued

close to the same time as the generalization began, this would indicate that little new information was encoded in the patterns during the generalization window, and the generalization was purely due to maintenance. Alternatively, if the plateau occurred substantially later, this would indicate that new information was encoded in the brain signal during this period, and there was information accumulation.

We defined plateau points as the time point at which the prediction-target distance no longer meaningfully decreased. We opted for a threshold of 5% for this, meaning that when the model had reached within 5% of the total reduction in distance, we classified it as reaching plateau. We chose a 5% threshold rather than the global minimum whose exact time point could be rather arbitrary due to noise-induced signal variation. We interpreted this plateau point as the time after which little more relevant information was encoded in the MEG signal. We then used a mixed effects linear regression to predict plateau points from the feature model used, with random intercepts for each concept.

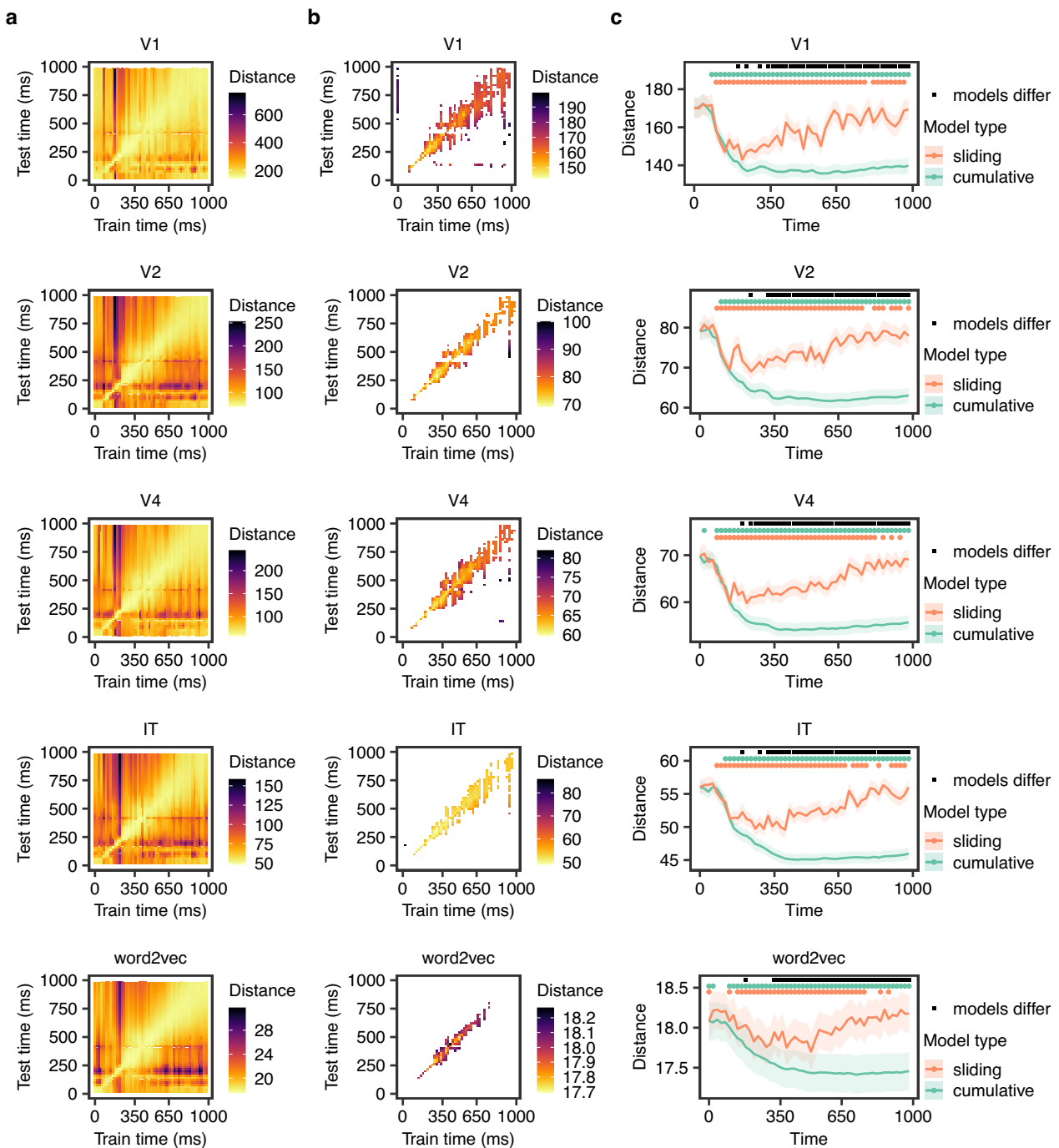
Figure 3 shows estimated plateau points in comparison to the points when the generalization window began. The plateaus were significantly later than the start of the generalization window for all models,  $p < 0.05$ . This indicates that further relevant information is encoded in the MEG signal during this period. In other words, there is information accumulation rather than only maintenance.

The estimated mean (95% CI) plateau was 382 (348–416) ms for V1, 418 (384–451) ms for V2, 415 (381–449) ms for V4, 447 (414–481) ms for IT and 439 (405–473) ms for word2vec. Apart from V2 and V4 ( $p = 0.9803$ ), and IT and word2vec ( $p = 0.4406$ ), the estimates were significantly different (all  $p < 0.0001$ , Tukey correction).

**Individual level results.** Following the grand average analyses, we investigated whether a consistent pattern could be identified at the individual level and explored individual differences therein. We again compared the sliding and cumulative models to investigate the progression of information processing of each participant. The results for word2vec are presented in Fig. 4 and the visual feature models are presented in Supplementary Figs. 1–4. All participants displayed decreasing prediction-target distance as a function of time with an eventual plateau for cumulative models, which yielded significantly better than chance predictions (based on a permutation test with FDR correction,  $p < 0.05$ ). The sliding models in word2vec produced significantly better than chance predictions at some time points (based on a permutation test with FDR correction,  $p < 0.05$ ) for the majority of participants (but not 5, 6, 7, 10, 12, 17).

A linear mixed model (with random intercepts for each concept) showed significant variation in plateau points between participants and feature models (Fig. 5a, Table 1). There was a significant main effect of the feature model ( $F(4, 5546) = 3.88$ ,  $p = 0.001$ ) following a similar pattern observed in the grand average analyses with IT and word2vec plateauing significantly later than V1 and V2 ( $p < 0.05$ ), see Table 2. We also observed a significant main effect of participants indicating inter-individual variability ( $F(18, 5546) = 9.23$ ,  $p = 0.001$ ).

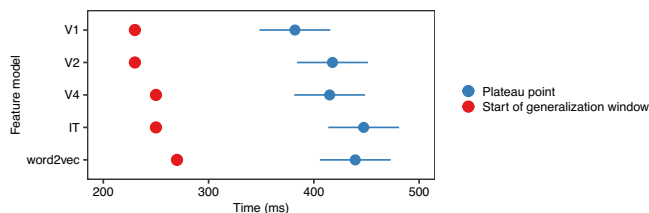
We explored how plateau timings in one feature model were related to those in other feature models. The visual feature models were positively correlated and higher correlations were observed between consecutive layers such that participants with earlier plateaus in V1 had earlier plateaus in V2; similarly for V2–V4 and V4–IT (Fig. 5b and Supplementary Figure 5). We also examined how the differences between participants in prediction-target distance changed as a function of time. Specifically, we looked at



**Fig. 2 Results of grand average analysis. a** Cross-temporal decoding results on grand average data. Here, models were trained on one 20-ms time point and tested on another 20-ms time point. The color corresponds to the prediction-target distance averaged over all targets. **b** Temporal cross-decoding results on grand average data. Only statistically significant values shown ( $p < 0.05$ , FDR corrected, based on permutation tests with 1000 permutations). **c** Distance between predictions and targets over time on grand average data, using two different types of models: a sliding model taking one 20-ms time point at a time, without overlap, and a cumulative window with width increasing at 20-ms increments. The lines represent the mean across concepts, and the edges of the shaded areas indicate plus and minus one standard error. The dots above each line plot indicate time points with statistically significant differences ( $p < 0.05$ ), based on permutation tests with 1000 permutations, FDR corrected). Note that due to differences in the feature space of the models, the magnitudes of the Euclidean distance values are on different scales, and should not be directly compared. Instead, the temporal patterns are the focus of interest. For further details on the calculation of distance, see "Methods".

the time points when differences between participants reached their maximum. The model based on word2vec appeared to reach maximum variation across participants later compared to visual feature models (Fig. 5c).

**Representational similarity analysis.** To investigate the brain areas involved in the information accumulation processes, we performed representational similarity analysis (RSA) between concept similarity in the brain (the brain-level concept-to-



**Fig. 3 Comparison of generalization window and plateau points.** Blue dots indicate estimated marginal means (95% confidence intervals) from a linear mixed model predicting plateau points from the feature model with random by-concept intercepts. Red dots show the start of the generalization window, identified from the cross-temporal decoding, for comparison.

concept dissimilarity matrices at different time points in sliding and cumulative approaches; Fig. 6a, b) and concept similarity in the feature models (the feature model concept-to-concept dissimilarity matrices; Fig. 7). We observed the highest RSA scores in the occipital regions of both hemispheres (Fig. 8a, b). Incrementally adding time points did not change the regions where the RSA scores were highest. The 0–420 ms and onward windows showed the highest correlations. This timing aligned with the decoding results described above. RSA figures for the visual feature models are presented in Supplementary Figs. 6–9.

## Discussion

In this study, we tracked the progression of information processing throughout visual object recognition. We identified a period of information generalization, and found that during this time there is information accumulation, during which concept representations are enriched. This interpretation is in line with Contini et al.<sup>8</sup>, who suggested two reasons for information generalization: (1) the encoded representation is maintained throughout this period, or (2) there are ongoing overlapping processes of differing duration, such that some information is maintained while the representation is enriched with further accumulation. By demonstrating information accumulation in this window, our work brings relevant new findings to complement earlier studies on object recognition<sup>28,29</sup> and semantic access in general<sup>7,30–37</sup>, especially regarding its temporal progression<sup>7,18,19,38</sup>.

We demonstrated information accumulation with the help of models that mapped between MEG responses and concepts, and showed that these models generalize to new concepts (following a zero-shot approach<sup>17</sup>). We compared a semantic feature model (word2vec) to models using visual features generated from CORnet, a model mimicking visual processing in primate visual cortex (V1, V2, V4, and IT). Through this comparison, and by using static images and multiple instances of each concept, we aimed to reduce the effects of mere stimulus characteristics on the observations and highlight neural effects related to semantic processing of the concept. Importantly, semantic features were derived from a large text corpus, and not directly from the limited set of visual stimuli shown to the participants. The visual feature representations were derived by applying the pre-trained CORnet visual model to the stimuli the participants were shown. When comparing visual and semantic models it is important to recognize that such models are always simplifications of the true underlying processes. In this particular case, the visual feature models may not capture all underlying visual processes. As such, differences between the visual models and the semantic model may not only indicate semantic processing, but rather inadequacies in the visual models<sup>26</sup>. Future work may avoid the confounding between visual and semantic processing by focusing on

written words<sup>38–40</sup>, however, written words are known to be a more challenging medium for neural decoding<sup>41</sup>.

The sliding approach indicated that there was information relevant to concepts starting at about 80–100 ms, which matches the timing reported in previous work<sup>7,18,26,42,43</sup>. Cross-temporal decoding indicated that there was also generalization in the brain signal from about 250 ms onwards. The cumulative models resulted in significantly lower prediction-target distance than the sliding models in all feature models, from 240–340 ms onwards, indicating there is information accumulation.

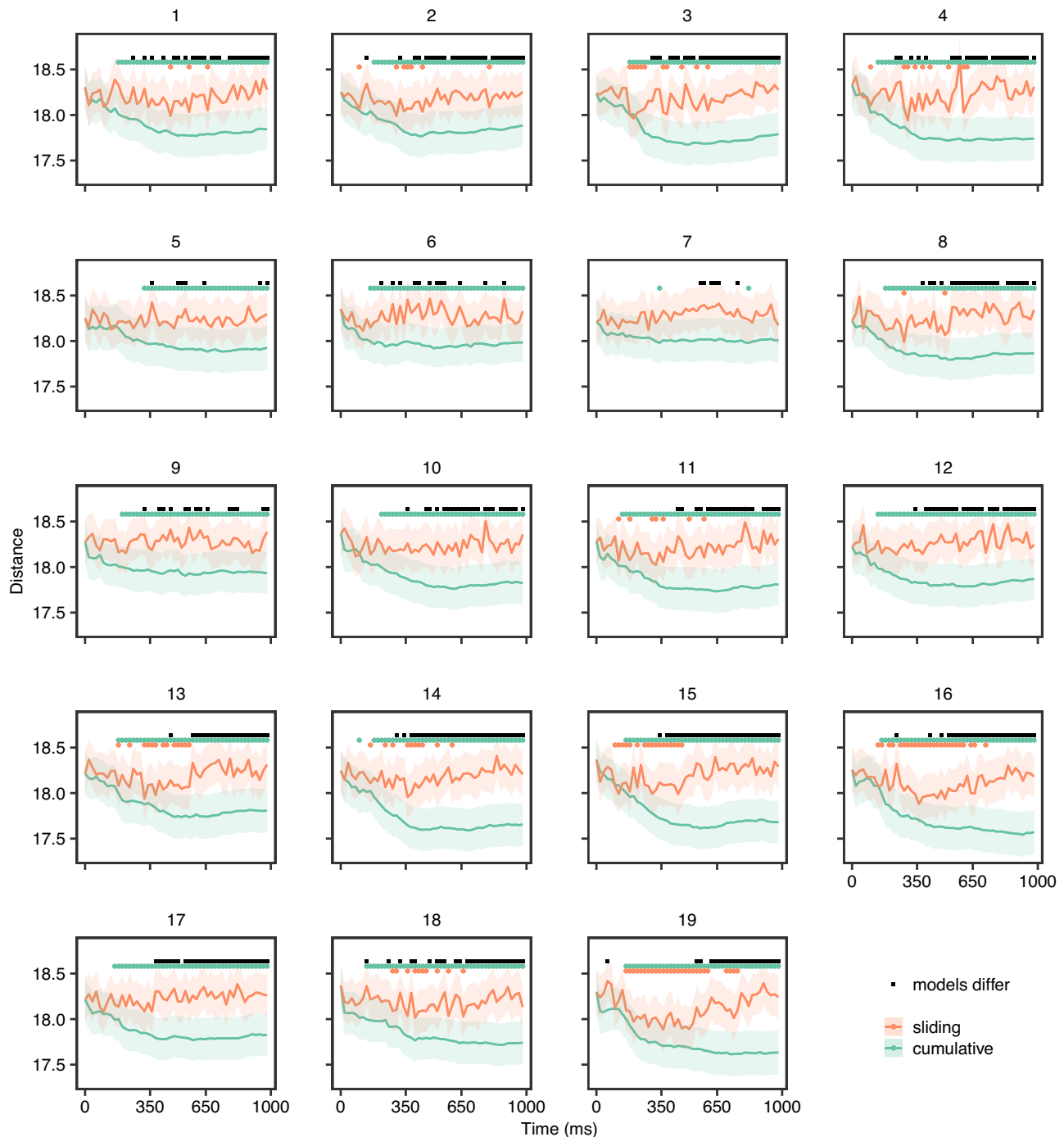
Models based on lower-level visual features (V1, V2) plateaued earlier than the semantic feature model (word2vec) or high-level visual feature models (IT). This pattern matches the level of processing. Furthermore, we found that plateau points of consecutive layers in the visual feature model were systematically correlated. Although word2vec showed moderate correlation with V4, there was indication that individual variability in word2vec decoding was delayed compared to the visual models. This suggests that the semantic feature model is capturing information other than visual feature correlates, and the decoding is not just relying on features correlated with low-level visual processing.

We propose that the plateau point of the semantic model could be interpreted as the time point after which the representation of the concept is no longer enriched. On group-level data, the mean plateau time across concepts was around 450 ms. This, coupled with the generalization shown between 270 and 750 ms, indicates that there is likely both accumulation of information (preceding plateau) and maintenance (following plateau). The timing of the plateau point varied between participants, with means ranging from approximately 350 to 500 ms.

Bo et al.<sup>44</sup> showed that visual activity appears in different regions up to around 360 ms. Peaks in conceptual processing (when controlling for visual features) have been shown to occur at different points between 180 and 540 ms<sup>26</sup>. The plateau points observed here, while not directly comparable due to differing methodologies, line up with these previous results.

Disentangling accumulation from maintenance, however, may not be as simple in the presence of noise. In a situation where there is only maintenance, but the recorded signals have substantial noise, a similar pattern of decreasing prediction-target distance might be observed, as adding more time points would counteract the noise and improve performance. Based on the behavior of the cross-temporal, cumulative, and sliding models, however, we consider this to be a less likely explanation than accumulation. Specifically, we refer to the following observations: First, the cross-temporal models indicated that while there was some generalization, this was predominantly between consecutive time points. If there was a constant signal with noise, we would not expect such a difference between consecutive and non-consecutive cross-temporal decoding performance. Second, the sliding models showed an increase and then decrease in predictive performance, rather than a sustained level of performance, counter to what would be expected in pure maintenance. Third, significant prediction by the sliding models continued well past the plateau of the cumulative models. If the cumulative models were simply accommodating for noise in the data, we could expect the increase in performance to continue until the signal was no longer predictive.

RSA indicated that occipital areas were relevant to semantic processing of pictures and showed temporal patterns in accordance with the decoding approach, with higher RSA scores at time points when decoding performed better. Interestingly, brain regions that are consistently reported in studies of picture naming<sup>45</sup>, such as the left temporal and left parietal cortices, did not strongly account for semantic relationships between target concepts. However, our results align with Simanova et al.<sup>43</sup>, who

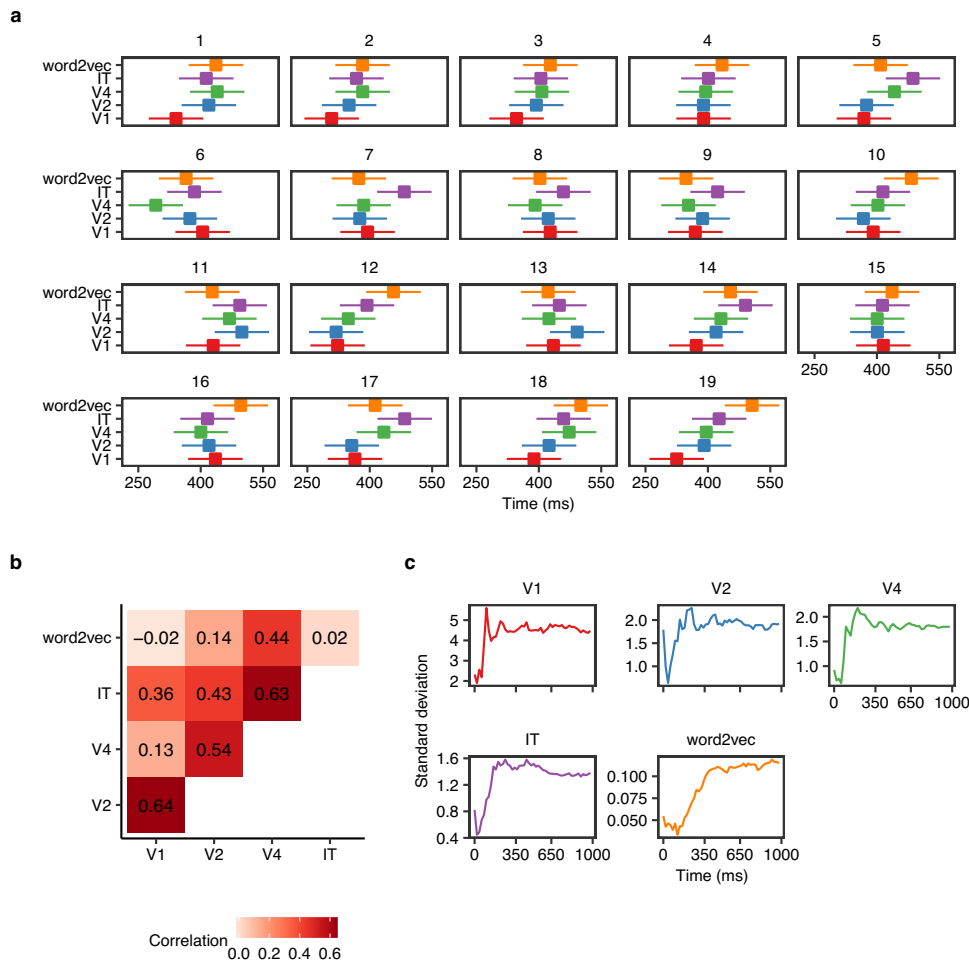


**Fig. 4 Distance between predictions and targets over time in individual participants for semantic model.** The lines represent the mean across concepts, and the edges of the shaded areas indicate plus and minus one standard error. The dots above each line plot indicate time points with statistically significant differences ( $p < 0.05$ ), based on permutation tests with 1000 permutations, FDR corrected). All participants displayed decreasing prediction-target distance as a function of time with an eventual plateau. Plateau points (for cumulative models) and troughs (for sliding fixed-length models) were defined as the time point after which prediction-target distance no longer decreased by more than 5%.

suggested that the predominance of occipital areas may be due to inherent visual similarities between semantically similar objects. In other words, the appearance of an object is tied to its semantic meaning, so it is unsurprising that brain regions related to visual processing emerged in the RSA analysis. While it is possible that silent identification of the pictures did not activate the phonological form of the concept as strongly as an overt naming task would have done, the lack of involvement of the typical language areas may be a reflection of the fact that semantic similarity is not

mirrored in phonological similarity (for example ‘cat’ and ‘dog’ are semantically near but phonologically distant).

The fact that people agree on names of objects and can communicate about them indicates that there are commonalities in semantic understanding. However, as each person has a unique life experience, the underlying semantic processes are also likely to vary. Previous studies have reported inter-individual variation in behavioral measures of naming speed<sup>46</sup>, neural correlates of semantic representation<sup>47</sup>, and gaze-behavior measures of visual



**Fig. 5** Variation of plateau points and prediction-target distance between participants. **a** Estimated plateau means (95% confidence intervals) for each participant based on linear mixed model. **b** Correlation between estimated plateau points of the different feature models. **c** Standard deviation between participants of prediction-target distance (mean over concepts) as a function of time, for each feature model.

**Table 1** Mixed-model regression analysis of plateau times.

Effect	df	F	p-value
Participant	18, 5546	3.88	<0.001
Feature model	4, 5546	9.23	<0.001
Participant × Feature model	72, 5546	1.13	0.271

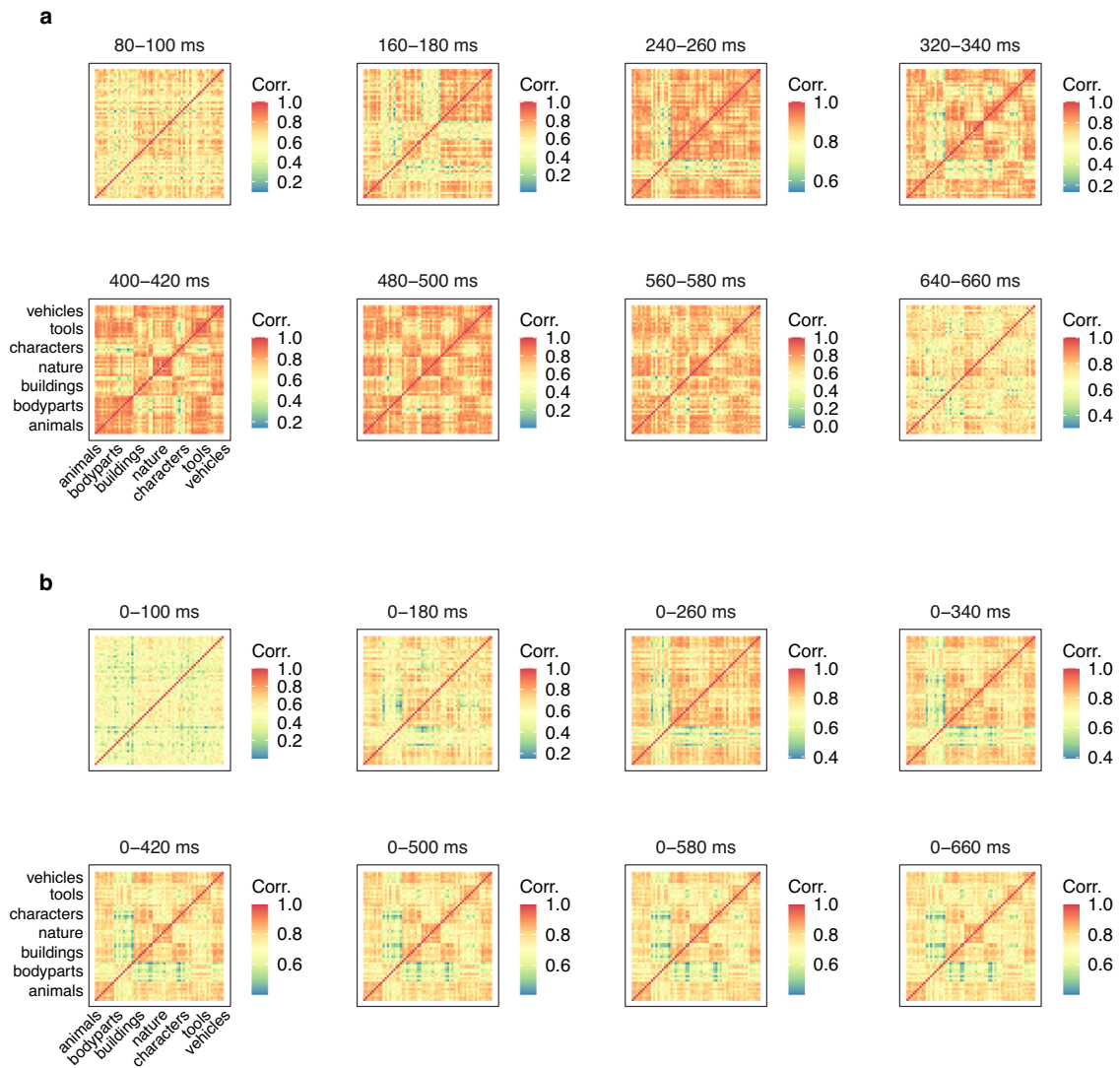
**Table 2** Estimated plateau points for each feature model.

Feature model	Min estimate	Max estimate	SE
V1	308 ms	435 ms	33 ms
V2	319 ms	499 ms	33 ms
V4	292 ms	473 ms	33 ms
IT	368 ms	494 ms	33 ms
word2vec	347 ms	506 ms	33 ms

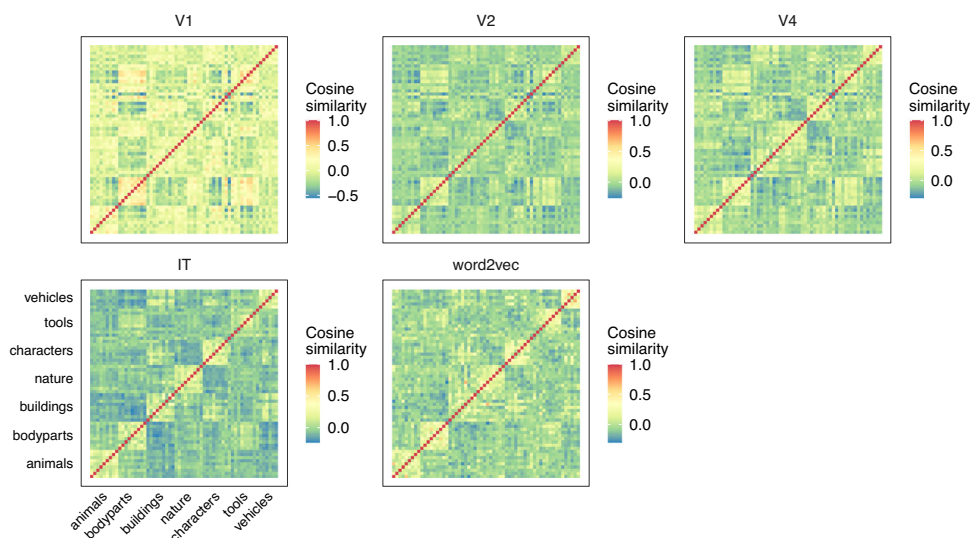
salience<sup>48</sup>. Individual variation has also been indirectly investigated through cross-decoding between individuals. This is performed by training models on data from one or more individuals and testing on data from another individual<sup>32,33,49</sup>. Generally, such cross-decoding has been less accurate than within-individual

decoding. As these studies predominantly used imaging methods that favor high spatial precision over temporal precision, the results likely indicate individual variation in the cortical areas involved in language processing, the existence of which has been known since early studies<sup>50</sup>. Individual differences in the temporal domain of semantic processing have also recently been indicated by Rupp et al.<sup>19</sup>, who reported individual variation in the windows when decoding performed best, suggesting differences in the progression of semantic understanding. Here, we found individual variability in accumulation of visual and semantic information. This finding is relevant to, for example, development of brain-computer interfaces, where individual variability may need to be taken into account. Variability between concepts is also an intriguing question for future studies but will likely require more repetitions of each concept than in the present study to ensure less noisy cortical time courses for individual concepts.

We have presented here a new perspective on the temporal dynamics of semantic understanding that opens future avenues of research. These include a deeper understanding of individual cognitive variation, addressing the link to behavioral measures, adapting to other modalities such as spoken or written words, and investigating concept processing in context by using more naturalistic stimuli such as sentences or stories. Such research will bring us towards a more complete model of language within the brain.

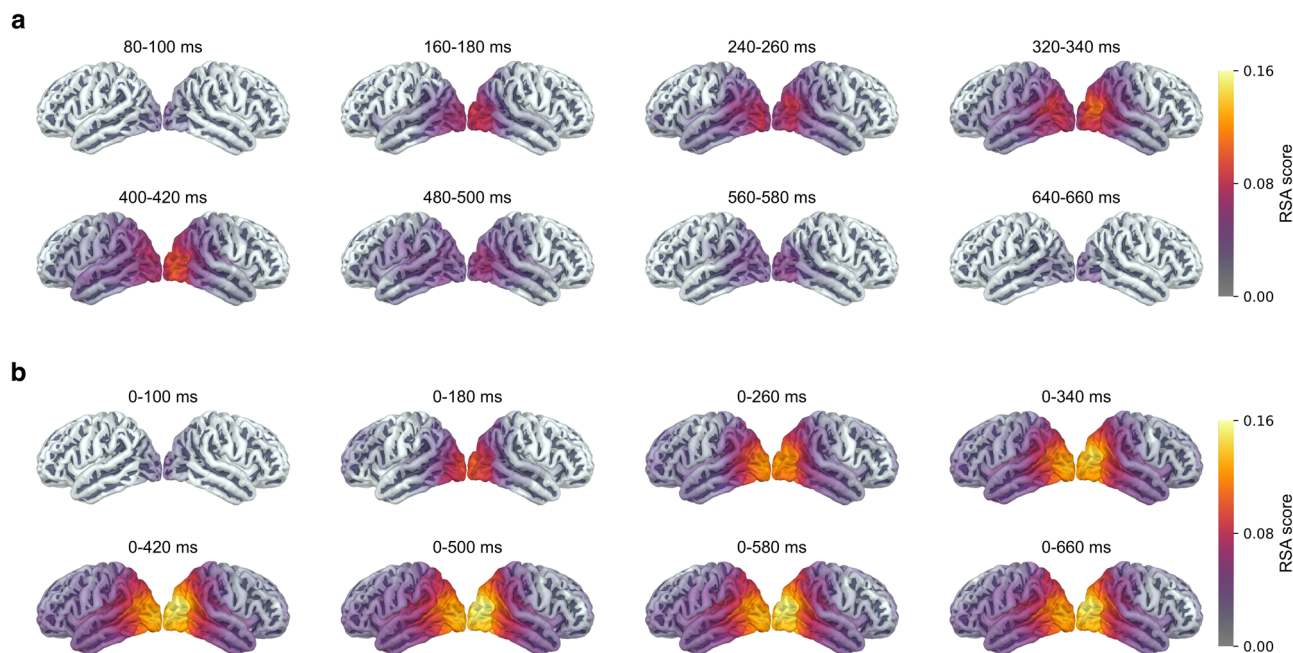


**Fig. 6 Concept-to-concept brain signal dissimilarity matrices (sensor-level) plotted over time, averaged across all participants. a** Sliding approach. **b** Cumulative approach.



**Fig. 7 Concept-to-concept similarity for different feature models.** Pairwise cosine similarity between feature vectors for each concept is shown and categorical structure can be seen along the diagonals.





**Fig. 8 RSA maps for the semantic feature model.** RSA maps illustrating the statistically significant clusters for different time windows, with **a** the sliding fixed-length approach and **b** the cumulative approach. For details on calculations of RSA scores see Methods. We used a cluster permutation test<sup>61</sup> across participants with a cluster threshold of  $p = 0.01$ , a cluster-wide significance threshold of  $p = 0.05$  and 5000 permutations in accordance with Hultén et al.<sup>38</sup>.

## Methods

**Participants.** Twenty native speakers of Finnish (females/males 10/10, age range 20–27, mean age 22) participated in the study. All participants were right handed (Edinburgh handedness questionnaire<sup>51</sup>) and had normal or corrected to normal vision. The study was approved by the Aalto University Research Ethics Committee and participants provided written informed consent prior to their participation. All ethical regulations relevant to human research participants were followed. Data from 1 participant was excluded due to technical issues with the MEG recordings, leaving data from 19 participants in the final analysis.

**Stimuli and procedure.** Stimuli consisted of 300 grayscale photographic images of 60 concrete Finnish nouns (five different depictions of each). To minimize the effects of low-level visual features on the neural responses, there were five different images depicting each concept. Overall, each concept was presented in picture form 18 times (across three sessions on different days), the responses of which were averaged.

The concepts belonged to 7 different categories: animals, body parts, buildings, nature, human characters, tools/artifacts, and vehicles. Details of the nouns are presented in Supplementary Table 1. There were nine concepts in each category except for vehicles, which had six concepts. In the experiment, the concepts were also presented in written and auditory forms in separate trials; the responses from those trials were not analyzed in this study.

Participants were tasked with viewing each picture and silently identifying and thinking about the depicted object. The stimuli were presented at a size of  $106 \times 106$  mm on a screen 140 cm from the participants' eyes, corresponding to a visual angle of  $4.3^\circ$ . Each trial started with a fixation cross displayed for 1000 ms. The picture was then shown for 300 ms, followed by a blank screen for a randomized duration of 700–1200 ms (Fig. 1a).

To ensure that participants remained engaged during the experiment, we included comprehension tasks after 10% of trials. In these tasks, participants used optical response pads to indicate

whether or not a written description was characteristic of the previously shown concept. As these tasks occurred after the trials, they did not interfere with the responses, thus all trials were included in the analysis.

**Data acquisition.** MEG measurements were conducted at the Aalto NeuroImaging MEG Core (Aalto University, Espoo, Finland) using a Vectorview whole-head MEG system (MEGIN (Elekta Oy), Helsinki, Finland). The system has 306 sensors (204 planar gradiometers, 102 magnetometers). The head position was continuously tracked during the experiment by 5 head position indicator (HPI) coils placed at known locations with respect to identifiable anatomical landmarks. Eye movements and blinks were captured using 2 electrode pairs (one pair positioned above and below the left eye, the other in the corner of each eye). The recording was bandpass-filtered at 0.03–330 Hz and sampled at 1000 Hz. Anatomical MRIs were obtained using Siemens Magnetom Skyra 3.0 T MRI scanner with a T1-weighted MP-RAGE sequence at the Aalto NeuroImaging Advanced Magnetic Imaging (AMI) Centre.

**Data preprocessing.** MEG data was first visually inspected and noisy channels were identified. External sources of noise were then removed using spatiotemporal signal space separation (tSSS)<sup>52</sup> with Elekta Maxfilter software (MEGIN Oy, Finland). For each participant, data from different sessions was transformed to the same head position. All subsequent analysis was performed using the MNE-Python software package<sup>53</sup>. The data was low-pass filtered at 40 Hz and split into 1200 ms epochs, the first 200 ms of which was the pre-stimulus baseline interval. To reduce contamination related to heartbeats, eye movements, and blinks, we performed independent component analysis (ICA). In order to minimize the effect of slow drifts on ICA decomposition, we used continuous data high-pass filtered at 1 Hz<sup>54</sup>. Components corresponding to heartbeats, eye movements, and blinks were visually identified and excluded from epochs. Epochs corresponding to the same concept were then averaged, the time

period between 0 and 1000 ms was extracted, and the signal was downsampled to create 20 ms bins. Only data from the gradiometers was used in the final analysis. This resulted in a matrix with 60 concepts  $\times$  204 channels  $\times$  50 time points for each participant.

We computed the source-level estimate of the average response to each concept using minimum norm estimates (MNE)<sup>53</sup>. Anatomical MRIs were used to reconstruct the cortical surface of each participant applying the FreeSurfer software package<sup>55–57</sup>. We used a single-layer boundary element model (BEM) with an icosahedron mesh of 2562 vertices in each hemisphere. When computing the inverse solution, a loose orientation constraint of 0.3 and depth weighting parameter of 0.8 were used. An empirical noise-covariance matrix was computed based on the pre-stimulus 200 ms interval to all concepts. To prepare the data for group-level analysis, participant-level source estimates of each concept were morphed to the FreeSurfer standard template brain (fsaverage).

**Semantic features.** Semantic vector representations for the stimuli were obtained using the word2vec tool with skip-gram architecture and negative sampling algorithm<sup>25</sup>. Each concept was represented as a vector of length 300 which defines its location in semantic space. The components of a vector are based on word co-occurrence statistics in large text corpora, the Finnish Internet Parsebank<sup>24</sup>, which is based on a large sample (1.5 billion words) of Finnish language websites. Co-occurrence was considered to take place when a word appeared within a window from 5 words before to 5 words after the word corresponding to the concept of interest.

**Visual features.** The CORnet model is a neural network architecture designed to simulate the processing of visual information in the primate brain. It consists of multiple layers of artificial neurons that are modeled after the neurons found in the primary visual cortex (V1), secondary visual cortex (V2), visual area V4, and the inferior temporal cortex (IT). V1, V2, V4, and IT form a hierarchy of visual processing, with each region responsible for processing increasingly complex visual information. Beginning with low-level features such as orientation and color, each region builds upon the previous one to construct a more complete representation of visual stimuli. This process culminates in high-level visual processing, such as object recognition<sup>27</sup>.

We created visual feature vectors by inputting the grayscale images into the CORnet-S model, and saving the outputs of each layer. We reduced the dimensionality of the visual feature vectors using PCA (principal components analysis). Following the zero-shot approach, we did this transformation in a cross-validated manner by first leaving out the exemplars of the left-out test concept, and then calculating the principal components on the training set. We then projected both the training and test vectors onto 295 principal components (the largest possible for the size of the training set) and averaged the transformed feature vectors across exemplars to arrive at one visual feature vector for each concept (per iteration of the cross-validation). For RSA, the procedure was the same but without leaving out concepts.

**Regression models.** We used a zero-shot decoding approach<sup>17</sup>. Regularized multivariate ridge regression, as implemented in scikit-learn<sup>58</sup>, was used to fit a model that predicts the semantic feature vector of a target concept based on brain response. The sensor-level MEG responses were first standardized with respect to concepts, such that the mean of each predictor (time point-channel pair) was 0 and the standard deviation was 1. We used leave-one-out cross validation, such that models were trained on

59 out of the 60 concepts and evaluated on the remaining concept the model had not been trained on, for all permutations.

For the regression models, we compared prediction and target vectors using Euclidean distance. The Euclidean distance matches the loss function of linear regression. As the Euclidean distance to the training items is minimized during model fitting, it is appropriate to use this metric to assess the predictive performance on the test items. Note that due to differences in the feature space of the models, the magnitudes of the Euclidean distance values are on different scales and not directly comparable. Instead, the temporal patterns are the focus of interest.

**Mapping brain response to semantic space as a function of time.** In accordance with Carlson et al.<sup>21</sup> and Grootswagers et al.<sup>22</sup>, we performed cross-temporal decoding, in which models were trained and tested on different time windows to check whether there is information generalization in the brain across different time points. For this we trained and tested models on pairs of 20-ms time windows on data averaged over all participants. To explore the progression of semantic understanding in more detail we compared two types of models on average sensor-level MEG responses. First, we used sliding windows of fixed length similar to Sudre et al.<sup>18</sup> and Rupp et al.<sup>19</sup>. Second, we developed a method to examine cumulatively widening windows (see details below). We evaluated models based on prediction-target distance (smaller distance indicates that the model better predicts the target concept). Both types of models were evaluated using leave-one-out cross validation.

For sliding fixed-length window models, regression models were trained and tested on fixed-length subsets of MEG data. A sliding window of 20 ms was used, with no overlap between adjacent windows. Each subset was evaluated independently of others. Similar models have been used by Sudre et al.<sup>18</sup>, Rupp et al.<sup>19</sup>, Hultén et al.<sup>38</sup> and we expected a pattern of gradual decrease, followed by an increase in prediction-target distance.

For cumulative window models, regression models were trained and tested on cumulative subsets of MEG data. The window size was sequentially increased by 20 ms. Thus, all previously encoded information was included in the estimation and model evaluation (Fig. 1b).

**Identifying individual differences.** To compare the progression of semantic understanding between participants, we used the cumulative models. We first calculated the progression of semantic information for all concepts for each participant. We then focused on the point of plateau, the time point after which there was less than 5% further decrease in prediction-target distance. We compared these plateau points using a linear mixed model, with random intercepts for concepts.

**Representational similarity analysis.** We performed representational similarity analysis (RSA)<sup>59</sup> on source localized MEG data to extract the brain areas accounting for highest similarity between the brain activation patterns and vector representations of the concepts for different time windows. This was done using the MNE-RSA software package<sup>60</sup>. RSA was performed for both sliding and cumulative time windows.

The model dissimilarity matrices (DSM) were obtained by calculating pairwise cosine distances between feature vectors. This was followed by the calculation of brain DSMs for each participant at each source-level vertex with a searchlight patch radius of 2 cm for the time window of interest. Brain DSMs were computed by calculating pairwise Pearson correlation coefficient between brain signals in response to different stimuli. The relationship between the brain DSMs and model DSMs was

quantified by Spearman rank correlation coefficients for each participant. This resulted in participant-level RSA maps. We then used a cluster permutation test<sup>61</sup> across participants with a cluster threshold of  $p = 0.01$ , a cluster-wide significance threshold of  $p = 0.05$  and 5000 permutations in accordance with Hultén et al.<sup>38</sup>.

**Statistics and reproducibility.** Statistical significance between regression model performance was evaluated using permutation tests with 1000 iterations as in Kivisaari et al.<sup>30</sup>. For each permutation, models were trained and evaluated. For the individual level analysis, this was done separately for each participant.  $p$ -values were calculated by the proportion of test statistics from the permuted data sets that were at least as high as the test statistics from the observed data.

When comparing prediction-target distance to chance-level, the mean distance over all concepts was used. When comparing the two types of models, a paired  $t$  statistic (across concepts) was calculated and compared to the permutation distribution for each time point.  $p$  values were corrected using false discovery rate (FDR) correction.

For RSA we used a cluster permutation test<sup>61</sup> across participants with a cluster threshold of  $p = 0.01$ , a cluster-wide significance threshold of  $p = 0.05$  and 5000 permutations in accordance with Hultén et al.<sup>38</sup>.

**Reporting summary.** Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The text corpus containing 1.5 billion Finnish words used to derive the statistical model cannot be publicly distributed due to the Finnish copyright law limitations. For further information see [https://turkunlp.org/finnish\\_nlp.html#parsebank](https://turkunlp.org/finnish_nlp.html#parsebank). The stimulus concepts are listed in the Supplementary Table 1. The MEG and MRI data are available upon reasonable request from the authors; the data is not publicly available as it contains personal information, and its reuse for other research purposes requires a new ethical pre-review. Numerical data for the figures that do not contain individual participant data is available at <https://zenodo.org/doi/10.5281/zenodo.1007637662>.

### Code availability

For preprocessing we used Elekta Maxfilter software (MEGIN Oy, Finland) and MNE-Python<sup>53</sup>. For the primary analysis we used Python and made substantial use of MNE-Python<sup>53</sup>, scikit-learn<sup>58</sup>, NumPy<sup>63</sup>, and Matplotlib<sup>64</sup>. In addition, we used R and the following packages: afex<sup>65</sup>, lme4<sup>66</sup>, qvalue<sup>67</sup>, ggplot2<sup>68</sup>. The feature vectors used in this study, together with the custom code used in the study can be accessed at <https://zenodo.org/doi/10.5281/zenodo.1007637662>. The code to compute RSA can be found at <https://github.com/wmvanvliet/mne-rsa>.

Received: 5 May 2022; Accepted: 20 November 2023;

Published online: 08 December 2023

### References

- Binder, J. R. & Desai, R. H. The neurobiology of semantic memory. *Trends Cogn. Sci.* **15**, 527–536 (2011).
- Binder, J. R., Desai, R. H., Graves, W. W. & Conant, L. L. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* **19**, 2767–2796 (2009).
- Lambon Ralph, M. A., Jefferies, E., Patterson, K. & Rogers, T. T. The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* **18**, 42–55 (2017).
- Liuzzi, A. G., Aglinskas, A. & Fairhall, S. L. General and feature-based semantic representations in the semantic network. *Sci. Rep.* **10**, 8931 (2020).
- Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P. & Salmelin, R. An MEG study of picture naming. *J. Cogn. Neurosci.* **10**, 553–567 (1998).
- Clarke, A. In *Psychology of Learning and Motivation* Vol. 70, 71–95 (Academic Press, 2019).
- Clarke, A., Devereux, B. J., Randall, B. & Tyler, L. K. Predicting the time course of individual objects with MEG. *Cereb. Cortex* **25**, 3602–3612 (2015).
- Contini, E. W., Wardle, S. G. & Carlson, T. A. Decoding the time-course of object recognition in the human brain: from visual features to categorical decisions. *Neuropsychologia* **105**, 165–176 (2017).
- Wardle, S. G. & Baker, C. I. Recent advances in understanding object recognition in the human brain: deep neural networks, temporal dynamics, and context. *F1000Research* **9**, 590 (2020).
- Mahon, B. Z. & Caramazza, A. Concepts and categories: a cognitive neuropsychological perspective. *Annu. Rev. Psychol.* **60**, 27–51 (2009).
- Martin, A. The representation of object concepts in the brain. *Annu. Rev. Psychol.* **58**, 25–45 (2007).
- Bruffaerts, R. et al. Redefining the resolution of semantic knowledge in the brain: advances made by the introduction of models of semantics in neuroimaging. *Neurosci. Biobehav. Rev.* **103**, 3–13 (2019).
- Joos, M. Description of language design. *J. Acoust. Soc. Am.* **22**, 701–707 (1950).
- Osgood, C. E., Suci, G. J. & Tannenbaum, P. H. *The Measurement of Meaning* (University of Illinois Press, 1978).
- Harris, Z. S. Distributional structure. *WORD* **10**, 146–162 (1954).
- Mitchell, T. M. et al. Predicting human brain activity associated with the meanings of nouns. *Science* **320**, 1191–1195 (2008).
- Palatucci, M., Pomerleau, D., Hinton, G. E. & Mitchell, T. M. Zero-shot learning with semantic output codes. In *Advances in Neural Information Processing Systems* (eds Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C. & Culotta, A.) Vol. 22, 1–9 (Curran Associates, Inc., 2009).
- Sudre, G. et al. Tracking neural coding of perceptual and semantic features of concrete nouns. *NeuroImage* **62**, 451–463 (2012).
- Rupp, K. et al. Semantic attributes are encoded in human electrocorticographic signals during visual object recognition. *NeuroImage* **148**, 318–329 (2017).
- Chen, Y. et al. The ‘when’ and ‘where’ of semantic coding in the anterior temporal lobe: temporal representational similarity analysis of electrocorticogram data. *Cortex* **79**, 1–13 (2016).
- Carlson, T. A., Tovar, D. A., Alink, A. & Kriegeskorte, N. Representational dynamics of object vision: the first 1000 ms. *J. Vis.* **13**, 1 (2013).
- Groetswagers, T., Wardle, S. G. & Carlson, T. A. Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J. Cogn. Neurosci.* **29**, 677–697 (2017).
- Ploran, E. J. et al. Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fMRI. *J. Neurosci.* **27**, 11912–11924 (2007).
- Luotolahti, J., Kanerva, J., Laippala, V., Pyysalo, S. & Ginter, F. Towards Universal Web Parsebanks. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, 211–220 (Uppsala University, 2015).
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. Distributed representations of words and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* **26**, 3111–3119 (2013).
- Giari, G., Leonardelli, E., Tao, Y., Machado, M. & Fairhall, S. L. Spatiotemporal properties of the neural representation of conceptual content for words and pictures—an MEG study. *NeuroImage* **219**, 116913 (2020).
- Kubilius, J. et al. CORnet: modeling the neural mechanisms of core object recognition. Preprint at *bioRxiv* <https://doi.org/10.1101/408385> (2018).
- Cichy, R. M., Pantazis, D. & Oliva, A. Resolving human object recognition in space and time. *Nat. Neurosci.* **17**, 455–462 (2014).
- Kaiser, D., Oosterhof, N. N. & Peelen, M. V. The neural dynamics of attentional selection in natural scenes. *J. Neurosci.* **36**, 10522–10528 (2016).
- Kivisaari, S. et al. Reconstructing meaning from bits of information. *Nat. Commun.* **10**, 927 (2019).
- Pereira, F. et al. Toward a universal decoder of linguistic meaning from brain activation. *Nat. Commun.* **9**, 963 (2018).
- Just, M. A., Cherkassky, V. L., Aryal, S. & Mitchell, T. M. A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS ONE* **5**, e8622 (2010).
- Shinkareva, S. V. et al. Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS ONE* **3**, e1394 (2008).
- Huth, A. G., Nishimoto, S., Vu, A. T. & Gallant, J. L. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* **76**, 1210–1224 (2012).
- Huth, A. G. et al. Decoding the semantic content of natural movies from human brain activity. *Front. Syst. Neurosci.* **10**, 81 (2016).
- Deniz, F., Nunez-Elizalde, A. O., Huth, A. G. & Gallant, J. L. The representation of semantic information across human cerebral cortex during listening versus reading Is invariant to stimulus modality. *J. Neurosci.* **39**, 7722–7736 (2019).

37. McCartney, B., Martinez-del Rincon, J., Devereux, B. & Murphy, B. A zero-shot learning approach to the development of brain-computer interfaces for image retrieval. *PLoS ONE* **14**, e0214342 (2019).
38. Hultén, A. et al. The neural representation of abstract words may arise through grounding word meaning in language itself. *Hum. Brain Mapp.* **42**, 4973–4984 (2021).
39. Leonardelli, E., Fait, E. & Fairhall, S. L. Temporal dynamics of access to amodal representations of category-level conceptual information. *Sci. Rep.* **9**, 239 (2019).
40. Deniz, F., Tseng, C., Wehbe, L., Dupré La Tour, T. & Gallant, J. L. Semantic representations during language comprehension are affected by context. *J. Neurosci.* **43**, 3144–3158 (2023).
41. Ghazaryan, G. et al. Trials and tribulations when attempting to decode semantic representations from MEG responses to written text. *Lang. Cogn. Neurosci.* <https://doi.org/10.1080/23273798.2023.2198245> (2023, in press).
42. Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A. & Oliva, A. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.* **6**, 27755 (2016).
43. Simanova, I., van Gerven, M., Oostenveld, R. & Hagoort, P. Identifying object categories from event-related EEG: Toward decoding of conceptual representations. *PLoS ONE* **5**, e14465 (2010).
44. Bo, K. et al. Decoding the temporal dynamics of affective scene processing. *NeuroImage* **261**, 119532 (2022).
45. Ala-Salomäki, H., Kujala, J., Liljeström, M. & Salmelin, R. Picture naming yields highly consistent cortical activation patterns: Test-retest reliability of magnetoencephalography recordings. *NeuroImage* **227**, 117651 (2021).
46. Shao, Z., Roelofs, A. & Meyer, A. S. Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *Q. J. Exp. Psychol.* **65**, 1927–1944 (2012).
47. Alfred, K. L., Hillis, M. E. & Kraemer, D. J. M. Individual differences in the neural localization of relational networks of semantic concepts. *J. Cogn. Neurosci.* **33**, 390–401 (2021).
48. de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S. & Gegenfurtner, K. R. Individual differences in visual salience vary along semantic dimensions. *Proc. Natl Acad. Sci. USA* **116**, 11687–11692 (2019).
49. Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M. & Just, M. A. Commonality of neural representations of words and pictures. *NeuroImage* **54**, 2418–2425 (2011).
50. Ojemann, G., Ojemann, J., Lettich, E. & Berger, M. Cortical language localization in left, dominant hemisphere. An electrical stimulation mapping investigation in 117 patients. *J. Neurosurg.* **71**, 316–326 (1989).
51. Oldfield, R. C. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
52. Taulu, S. & Simola, J. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys. Med. Biol.* **51**, 1759–1768 (2006).
53. Gramfort, A. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, 267 (2013).
54. Jas, M. et al. A reproducible MEG/EEG group study with the MNE software: recommendations, quality Assessments, and good practices. *Front. Neurosci.* **12**, 1–18 (2018).
55. Fischl, B. FreeSurfer. *NeuroImage* **62**, 774–781 (2012).
56. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. *NeuroImage* **9**, 179–194 (1999).
57. Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis. *NeuroImage* **9**, 195–207 (1999).
58. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
59. Kriegeskorte, N. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
60. van Vliet, M. MNE-RSA: representational similarity analysis. <https://users.aalto.fi/vanvliet/mne-rsa/> (2022).
61. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
62. Ghazaryan, G. et al. Analysis code relating to Ghazaryan et al. 2023: Cortical time-course of evidence accumulation during semantic processing. <https://zenodo.org/doi/10.5281/zenodo.10076376> (2023).
63. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
64. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
65. Singmann, H., Bolker, B., Westfall, J., Aust, F. & Ben-Shachar, M. S. afex: Analysis of factorial experiments. <https://CRAN.R-project.org/package=afex> (2023).
66. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
67. Storey, J. D., Bass, A. J., Dabney, A. & Robinson, D. qvalue: Q-value estimation for false discovery rate control. <https://github.com/jdstorey/qvalue> (2020).
68. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).

## Acknowledgements

We would like to thank Jenna Kanerva and Filip Ginter at the University of Turku for development of the Finnish language word2vec model. We also thank Ali Faisal for assistance in the early stages of the project and Anni Nora for helpful comments on a draft. Computational resources were provided by the Aalto Science-IT project. This research was funded by the Research Council of Finland (#315553 and #355407 to R.S., #310988 and #343385 to M.v.V., #286070 to S.K., #287474 to A.H.), the Sigrid Jusélius Foundation (to R.S.), and the Finnish Cultural Foundation (to G.G.).

## Author contributions

Conceptualization: A.H. and R.S.; data collection: L.L. and A.H.; methodology: G.G., M.v.V., L.L., T.L.-K., S.L.K., A.H., and R.S.; interpretation: G.G., M.v.V., L.L., T.L.-K., S.L.K., A.H., and R.S.; writing—original draft, G.G.; writing—review & editing, G.G., M.v.V., L.L., T.L.-K., S.L.K., A.H., and R.S.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-023-05611-6>.

**Correspondence** and requests for materials should be addressed to Gayane Ghazaryan.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Stefano Palminteri, Gene Chong, and Luke R. Grinham.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023