How our brains utilize real-world structures to create coherent visual experiences

Dissertation

Zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

am Fachbereich Erziehungswissenschaft und Psychologie
der Freien Universität Berlin

Vorgelegt von
Lixiang Chen, M.Sc.

Berlin, 2024

First reviewer:

Prof. Dr. Radoslaw Martin Cichy

Second reviewer:

Prof. Dr. Surya Gayet

Date of defense: 02.10.2024

Contents

Acknowledgements

First of all, I would like to express my deep gratitude to my supervisor, Prof. Radoslaw Martin Cichy, for providing me with the opportunity to pursue my PhD in his lab. I sincerely appreciate his support and guidance throughout my doctoral studies.

I am deeply grateful to my second supervisor, Prof. Daniel Kaiser, for his guidance, patience, and encouragement during my PhD. I greatly appreciate the time and effort he dedicated to my research projects.

I would like to thank all the members of Cichy's lab and Kaiser's lab for their valuable suggestions and feedback on my research. I also want to thank Daniela Satici-Thies for her assistance with administrative questions.

I would like to thank Prof. Surya Gayet, Prof. Felix Blankenburg, and Dr. Jana Lüdtke for kindly agreeing to join my doctoral committee. I am also grateful to Prof. Surya Gayet for taking the time to review my dissertation.

I am grateful to the Freie Universität Berlin - China Scholarship Council Program for Doctoral Researchers (FUB-CSC Program) for financial support. I also appreciate the HPC Service of FUB-IT, Freie Universität Berlin, for providing computing resources.

Finally, I would like to express my heartfelt appreciation to my parents and my sister for their unconditional love and support. Many thanks to Xu for accompanying me throughout this journey.

Abstract

We live in a structured world, where objects rarely exist in isolation but are often surrounded by similar environments. When objects consistently co-occur with certain objects and scene contexts, our neural systems can implicitly extract and learn such regularities in real-world environments. Predictive processing theories propose that our brains can use learned statistical regularities to predict the structure of incoming sensory input across space and time during visual processing. The predictions may allow us to efficiently recognize objects and understand scenes, thus forming coherent visual experiences in natural vision.

In this dissertation, we conducted three studies to explore how our brains use real-world structures to create coherent visual experiences using neuroimaging techniques (EEG & fMRI) and multivariate pattern analyses (MVPA). Study 1 investigated how scene context affects object processing across time by recording EEG signals while participants viewed semantically consistent or inconsistent objects within scenes. The results reveal that semantically consistent scenes facilitate object representations, but this facilitation is task-dependent rather than automatic. In Study 2, we investigated how cortical feedback mediates the integration of visual information across space by manipulating the spatiotemporal coherence of naturalistic video stimuli shown in both visual hemifields. By analytically combining EEG and fMRI data, we demonstrated that spatial integration of naturalistic visual inputs is mediated by cortical feedback in alpha dynamics that fully traverse the visual hierarchy. In Study 3, we further investigated what level of spatiotemporal coherence is needed to trigger such integration-related alpha dynamics. The findings suggest that integration-related alpha dynamics have some flexibility so that they can accommodate information from videos

belonging to the same basic-level category. Together, the dissertation provides multimodal evidence demonstrating that contextual information facilitates object perception and scene integration, highlighting the critical role of predictions related to real-world regularities in constructing coherent visual experiences.

Zusammenfassung

Wir leben in einer strukturierten Welt, in der Objekte selten isoliert existieren, sondern oft von ähnlichen Umgebungen umgeben sind. Wenn Objekte konsequent mit bestimmten Objekten und Szenenkontexten zusammen auftreten, können unsere neuronalen Systeme solche räumlich-zeitlichen Regelmäßigkeiten in realen Umgebungen implizit extrahieren und erlernen. Theorien zur prädiktiven Verarbeitung gehen davon aus, dass unser Gehirn erlernte statistische Regelmäßigkeiten nutzen kann, um die Struktur eingehender sensorischer Eingaben über Raum und Zeit während der visuellen Verarbeitung vorherzusagen. Die Vorhersagen könnten es uns ermöglichen, Objekte effizient zu erkennen und Szenen zu verstehen und so kohärente visuelle Erfahrungen im natürlichen Sehen zu bilden.

In dieser Dissertation haben wir drei Studien durchgeführt, um zu untersuchen, wie unser Gehirn reale Strukturen nutzt, um kohärente visuelle Erfahrungen zu schaffen, indem wir Techniken des Neuroimaging (EEG und fMRT) sowie multivariate Musteranalysen einsetzten. Studie 1 untersuchte, wie der Szenenkontext die Objektverarbeitung über die Zeit hinweg beeinflusst, indem EEG-Signale aufgezeichnet wurden, während die Teilnehmer semantisch konsistente oder inkonsistente Objekte in Szenen betrachteten. Die Ergebnisse zeigen, dass semantisch konsistente Szenen die Objektrepräsentation erleichtern; diese Erleichterung ist jedoch eher aufgabenabhängig als automatisch. In Studie 2 untersuchten wir, wie kortikales Feedback die Integration visueller Informationen über den Raum hinweg vermittelt, indem wir die räumlich-zeitliche Kohärenz naturalistischer Videostimuli in beiden visuellen Hemifeldern manipulierten. Durch die analytische Kombination von EEG- und fMRI-Daten konnten wir zeigen, dass die räumliche Integration naturalistischer visueller Inputs durch kortikales Feedback in einer Alpha-Dynamik vermittelt wird, die

die visuelle Hierarchie vollständig durchläuft. In Studie 3 untersuchten wir weiter, welches Maß an raum-zeitlicher Kohärenz erforderlich ist, um solche integrationsbezogenen Alphadynamiken auszulösen. Die Ergebnisse deuten darauf hin, dass integrationsbezogene Alphadynamiken eine gewisse Flexibilität aufweisen, so dass sie Informationen aus Videos aufnehmen können, die zur gleichen Basiskategorie gehören. Zusammenfassend liefert die Dissertation multimodale Beweise dafür, dass Kontextinformationen die Objektwahrnehmung und Szenenintegration erleichtern, und unterstreicht die entscheidende Rolle von Vorhersagen im Zusammenhang mit Regelmäßigkeiten in der realen Welt bei der Konstruktion kohärenter visueller Erfahrungen.

List of abbreviations

EEG: electroencephalography

ERP: event-related potential

EVC: early visual cortex (V1, V2, V3)

fMRI: functional magnetic resonance imaging

hMT: human middle temporal complex

IT: inferior temporal cortex

LDA: linear discriminant analysis

LOC: lateral occipital complex

MPA: medial place area

MVPA: multivariate pattern analysis

OPA: occipital place area

PCA: principal component analysis

pRF: population receptive field

PPA: parahippocampal place area

RDM: representational dissimilarity matrix

RSA: representational similarity analysis

SVM: support vector machine

TMS: transcranial magnetic stimulation

V1: primary visual cortex

List of original research articles

1. Chen, L., Cichy, R. M., & Kaiser, D. (2022). Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations. Cerebral Cortex, 32(16), 3553–3567. https://doi.org/10.1093/cercor/bhab433

2. Chen, L., Cichy, R. M., & Kaiser, D. (2023). Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision. Science Advances, 9(45), eadi2321. https://doi.org/10.1126/sciadv.adi2321

3. Chen, L., Cichy, R. M., & Kaiser, D. (2024). Coherent categorical information triggers integration-related alpha dynamics. Journal of Neurophysiology, 131(4), 619–625. https://doi.org/10.1152/jn.00450.2023

1 Introduction

The real world is structured. For instance, objects repeatedly co-occur with other objects (e.g. a computer with a keyboard) and appear in specific positions in scene contexts (e.g., a computer in the office). Over time, as we navigate life, information relating to spatiotemporal regularities in real-world environments is implicitly encoded in our information processing systems. Predictive processing theories cast visual perception as a process of probabilistic, knowledge-driven inference (Bastos et al., 2012; Friston, 2005; G. B. Keller & Mrsic-Flogel, 2018; Rao & Ballard, 1999; Walsh et al., 2020). Accordingly, predictions based on learned real-world statistical regularities should guide our visual perception and facilitate the construction of coherent perceptual experiences.

Statistical regularities may yield robust predictions regarding what objects tend to appear in what scenes, facilitating our object recognition and scene understanding in natural vision (Bar, 2004; Oliva & Torralba, 2007). For instance, we identify an object more quickly and accurately when it is within a consistent scene, e.g., a computer in the office, than within an inconsistent scene, e.g., a computer in the bathroom (Davenport, 2007; Davenport & Potter, 2004; Munneke et al., 2013). Statistical regularities also allow our sensory system to make reliable predictions about the structure of visual inputs across space and time (Bar, 2009; Goettker et al., 2021; Henderson, 2017; Kaiser & Cichy, 2021), facilitating our brains to integrate visual information from different locations into a whole percept. The integration enables us to efficiently understand the meaning of scenes and events in natural vision. In this dissertation, we investigated how real-world statistical regularities facilitate our brains to construct coherent visual experiences.

1.1 Predictive processing in visual perception

In recent years, predictive processing has emerged as an increasingly influential model of how the brain processes sensory information. Traditionally, visual perception has been conceptualized as a hierarchical feedforward process, starting from light-sensitive cells in the retina, progressing through simple contrast and edge detectors in early visual regions, and subsequently forming more complex representations in higher visual areas (DiCarlo et al., 2012; DiCarlo & Cox, 2007; Riesenhuber & Poggio, 1999). Recent predictive processing theories challenge this view, casting that visual processing relies on the dynamic convergence of bottom-up stimulus analysis and top-down predictions generated by internal models (Bastos et al., 2012; Friston, 2005; G. B. Keller & Mrsic-Flogel, 2018; Rao & Ballard, 1999; Walsh et al., 2020). Specifically, predictive processing theories posit that the brain contains internal generative models that continuously predict the sensory input it receives. These predictions are then sent to lower-level areas of the visual hierarchy, suppressing congruent sensory signals, and allowing only residual, unexplained sensory information to be forwarded to higher-level regions as prediction errors. Subsequently, the generative models utilize these prediction-error signals to adjust the probability assigned to perceptual hypotheses, iterating across all hierarchical levels until the network achieves a coherent representation of the sensory input. In this framework, perception is proposed as the process of identifying the perceptual hypothesis that most accurately predicts sensory input and minimizing prediction errors.

Recent studies provide some neural evidence supporting the predictive processing theories using invasive and non-invasive techniques. In the predictive processing framework, prior expectations may initiate stimulus-related predictions before the input

of sensory information (Wyart et al., 2012). In line with this, Bell et al. (2016) observed that the expected stimulus was decodable from neural activity in the inferior temporal cortex (IT) of macaques before stimulus onset, using multivariate pattern analyses (MVPA), in a single-unit recording study. In addition, recent studies have shown prediction responses in the sensory cortex when bottom-up input is absent. Kanisza illusion was often used, wherein the illusion is induced by circles with missing wedges ("Pac-Man" inducers). In a single-unit recordings study, Lee and Nguyen (2001) presented monkeys with illusory contours in the receptive fields of V1 and V2 neurons. They found that both V1 and V2 neurons responded to the illusory contours, with V2 neurons showing consistent responses earlier. Similarly, Kok and de Lange (2014) reported increased neural activity in the subregion of V1 retinotopically corresponding to the illusory contours using fMRI and population receptive field (pRF) mapping. These results suggest there are top-down predictions to V1 from higher-order regions which encompass the entire shape. Furthermore, fMRI studies using lower-right quadrant occluded natural scenes (Morgan et al., 2019; Muckli et al., 2015; Smith & Muckli, 2010), have shown that scene information is decodable from response patterns in the non-stimulated region of V1 (i.e., retinotopically mapped to the occluded quadrant), suggesting predictive feedback to V1 from higher-order regions containing a representation of the whole scene image.

A main challenge for testing predictive processing theories is to strictly differentiate between top-down predictions and bottom-up processing signals. Early animal studies have indicated that feedforward signals arrive in the middle layers, while top-down feedback primarily targets deep and superficial layers, bypassing the middle layers (Felleman & Van Essen, 1991; Harris & Mrsic-Flogel, 2013; Rockland & Pandya, 1979). Recent evidence supports that feedforward and feedback information may also

appear at different cortical depths in the human brain using non-invasive laminar fMRI (Kok et al., 2016; Muckli et al., 2015; Self et al., 2019). These suggest that feedforward and feedback signals may be separated at different cortical depths. Another potential approach for achieving the differentiation is based on brain rhythms. Feedforward and feedback information flows were proposed to be coded by oscillatory activity in different frequency bands: high-frequency gamma rhythms mediate bottom-up feedforward propagation, whereas low-frequency alpha/beta rhythms carry predictive feedback to low-level areas (Bastos et al., 2015; Fries, 2015; Michalareas et al., 2016; van Kerkoerle et al., 2014). Supporting evidence has been shown in previous animal and human studies. In a monkey study, Kerkoerle et al. (2014) recorded neuronal activity in different layers of visual areas and investigated the propagation of alpha and gamma activities across cortical layers. They observed that gamma rhythms first emerged in the middle layer (layer 4) and then propagated to the superficial and deep layers, whereas alpha activity propagated in the opposite direction: from superficial (layers 1, 2) and deep layers (layer 5) to layer 4. Furthermore, they found that gamma rhythms propagated from V1 to V4, whereas the alpha rhythms propagated in the opposite direction. Their findings suggest that gamma activity is a signature of feedforward information propagation, while alpha activity is feedback signaling. Similar effects of alpha and gamma rhythms in visual perception were observed by Bastos et al. (2015) and Michalareas et al. (2016) using primate animals. In addition, using non-invasive electroencephalography (EEG), some recent studies found shared object/scene representations in the alpha frequency band between imagery and perception in the human brain (Stecher & Kaiser, 2023; Xie et al., 2020), suggesting alpha rhythms mediate top-down processing during imagery. The studies suggest EEG

oscillatory activities in different frequency bands may be used to distinguish feedback and feedforward information flows in the human brain.

1.2 Scene semantics

Objects usually occur in specific scenes in the real world. As "objects in scenes" is similar to "words in sentences", semantics have been used to describe the scene-object relationship, determining whether an object aligns with the overall meaning of a scene (Võ et al., 2019).

Behavioral studies have shown that semantic relations between objects and scenes influence object identification. Early studies using line drawings as stimuli, where they paired target objects with consistent or inconsistent scenes, showed that objects were detected more quickly and accurately when they were in a consistent scene (Biederman et al., 1982; Boyce et al., 1989; Boyce & Pollatsek, 1992; Palmer, 1975). Recent studies using scene photographs have reported similar facilitation effects (Davenport, 2007; Davenport & Potter, 2004; Munneke et al., 2013). Consistent with these findings, eye-tracking studies have indicated that inconsistent objects received longer and more frequent fixations compared to consistent objects (Cornelissen & Võ, 2017; Võ & Henderson, 2009, 2011), suggesting that objects are perceived more rapidly within a consistent scene. Similar behavioral facilitation effects emerge when the scene, rather than the object, is task-relevant. Davenport et al. (2007; 2004) found that scenes were identified with higher accuracy when presented with a consistent foreground object than with an inconsistent object. These behavioral results suggest that objects and scenes are processed in a highly interactive manner.

Recent neuroimaging studies provide neural evidence for the interactions between object and scene processing in the brain (Brandman & Peelen, 2017, 2019; Wischnewski & Peelen, 2021). In an fMRI study, Brandman and Peelen (2017) presented participants with images containing degraded objects in scenes, degraded objects alone, and scenes alone, to investigate context-based object perception in the brain. They found that scenes enhanced the representations of degraded objects in object-selective lateral occipital cortex (LOC), and the facilitatory effect in LOC was correlated with activity in the scene-selective regions: occipital place area [OPA; or transverse occipital sulcus (TOS)], medial place area [MPA; or retrosplenial complex (RSC)], and parahippocampal place area (PPA). Similarly, using degraded scenes with objects, degraded scenes alone, and objects alone as stimuli, they explored the effect of objects on scene processing (Brandman & Peelen, 2019). They found a significant improvement of scene representations in the scene-selective cortex: OPA and PPA, when objects were presented. Additionally, in a recent transcranial magnetic stimulation (TMS) study, Wischnewski and Peelen (2021) stimulated early visual cortex (EVC), LOC, and OPA separately at different times after stimulus onset to further investigate the causal role of these regions in context-based object processing. They found that OPA was causally involved in context-based object processing at 160–200 ms, and LOC was involved later at 260–300 ms, after stimulus onset. The results suggest that scene information represented in the OPA may provide feedback to LOC to facilitate object representations. These studies have provided evidence for interactive facilitation between scenes and objects when they are relevant to the current task demands. In Study 1, we further explored whether the facilitatory effect is task-dependent or automatic.

In addition, recent electrophysiological studies have shown that N300 and N400 event-related potential (ERP) components are relevant to semantic violations between objects

and scenes. These studies showed objects within consistent or inconsistent scenes in either a sequential or simultaneous way (Coco et al., 2020; Draschkow et al., 2018; Ganis & Kutas, 2003; Mudrik et al., 2010; Võ & Wolfe, 2013). Using a sequential design, Võ and Wolfe (2013) presented a scene preview and then showed a consistent (e.g., a computer mouse on an office table) or an inconsistent object (e.g., a bar of soap on an office table) in an appropriate location within the scene. They found that objects in inconsistent scenes evoked higher N300 and N400 responses than objects in consistent scenes. Studies employing a simultaneous design, where participants viewed scene images containing either a consistent or an inconsistent object (Mudrik et al., 2010, 2014; Truman & Mudrik, 2018), have reported similar N300 and/or N400 effects. As the N400 component is often found in language studies, the scene-object consistency related N400 was interpreted as more conceptual, semantic processing in previous studies (Mudrik et al., 2010, 2014; Võ & Wolfe, 2013). In contrast, the earlier N300 effect is often linked to differences in perceptual processing between typically and atypically positioned objects (Kumar et al., 2021; Mudrik et al., 2010; Schendan & Maher, 2009). On this view, scene-object consistency related differences in the N300 component arise from differences in the visual analysis of objects and scenes, rather than from post-perceptual processing of (in)consistency. If such differences in the N300 waveform indeed index changes in perceptual processing, the N300 effect should be accompanied by differences in the neural representation of consistent and inconsistent objects. This question was explored in Study 1.

1.3 Visual integration

Wherever we go, our brains constantly integrate complex visual inputs from the environment into a unified whole. Thus, we always have seamless and coherent visual

experiences. Visual integration is a crucial perceptual process involving the ability to link individual local attributes of a scene to form a larger, more complex global structure. The integration allows us to efficiently understand the meaning of scenes and events in natural vision, guiding our behavior.

In our visual system, sensory inputs cannot be initially integrated across space, as early visual areas have limited receptive fields and can only access spatially confined regions of the visual space. Therefore, these inputs must be spatially integrated at higher-order regions of the visual hierarchy to form more spatially extensive global aspects of the environment. Previous studies have probed neural integration of local edges into glocal shapes using a spatial array of oriented edges in which the edges can or cannot be integrated into a global form (Altmann et al., 2003; Kourtzi et al., 2003; Mannion et al., 2013). Real-world inputs are more complex and contain richer information. It is unclear how our brains integrate information across space in natural vision.

Given the highly predictable spatial (Geisler, 2008; Kaiser, Quek, et al., 2019; Kaiser & Cichy, 2021; Võ et al., 2019) and temporal (Goettker et al., 2021; Hogendoorn, 2022) structure of natural visual inputs, predictive processing provides particularly pervasive accounts for scene recognition (Bar, 2004, 2009; Lange et al., 2018), natural visual exploration (Goettker et al., 2021; Henderson, 2017), and neural scene representation (Kaiser, Turini, et al., 2019; Kaiser & Cichy, 2021; Naselaris et al., 2009). Although research agrees on the importance of cortical feedback for visual integration and the generation of coherent precepts, it is still unclear how such feedback is mechanistically implemented in the brain. In Studies 2 and 3, we probed this question using fMRI and EEG.

1.4 Research questions

The dissertation investigated how our brains utilize real-world regularities to create coherent visual experiences. We focused on the neural correlates of context-based object processing and global scene integration. Specifically, we explored the following questions: 1) how scenes facilitate cortical object representations and whether the facilitatory effect is automatic; 2) how visual inputs are integrated across space in the brain and whether cortical feedback signals mediate the integration; 3) what is required to trigger spatial integration in the brain, separately in three studies.

In Study 1, we recorded EEG signals, while participants viewed scenes and consistent/inconsistent objects sequentially in two experiments: task-relevant (color discrimination) vs. task-relevant (object recognition). We probed whether the differences in the N300/N400 components between consistent and inconsistent scene-object combinations reflect differences in the cortical representation of objects, and whether the facilitations between scenes and objects are automatic or task-dependent, using both univariate analyses and multivariate decoding analyses on the EEG-evoked response patterns.

In Study 2, we manipulated the degree to which stimuli could be integrated across space through the spatiotemporal coherence of naturalistic video stimuli shown in the two visual hemifields, in both EEG and fMRI experiments. We probed how visual inputs are integrated across space and whether the integration process is mediated by rhythmic feedback signals in the brain, using multivariate decoding analyses on fMRI multi-voxel patterns and EEG spectral patterns.

In Study 3, we manipulated the degree of spatiotemporal coherence of stimuli and presented naturalistic stimuli using the paradigm from Study 2 in an EEG experiment. We further investigated what level of spatiotemporal coherence in the stimuli is needed

to trigger integration-related rhythmic signals using multivariate decoding analyses on EEG spectral patterns.

2 Summary of dissertation studies

In this chapter, I will summarize the three studies on which this dissertation is based.

2.1 Study 1: Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations (Chen et al., 2022)

Previous studies have shown that consistent scenes enhance the cortical representations of objects (Brandman & Peelen, 2017; Wischnewski & Peelen, 2021), and there are differences in the EEG-evoked N300/N400 components between consistent and inconsistent scene-object combinations (Coco et al., 2020; Draschkow et al., 2018; Ganis & Kutas, 2003; Mudrik et al., 2010; Võ & Wolfe, 2013). In this study, we conducted two EEG experiments to explore whether the scene-object N300/N400 differences reflect changes in the cortical representations of objects, and whether the facilitations between objects and scenes are automatic.

We selected scene images from eight categories, which were grouped into four pairs. For each scene pair, we chose four objects: two were consistent with one scene category (e.g., computers & printers with offices), and two were consistent with another scene category (e.g., microwaves & rice cookers with kitchens). We recorded EEG signals while participants viewed scene and object images during the experiments. Each trial began with a central fixation point, followed by a scene image (without the critical object). Then, a red dot appeared in the scene, indicating where the critical object would appear. Subsequently, we presented either a consistent or an inconsistent object at the dot position. Participants performed a color discrimination task in the first experiment (task-irrelevant). In some trials, the red dot changed to blue instead of the object

appearing. Participants were instructed to press a button when they detected the color change. In the second experiment (task-relevant), we used an object recognition task. After the object disappeared, we presented an object exemplar on the screen, either the same or a different exemplar from the same basic-level category. Participants were instructed to judge whether it was the exemplar they had seen in the trial.

To replicate the previous scene-object N300/N400 effects (Coco et al., 2020; Draschkow et al., 2018; Ganis & Kutas, 2003; Mudrik et al., 2010; Võ & Wolfe, 2013), we chose nine mid-central channels (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, and CP2) and compared averaged evoked responses across nine channels at each time point (from -100 to 800 ms relative to object onset) between consistent and inconsistent scene-object combinations, using preprocessed EEG data. We replicated previous findings that inconsistent scene-object combinations evoked stronger N300 and N400 components in both experiments.

Next, we investigated the influence of scenes on object representations across time using both timeseries decoding (Boring et al., 2020; Kaiser & Nyga, 2020) and cumulative decoding analyses (Kaiser et al., 2020b; Ramkumar et al., 2013) on EEG-evoked response patterns. Timeseries decoding was performed using data from a sliding time window (50 ms), and cumulative decoding was conducted using aggregated data from all time points before the current time point in the epoch. Specifically, we decoded between two consistent objects (e.g., computers & printers) or two inconsistent objects (e.g., microwaves & rice cookers) within each scene category (e.g., offices) using evoked response patterns across channels. In both experiments, we found consistent and inconsistent objects were decodable from ~100 ms after object onset. Consistent and inconsistent objects were equally decodable in Experiment 1 (task-irrelevant), but

consistent objects were decoded better than inconsistent objects when objects were useful to the task (Experiment 2). Using a 2 (task-irrelevant vs. task-relevant objects) × 2 (ERP vs. object decoding) mixed ANOVA, we found the effect of task relevance was significantly larger in the cumulative decoding analysis than it was in the ERP analysis.

Similarly, we also performed decoding analyses to investigate whether objects affect scene representations. Specifically, we decoded between every two scene categories separately for the consistent and inconsistent conditions. In both experiments where scenes were task-irrelevant, significant decoding for both consistent and inconsistent scenes emerged from ~50 ms after scene presentation, but no significant differences were found between consistent and inconsistent scenes.

In summary, our results showed that differences in the N300/N400 components are accompanied by differences in the object decoding between consistent and inconsistent only in the task-relevant experiment but not in the task-irrelevant experiment, suggesting that the differences in the N300/N400 components do not reflect differences in perceptual object representations. Furthermore, the decoding effects of objects and scenes across two experiments suggest that the facilitations between objects and scenes are task-dependent rather than automatic.

2.2 Study 2: Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision (Chen et al., 2023)

Our brains must integrate sensory inputs across visual fields to create coherent visual experiences. On the classic hierarchical view of the human visual system, such integration is thought to be solved during feedforward processing (DiCarlo et al., 2012;

DiCarlo & Cox, 2007; Riesenhuber & Poggio, 1999). Here, we challenge this view, asking whether the integration is mediated by cortical feedback.

We selected eight natural original videos (3 seconds) as stimuli. In both EEG and fMRI experiments, we manipulated the degree to which stimuli could be integrated across space by presenting original videos through one aperture or two apertures left and right of the central fixation. We designed four conditions: left-only, right-only, consistent, and inconsistent. In the left- and right-only conditions, we presented videos through only one aperture, which can provide a baseline for processing visual inputs from one hemifield. In the consistent condition, we presented the same original video through two apertures, which could be easily integrated into a unified percept; in the inconsistent condition, the two parts were from different original videos, which could not be integrated into a unified percept. In each trial, we presented a central fixation for 0.5 seconds, then showed the videos through one or two apertures, as mentioned above, for 3 seconds. During the video, the central fixation changed color every 200 ms, and we asked participants to maintain central fixation and judge whether a yellow or green dot was included in the sequence. We used this paradigm to present stimuli in both EEG and fMRI experiments.

We first probed whether visual integration is mediated by rhythmic feedback signals using EEG data. Previous studies proposed that high-frequency gamma rhythms propagate bottom-up feedforward information, whereas low-frequency alpha/beta rhythms carry feedback signals from higher-order regions (Bastos et al., 2015; Fries, 2015; Michalareas et al., 2016; van Kerkoerle et al., 2014). Accordingly, we probed our question using EEG decoding analysis on spectral patterns in different frequency bands. Specifically, after preprocessing, we performed spectral analyses using the fast Fourier

transform (FFT) to obtain the power value of each frequency (alpha: 8–12 Hz; beta: 13–30 Hz; gamma: 31–70 Hz) separately for each EEG channel in each trial. We then decoded between the eight video stimuli within each condition using spectral power patterns across channels separately for alpha, beta, and gamma frequency bands. We found that video stimuli in the single-video conditions (left-/right-only) were decodable only from gamma activity. In the two-video conditions, inconsistent video stimuli were only decodable in the gamma band, whereas consistent stimuli were only decodable from alpha activity. There were significant differences between consistent and inconsistent conditions in these two frequency bands. The results indicate that when consistent visual inputs allow for integration into a unified percept, stimulus information is coded in the feedback-related alpha activity; when inconsistent inputs can not be integrated, stimulus information is coded in the feedforward-related gamma activity.

Next, we localized the integration-related regions using preprocessed fMRI data. We defined seven regions of interest (ROIs): three from early visual cortex (V1, V2, V3), one motion-selective region (hMT), and three from scene-selective cortex (OPA, MPA, PPA). We decoded between the eight video stimuli within each condition using multi-voxel patterns in each ROI separately for each brain hemisphere. The decoding results were averaged across hemispheres. We found that single-video stimuli were decodable in all seven regions, and the effects were only shown in contralateral brain regions, except for the hMT. Both consistent and inconsistent stimuli were decodable in all regions. Critically, consistent stimuli were decoded better than inconsistent stimuli in the MPA and PPA, and a similar trend was observed in the hMT. Using searchlight decoding analysis, we probed the effects in the whole brain. We found that the decoding differences between consistent and inconsistent conditions were only located in the

regions overlapping or close to hMT and scene-selective cortex. The results indicate that hMT and scene-selective cortex (MPA and PPA) aggregate spatiotemporally consistent information across hemifields, suggesting that these regions are likely the source of the integration-related feedback.

Finally, we linked the EEG spectral representations with fMRI spatial representations using representational similarity analysis (Cichy et al., 2014; Cichy & Oliva, 2020; Kriegeskorte et al., 2008). We first performed pair-wise decoding analysis between each pair of stimuli within each condition to construct EEG representational dissimilarity matrices (RDMs) and fMRI RDMs in each subject. To increase the power, we averaged the fMRI RDMs across subjects, and then correlated the group-averaged fMRI RDMs in each region with subject-specific EEG RDMs in each frequency band. We found that the consistent condition had a higher correlation between alpha activity and representations in V1, and a similar trend was observed in V2 and V3. In our design, V1 received identical feedforward sensory inputs across consistent and inconsistent conditions; therefore, the correspondence difference observed here should be attributed to the feedback from higher-order regions that have access to both contralateral and ipsilateral information.

In summary, Study 2 suggests that feedback-related alpha activity mediates visual integration across space. When spatiotemporal consistent inputs allow for integration into a unified percept, cortical feedback in the alpha frequency band reaches the earliest stage of cortical visual processing.

2.3 Study 3: Coherent categorical information triggers integration-related alpha dynamics (Chen et al., 2024)

We demonstrated that feedback-related alpha dynamics mediate the spatial integration of visual inputs in Study 2. Here, we further investigated what level of spatiotemporal coherence is required to trigger integration-related alpha dynamics using the same paradigm in an EEG experiment.

We selected a set of natural videos (3 s) and manipulated the spatiotemporal coherence of stimuli by presenting videos through two apertures left and right of the central fixation. We designed four conditions: video-level consistent, basic-level consistent, superordinate consistent, and inconsistent. Specifically, in the video-level consistent condition, two parts of stimuli were from the same video; in the basic-level consistent condition, the two parts were from two different videos belonging to the same basic category; in the superordinate consistent condition, the two parts were from two different videos belonging to the same superordinate category; and in the inconsistent condition, the two parts were from two very different videos (different superordinate categories). We presented the stimuli in the same manner as we did in Study 2.

We first performed spectral analysis on the preprocessed EEG data. Then, we extracted the spectral power patterns across channels to decode between video stimuli within each condition separately for the alpha (8–12 Hz), beta (13–30 Hz), and gamma (31–70 Hz) frequency bands. We found that stimuli in the superordinate consistent and inconsistent conditions were not decodable from the activity in all three frequency bands. Video-level consistent and basic-level consistent stimuli were decodable from alpha activity, and the stimuli in these two conditions were decoded better than the stimuli in the superordinate consistent and inconsistent conditions in the alpha frequency band. The stimuli in all conditions were decodable from EEG broadband response patterns in the first 500 ms, indicating that the difference in alpha decoding across conditions was

unrelated to an absence of stimulus representation in the more incoherent conditions. In addition, we did not observe similar effects when we performed decoding analysis on spectral phase patterns, suggesting that the integration-related alpha effects were specific to spectral power.

In summary, Study 3 suggests that rhythmic alpha activity mediates spatial integration, and such integration exhibits some flexibility so that broadly consistent videos from the same category can trigger integration-related alpha dynamics.

3 Discussion

3.1 Summary

This dissertation investigated how real-world structures facilitate our brains' construction of coherent visual experiences. Specifically, we probed the neural correlates of context-based object perception and global visual integration through three studies.

Study 1 probed how scenes facilitate the representations of consistent objects across time and whether the facilitations are automatic. We reveal that the N300/N400 components related to scene-object consistency do not index perceptual representations; Furthermore, the predictions generated from scene representations enhance object representation in tasks requiring detailed object recognition.

Study 2 investigated how visual inputs from two visual hemifields are integrated in the brain and whether cortical feedback signals mediate the integration. The results suggest that feedback-related alpha dynamics mediate spatial integration of visual inputs. Such feedback activity potentially originates from hMT and scene-selective cortex, targeting early visual cortex.

In Study 3, we further probed what level of spatiotemporal coherence is required to trigger integration-related alpha dynamics. The findings reveal that such integration processes are flexible enough to accommodate information from different exemplars of the same basic-level category.

Together, this dissertation demonstrates that our brains generate predictions about real-world statistical regularties to facilitate local and global visual processing. Such

predictions play a crucial role in the construction of our coherent experiences in natural vision.

3.2 Scene-object consistency related N300 and N400 effects

In Study 1, we successfully replicated previous findings that stronger N300 and N400 responses were evoked by inconsistent scene-object combinations (Coco et al., 2020; Draschkow et al., 2018; Ganis & Kutas, 2003; Mudrik et al., 2010; Võ & Wolfe, 2013). The N300 component is often associated with high-level visual processing, including object recognition (Schendan & Kutas, 2002), shape perception (Schendan & Kutas, 2007), and canonical view of objects (Schendan & Kutas, 2003). In line with this, the N300 difference between consistent and inconsistent scene-object combinations has been previously interpreted as indicative of differences in perceptual processing (Dyck & Brodeur, 2015; Kumar et al., 2021; Mudrik et al., 2010; Sauvé et al., 2017; Schendan & Maher, 2009), particularly through the contextual facilitation theory (Bar, 2004; Bar et al., 2006; Bar & Ullman, 1996). According to this theory, the presentation of a scene can rapidly activate gist-consistent schemas. Subsequently, comparing these schemas and incoming object information facilitates a reduction in perceptual uncertainty during object recognition. When the object does not align with the scene gist, its identification may be impeded, potentially evoking a higher N300 component during perceptual processing.

Our results in Study 1 challenge this interpretation of the N300 effect. Based on contextual facilitation theory, the N300 effect should reflect differences in object representations between consistent and inconsistent conditions. However, in our results, the differences in N300 did not consistently correspond with differences in object representations. Specifically, a higher N300 component was elicited by inconsistent

scene-object combinations in both task-irrelevant and task-relevant designs. In contrast, in multivariate decoding analyses, better decoding for consistent objects was observed only when objects were task-relevant. The differences in the decoding effect were also larger than the differences in the N300 effect between consistent and inconsistent conditions. Therefore, the N300 here may serve as a generic marker of inconsistency or a purely attentional response to a violation of expectation, rather than reflecting perceptual processing. In contrast, the N400 is widely accepted as a signature of post-perceptual semantic processing. It is elicited by written words, pseudo-words, sounds, and mathematical symbols (Kutas & Federmeier, 2011). A recent study has shown that N400 effects are qualitatively similar to N300 effects (Draschkow et al., 2018), further supporting the notion that N300 differences do not directly indicate changes in perceptual processing.

3.3 Interaction between scene and object processing

The results from Study 1 support a task-dependent facilitation between scenes and objects. Specifically, we showed that consistent objects were decoded more accurately when participants performed the object recognition task than the color discrimination task. Additionally, scenes were task-irrelevant in both experiments. We found that scenes containing consistent objects and scenes containing inconsistent objects were decoded equally well in both experiments. In line with our results, previous studies have also reported neural facilitation between scenes and objects when such facilitation was beneficial for the current task demands. For instance, consistent scenes enhanced cortical object representations when participants were asked to memorize the objects (Brandman & Peelen, 2017), and consistent objects improved cortical representations of scene layout during a scene repetition task (Brandman & Peelen, 2019). A recent

study tested cortical object representations under different task demands (Kaiser et al., 2021). They found that spatially consistent scene context facilitated object representation more effectively than spatially inconsistent scene context during an object category task. In contrast, when participants performed a scene-relevant task, object representations were comparable for both spatially consistent and inconsistent scene contexts. Our findings suggest that the visual system flexibly and strategically utilizes contextual information: when contextual information aligns with current task demands, it enhances the representations of scenes and objects.

High-level representations of both objects and scenes were shown to emerge within 200 ms after stimulus onset (Brandman & Peelen, 2023; Cichy et al., 2014). Scene-based enhancement for object representations was observed around 300 ms after stimulus onset in our Study 1 and Brandman & Peelen (2017), while object-based enhancement for scene representations was found at a similar latency in Brandman & Peelen (2023). These findings suggest a more parallel and interactive view of context-based object and scene processes. First, objects and scenes may be processed independently within distinct pathways in the first 200 ms through hierarchical predictive processing. Subsequently, scene-selective and object-selective regions compute semantic representations and mutually influence each other. This interaction between higher-level regions may modulate the feedback propagations to lower-level regions of the visual pathway and further enhance the processing of consistent objects and scenes. The differences between conditions in decoding analysis across experiments in Study 1 suggest that the later interaction between scene-selective and object-selective cortical areas may be modulated by task demands. Specifically, scene representations in the scene-selective cortex may generate predictions about probable objects related to the scene and then send the predictions to object-selective cortex, only when the tasks

require detailed object perception. The predictive mechanism may be less engaged or absent when tasks do not require detailed object perception.

3.4 Rhythmic activities mediate visual integration

Our results suggest that alpha rhythms may encode stimulus-related feedback during integration. Alpha oscillations are the most predominant rhythmic activity in the human brain. Initially considered idling activities, alpha oscillations are most prominent when eyes are closed but suppressed when eyes are open (Pfurtscheller et al., 1996; Romei et al., 2008). Furthermore, they have been associated with functional inhibition, as evidenced by a decrease in alpha power within task-relevant brain regions and an increase within task-irrelevant ones (Jensen & Mazaheri, 2010; Kelly et al., 2006; Romei et al., 2010). Contrary to the passive or inhibitory functions, our results support that alpha oscillations play an active role in cognitive processes. In Studies 2 and 3, we observed spatiotemporally consistent stimuli were decodable from EEG spectral alpha activity. This aligns with studies showing alpha encodes object/scene information during visual perception and imagery (Kaiser, 2022; Stecher & Kaiser, 2023; Xie et al., 2020). Furthermore, we showed that consistent stimuli had a stronger correlation between alpha activity and representations in early visual cortex compared to inconsistent stimuli, suggesting cortical feedback from high-level regions sent to early visual cortex duing integration process. The role of feedback signaling for cortical alpha dynamics was also reported in recent primate and human studies (Bastos et al., 2015; Hetenyi et al., 2024; Michalareas et al., 2016; van Kerkoerle et al., 2014).

Our results suggest that gamma rhythms may encode stimulus-related feedforward information. Gamma rhythms were initially linked to visual grouping and binding (Elliott & Müller, 1998; Gray & Singer, 1989; Tallon-Baudry et al., 1999). Later, they

were associated with attentional selection (Bichot et al., 2005; Fries et al., 2001) and the enhancement of stimulus strength and salience (Friedman-Hill et al., 2000; Swettenham et al., 2009). Drawing from these studies, researchers have proposed that gamma rhythms serve as a mechanism for the feedforward propagation of unpredicted information ("prediction error") within the predictive processing framework (Bastos et al., 2012; Fries, 2015). In Study 2, we showed that when video stimuli were inconsistent across space (single video stimuli and inconsistent stimuli) and could not be integrated into a unified percept, they were only decodable from gamma activity. As the inconsistent stimuli in our experiments do not adhere to typical real-world regularities, the brain is unlikely to accurately predict them based on prior knowledge. Feedforward prediction-error signals may be informative and useful for processing such unexpected stimuli.

In addition, our results suggest that feedback may dominate feedforward processing during the integration process. An absence of representations in the gamma band accompanied the increased involvement of alpha rhythms in coding the consistent visual stimuli in Study 2. An explanation for the lack of decoding from feedforward-related gamma activity in the consistent condition is that accurate top-down predictions for stimuli efficiently suppress feedforward activities. In our experiments, video stimuli were presented over an extended duration, with a lack of rapid or unexpected visual events. The extended presentation may effectively silence feedforward propagation within the gamma activity for the consistent stimuli.

Overall, our results align with the multiplexing hypothesis of predictive processing: feedforward and feedback information flows are coded in high-frequency gamma and low-frequency alpha/beta bands, respectively (Bastos et al., 2015; Fries, 2015;

Michalareas et al., 2016; van Kerkoerle et al., 2014). Unlike previous studies that employed invasive techniques (Bastos et al., 2015; Michalareas et al., 2016; van Kerkoerle et al., 2014), our studies indicate that EEG spectral activities in different frequencies are also sensitive enough to distinguish feedforward and feedback signals in the human brain.

3.5 Feedback traverses the visual hierarchy during visual integration

Higher correspondence between alpha activity and V1 representations in Study 2 suggests that V1 receives feedback information from high-level regions. V1 is the first stage of cortical visual processing, and each hemisphere of V1 processes sensory information from the contralateral visual hemifield. In line with our results, V1 has also been shown to receive various types of feedback during mental imagery (Ragni et al., 2021; Winlove et al., 2018), cross-modal perception (Vetter et al., 2014, 2020), and the ˙interpolation of missing scene information (Morgan et al., 2019; Muckli et al., 2015; Smith & Muckli, 2010). Our results also indicate that the feedback signals emerging in V1 originate from high-level regions with access to the information in both hemifields, suggesting long-distance feedback. This aligns with studies showing contextual signals in V1 exhibit substantial delays compared to feedforward processing (A. J. Keller et al., 2020; Kirchberger et al., 2023; Papale et al., 2022). Such feedback processes may use the spatial resolution of V1 as a flexible sketchpad mechanism (Dehaene & Cohen, 2007; Williams et al., 2008) for recreating detailed feature mappings inferred from the global context.

The fMRI results in Study 2 suggest that scene-selective MPA and PPA, and motion-selective hMT, are potential sources of feedback signals to early visual cortex. MPA and PPA had stronger representations for spatiotemporally consistent stimuli than

inconsistent stimuli. Scene-selective cortex is a logical candidate for providing contextual feedback as it is sensitive to the typical spatial configuration of scene stimuli (Bilalić et al., 2019; Kaiser et al., 2020a; Kaiser & Cichy, 2021; Mannion et al., 2014). The sensitivity allows these regions to generate feedback signals that convey information about whether and how stimuli should be integrated at lower levels of the visual hierarchy. Such feedback signals may arise from adaptively comparing contralateral feedforward information with ipsilateral information from interhemispheric connections. In the inconsistent condition, the ipsilateral information received from the other hemisphere does not match typical real-world regularities and may therefore not trigger the integration. Conversely, when stimuli are consistent, information from other hemisphere becomes critical for facilitating integration across visual hemifields. This idea aligns with previous studies showing increased inter-hemispheric connectivity when object or word information needs to be integrated across visual hemifields (Mima et al., 2001; Stephan et al., 2007).

Motion-selective hMT is another potential region for generating integration-related feedback during dynamic visual integration. Not only did this region show enhanced representations for spatiotemporally consistent stimuli, but it also had representations for both contralateral and ipsilateral visual inputs in the fMRI data of Study 2. The hMT was also shown to be sensitive to motion (Tootell et al., 1995; Watson et al., 1993) and exhibit bilaterally representations for visual information (Cohen et al., 2019). These functions were well-suited for integrating consistent motion patterns across visual hemifields.

Our results suggest that scene-selective and motion-selective cortical areas may jointly generate feedback signals that integrate information about coherent scene content (MPA

and PPA) and coherent motion patterns (hMT). The generated feedback traverses to the earliest stage of cortical visual processing.

3.6 Categorical information triggers integration-related neural dynamics

Our findings in Study 3 demonstrate that integration-related alpha dynamics can be elicited not only by segments from the same video but also by segments from different videos within the same basic-level category. This indicates a spectral signature linked to the category-level feedback information used for spatial integration. According to Rosch (1978), the basic level is the most informative, offering the highest degree of cue validity by maximizing attributes shared within the category while minimizing those shared with other categories. The brain potentially uses this optimal balance during integration processes. Nonetheless, our study can not determine whether integration-related alpha activity is driven by an abstract coherence at the basic-level category, potentially encoded in the high-level visual cortex (Proklova et al., 2016; Walther et al., 2009), or by the spatiotemporal coherence of visual features associated with a category (Coggan et al., 2019, 2022). Clarifying the distinct role of the basic level in integration would necessitate a systematic comparison of integration across basic, superordinate, and subordinate levels.

3.7 Replicability and Reproducibility

Reproducibility is fundamental for scientific research. The past decade has seen a growing concern about the reproducibility and replicability of research findings in cognitive neuroscience (Poldrack et al., 2017). Reproducibility is the capability to obtain the same results using the same data and code, while replicability is the ability to obtain consistent results using different datasets (Nichols et al., 2017). Several factors

may contribute to the low rates of reproducibility and replicability, including flexibility in data analysis and results reporting, low statistical power, and hypotheses being formed based on results.

We uploaded our data and code related to our publications to open science platforms (e.g., OSF, Zenodo), allowing researchers to reanalyze the data and reproduce the results. Additionally, we examined the replicability of EEG effects in the dissertation. First, we replicated the scene-object consistency related N300/N400 ERP effects using a previous paradigm (Võ & Wolfe, 2013) in Study 1. We also examined the replicability of EEG rhythmic results by using the same paradigm and data analyses in Studies 2 and 3. We successfully replicated the decoding effects for the consistent stimuli in the alpha band. These results indicate that broadband response and low-frequency rhythmic activity in the EEG data may be relatively stable.

However, we failed to replicate the gamma effects for the spatiotemporally incoherent stimuli. Several factors may contribute to this. First, variations in participant groups and the use of different EEG systems may impact the detection of gamma effects, given that gamma activity is relatively weak and unreliable in EEG recordings. Second, the two studies employed distinct stimulus selection strategies. Study 2 intentionally maximized incoherence by selecting highly dissimilar videos, whereas Study 3 did not emphasize such dissimilarity in the stimulus selection. Highly spatiotemporally inconsistency in stimuli may be necessary to induce reliable gamma activity.

3.8 Limitations and future directions

In Study 1, the presentation time for the object was different between the two experiments. In the task-irrelevant experiment, it was 500 ms, whereas in the task-

relevant experiment, we reduced it to 83 ms to increase the difficulty of the object recognition task. This raises the possibility that the longer presentation time, rather than the lack of behavioral relevance, in Experiment 1 abolished the decoding effects. Further experiments are needed to establish a clear distinction between these explanations. However, if the N300 and decoding effects reflect the same perceptual representations, both should be affected by presentation time. We observed differences in the decoding effects but not in the ERP effects between experiments. Thus, our effects are unlikely to be attributed to the difference in presentation time.

In Study 2, we speculated that scene-selective cortex and hMT are the sources of the feedback to the early visual cortex during visual integration based on our searchlight and ROI decoding results. However, we did not provide solid evidence to support the speculation. To validate this, future studies could estimate functional connectivity (e.g., psychophysiological interaction, PPI, dynamic causal modeling, DCM) between scene-selective cortex/hMT and early visual areas on fMRI data or use non-invasive brain stimulation techniques (e.g., TMS, tDCS) to stimulate scene-selective cortex/hMT during the integration experiment.

In Studies 2 and 3, we used a fixation task to investigate the neural correlates of automatic integration. Such automatic integration is mainly based on phenomenological experience. However, we did not provide evidence for the occurrence of integration. Future research should explore how integration effects change when participants actively engage with the stimuli that need to be integrated. As the stimuli we used contain information in multiple dimensions (e.g., low-level and mid-level visual features, motion patterns), future studies could separate these and probe the critical features enabling integration and what is integrated.

In Study 1, we investigated context-based object processing. This requires integration between objects and scenes, which are processed in spatially distinct pathways. The interaction between pathways may only occur when objects/scenes are relevant to the task. By contrast, Studies 2 and 3 investigated integration across different parts of scenes. Such integration may occur more automatically because different parts of scenes were processed mainly within the same pathway. Future studies could explore predictions within and between pathways under different tasks to clarify this.

The dissertation has demonstrated that rhythmic feedback signals mediate spatial integration in the brain. To create coherent visual experiences, our brains need to integrate visual information not only across space but also across time. Given that real-world inputs are highly predictable over time, cortical feedback may also play a crucial role in the temporal integration of sensory information. Future research could use spectral EEG and design similar experiments to explore this question.

3.9 Conclusions

This dissertation comprises three studies investigating how our brains utilize real-world regularities to facilitate object recognition and visual integration. Study 1 reveals that contextual information enhances object representations. Studies 2 and 3 demonstrate that top-down predictions in the alpha frequency mediate the integration of visual information across space. Collectively, our findings in the dissertation highlight that predictions based on real-world regularities are crucial for constructing coherent visual experiences.

4 References

Altmann, C. F., Bülthoff, H. H., & Kourtzi, Z. (2003). Perceptual Organization of Local Elements into Global Shapes in the Human Visual Cortex. *Current Biology*, *13*(4), 342–349. https://doi.org/10.1016/S0960-9822(03)00052-6

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. https://doi.org/10.1038/nrn1476

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. https://doi.org/10.1098/rstb.2008.0310

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hämäläinen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454. https://doi.org/10.1073/pnas.0507062103

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, *25*(3), 343–352. https://doi.org/10.1068/p250343

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., De Weerd, P., Kennedy, H., & Fries, P. (2015). Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. *Neuron*, *85*(2), 390–401. https://doi.org/10.1016/j.neuron.2014.12.018

Bell, A. H., Summerfield, C., Morin, E. L., Malecek, N. J., & Ungerleider, L. G. (2016). Encoding of Stimulus Probability in Macaque Inferior Temporal Cortex. *Current Biology*, *26*(17), 2280–2290. https://doi.org/10.1016/j.cub.2016.07.007

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and Serial Neural Mechanisms for Visual Search in Macaque Area V4. *Science*, *308*(5721), 529–534. https://doi.org/10.1126/science.1109676

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143–177. https://doi.org/10.1016/0010-0285(82)90007-X

Bilalić, M., Lindig, T., & Turella, L. (2019). Parsing rooms: The role of the PPA and RSC in perceiving object relations and spatial layout. *Brain Structure and Function*, *224*(7), 2505–2524. https://doi.org/10.1007/s00429-019-01901-0

Boring, M. J., Ridgeway, K., Shvartsman, M., & Jonker, T. R. (2020). Continuous decoding of cognitive load from electroencephalography reveals task-general and task-specific correlates. *Journal of Neural Engineering*, *17*(5), 056016. https://doi.org/10.1088/1741-2552/abb9bc

Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 531–543. https://doi.org/10.1037/0278-7393.18.3.531

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 556–566. https://doi.org/10.1037/0096-1523.15.3.556

Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fMRI and MEG decoding. *Journal of Neuroscience*, *37*(32), 7700–7710. https://doi.org/10.1523/JNEUROSCI.0582-17.2017

Brandman, T., & Peelen, M. V. (2019). Signposts in the fog: Objects facilitate scene representations in left scene-selective cortex. *Journal of Cognitive Neuroscience*, *31*(3), 390–400. https://doi.org/10.1162/jocn_a_01258

Brandman, T., & Peelen, M. V. (2023). Objects sharpen visual scene representations: Evidence from MEG decoding. *Cerebral Cortex*, bhad222. https://doi.org/10.1093/cercor/bhad222

Chen, L., Cichy, R. M., & Kaiser, D. (2022). Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations. *Cerebral Cortex*, *32*(16), 3553–3567. https://doi.org/10.1093/cercor/bhab433

Chen, L., Cichy, R. M., & Kaiser, D. (2023). Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision. *Science Advances*, *9*(45), eadi2321. https://doi.org/10.1126/sciadv.adi2321

Chen, L., Cichy, R. M., & Kaiser, D. (2024). Coherent categorical information triggers integration-related alpha dynamics. *Journal of Neurophysiology*, *131*(4), 619–625. https://doi.org/10.1152/jn.00450.2023

Cichy, R. M., & Oliva, A. (2020). A M/EEG-fMRI Fusion Primer: Resolving Human Brain Responses in Space and Time. *Neuron*, *107*(5), 772–781. https://doi.org/10.1016/j.neuron.2020.07.001

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), Article 3. https://doi.org/10.1038/nn.3635

Coco, M. I., Nuthmann, A., & Dimigen, O. (2020). Fixation-related brain potentials during semantic integration of object–scene information. *Journal of Cognitive Neuroscience*, *32*(4), 571–589. https://doi.org/10.1162/jocn_a_01504

Coggan, D. D., Baker, D. H., & Andrews, T. J. (2019). Selectivity for mid-level properties of faces and places in the fusiform face area and parahippocampal place area. *European Journal of Neuroscience*, *49*(12), 1587–1596. https://doi.org/10.1111/ejn.14327

Coggan, D. D., Watson, D. M., Wang, A., Brownbridge, R., Ellis, C., Jones, K., Kilroy, C., & Andrews, T. J. (2022). The representation of shape and texture in category-selective regions of ventral-temporal cortex. *European Journal of Neuroscience*, *56*(3), 4107–4120. https://doi.org/10.1111/ejn.15737

Cohen, D., Goddard, E., & Mullen, K. T. (2019). Reevaluating hMT+ and hV4 functional specialization for motion and static contrast using fMRI-guided repetitive transcranial magnetic stimulation. *Journal of Vision*, *19*(3), 11. https://doi.org/10.1167/19.3.11

Cornelissen, T. H., & Võ, M. L. (2017). Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception, & Psychophysics*, *79*(1), 154–168. https://doi.org/10.3758/s13414-016-1203-7

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, *35*(3), 393–401. https://doi.org/10.3758/BF03193280

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x

Dehaene, S., & Cohen, L. (2007). Cultural Recycling of Cortical Maps. *Neuron*, *56*(2), 384–398. https://doi.org/10.1016/j.neuron.2007.10.004

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–341. https://doi.org/10.1016/j.tics.2007.06.010

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How Does the Brain Solve Visual Object Recognition? *Neuron*, *73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Draschkow, D., Heikel, E., Võ, M. L., Fiebach, C. J., & Sassenhagen, J. (2018). No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing. *Neuropsychologia*, *120*, 9–17. https://doi.org/10.1016/j.neuropsychologia.2018.09.016

Dyck, M., & Brodeur, M. B. (2015). ERP evidence for the influence of scene context on the recognition of ambiguous and unambiguous objects. *Neuropsychologia*, *72*, 43–51. https://doi.org/10.1016/j.neuropsychologia.2015.04.023

Elliott, M. A., & Müller, H. J. (1998). Synchronous Information Presented in 40-HZ Flicker Enhances Visual Feature Binding. *Psychological Science*, *9*(4), 277–283. https://doi.org/10.1111/1467-9280.00055

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, N.Y.*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1-a

Friedman-Hill, S., Maldonado, P. E., & Gray, C. M. (2000). Dynamics of Striate Cortical Activity in the Alert Macaque: I. Incidence and Stimulus-dependence of Gamma-band Neuronal Oscillations. *Cerebral Cortex*, *10*(11), 1105–1116. https://doi.org/10.1093/cercor/10.11.1105

Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron*, *88*(1), 220–235. https://doi.org/10.1016/j.neuron.2015.09.034

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science (New York, N.Y.)*, *291*(5508), 1560–1563. https://doi.org/10.1126/science.1055465

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, *16*(2), 123–144. https://doi.org/10.1016/S0926-6410(02)00244-6

Geisler, W. S. (2008). Visual Perception and the Statistical Properties of Natural Scenes. *Annual Review of Psychology*, *59*(1), 167–192. https://doi.org/10.1146/annurev.psych.58.110405.085632

Goettker, A., Pidaparthy, H., Braun, D. I., Elder, J. H., & Gegenfurtner, K. R. (2021). Ice hockey spectators use contextual cues to guide predictive eye movements. *Current Biology*, *31*(16), R991–R992. https://doi.org/10.1016/j.cub.2021.06.087

Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences*, *86*(5), 1698–1702. https://doi.org/10.1073/pnas.86.5.1698

Harris, K. D., & Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature*, *503*(7474), Article 7474. https://doi.org/10.1038/nature12654

Henderson, J. M. (2017). Gaze Control as Prediction. *Trends in Cognitive Sciences*, *21*(1), 15–23. https://doi.org/10.1016/j.tics.2016.11.003

Hetenyi, D., Haarsma, J., & Kok, P. (2024). *Pre-stimulus alpha oscillations encode stimulus-specific visual predictions* [Preprint]. Neuroscience. https://doi.org/10.1101/2024.03.13.584593

Hogendoorn, H. (2022). Perception in real-time: Predicting the present, reconstructing the past. *Trends in Cognitive Sciences*, *26*(2), 128–141. https://doi.org/10.1016/j.tics.2021.11.003

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, *4*, 186. https://doi.org/10.3389/fnhum.2010.00186

Kaiser, D. (2022). Spectral brain signatures of aesthetic natural perception in the α and β frequency bands. *Journal of Neurophysiology*, *128*(6), 1501–1505. https://doi.org/10.1152/jn.00385.2022

Kaiser, D., & Cichy, R. M. (2021). Parts and wholes in scene processing. *Journal of Cognitive Neuroscience*, *34*(1), 4–15. https://doi.org/10.1162/jocn_a_01788

Kaiser, D., Häberle, G., & Cichy, R. M. (2020a). Cortical sensitivity to natural scene structure. *Human Brain Mapping*, *41*(5), 1286–1295. https://doi.org/10.1002/hbm.24875

Kaiser, D., Häberle, G., & Cichy, R. M. (2020b). Real-world structure facilitates the rapid emergence of scene category information in visual brain signals. *Journal of Neurophysiology*, *124*(1), 145–151. https://doi.org/10.1152/jn.00164.2020

Kaiser, D., Häberle, G., & Cichy, R. M. (2021). Coherent natural scene structure facilitates the extraction of task-relevant object information in visual cortex. *NeuroImage*, *240*, 118365. https://doi.org/10.1016/j.neuroimage.2021.118365

Kaiser, D., & Nyga, K. (2020). Tracking cortical representations of facial attractiveness using time-resolved representational similarity analysis. *Scientific Reports*, *10*(1), 1–10. https://doi.org/10.1038/s41598-020-74009-9

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences*, *23*(8), 672–685. https://doi.org/10.1016/j.tics.2019.04.013

Kaiser, D., Turini, J., & Cichy, R. M. (2019). A neural mechanism for contextualizing fragmented inputs during naturalistic vision. *Elife*, *8*, e48182. https://doi.org/10.7554/eLife.48182

Keller, A. J., Roth, M. M., & Scanziani, M. (2020). Feedback generates a second receptive field in neurons of the visual cortex. *Nature*, *582*(7813), Article 7813. https://doi.org/10.1038/s41586-020-2319-4

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive Processing: A Canonical Cortical Computation. *Neuron*, *100*(2), 424–435. https://doi.org/10.1016/j.neuron.2018.10.003

Kelly, S. P., Lalor, E. C., Reilly, R. B., & Foxe, J. J. (2006). Increases in Alpha Oscillatory Power Reflect an Active Retinotopic Mechanism for Distracter Suppression During Sustained Visuospatial Attention. *Journal of Neurophysiology*, *95*(6), 3844–3851. https://doi.org/10.1152/jn.01234.2005

Kirchberger, L., Mukherjee, S., Self, M. W., & Roelfsema, P. R. (2023). Contextual drive of neuronal responses in mouse V1 in the absence of feedforward input. *Science Advances*, *9*(3), eadd2498. https://doi.org/10.1126/sciadv.add2498

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Current Biology*, *26*(3), 371–376. https://doi.org/10.1016/j.cub.2015.12.038

Kok, P., & de Lange, F. P. (2014). Shape Perception Simultaneously Up- and Downregulates Neural Activity in the Primary Visual Cortex. *Current Biology*, *24*(13), 1531–1535. https://doi.org/10.1016/j.cub.2014.05.042

Kourtzi, Z., Tolias, A. S., Altmann, C. F., Augath, M., & Logothetis, N. K. (2003). Integration of Local Features into Global Shapes: Monkey and Human fMRI Studies. *Neuron*, *37*(2), 333–346. https://doi.org/10.1016/S0896-6273(02)01174-1

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Kumar, M., Federmeier, K. D., & Beck, D. M. (2021). The N300: An index for predictive coding of complex visual objects and scenes. *Cerebral Cortex Communications*, *2*(2), tgab030. https://doi.org/10.1093/texcom/tgab030

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(Volume 62, 2011), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lange, F. P. de, Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences*, *98*(4), 1907–1911. https://doi.org/10.1073/pnas.98.4.1907

Mannion, D. J., Kersten, D. J., & Olman, C. A. (2013). Consequences of polar form coherence for fMRI responses in human visual cortex. *NeuroImage*, *78*, 152–158. https://doi.org/10.1016/j.neuroimage.2013.04.036

Mannion, D. J., Kersten, D. J., & Olman, C. A. (2014). Regions of mid-level human visual cortex sensitive to the global coherence of local image patches. *Journal of Cognitive Neuroscience*, *26*(8), 1764–1774. https://doi.org/10.1162/jocn_a_00588

Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.-M., Kennedy, H., & Fries, P. (2016). Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. *Neuron*, *89*(2), 384–397. https://doi.org/10.1016/j.neuron.2015.12.018

Mima, T., Oluwatimilehin, T., Hiraoka, T., & Hallett, M. (2001). Transient Interhemispheric Neuronal Synchrony Correlates with Object Recognition. *Journal of Neuroscience*, *21*(11), 3942–3948. https://doi.org/10.1523/JNEUROSCI.21-11-03942.2001

Morgan, A. T., Petro, L. S., & Muckli, L. (2019). Scene Representations Conveyed by Cortical Feedback to Early Visual Cortex Can Be Described by Line Drawings. *Journal of Neuroscience*, *39*(47), 9410–9423. https://doi.org/10.1523/JNEUROSCI.0852-19.2019

Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., Goebel, R., & Yacoub, E. (2015). Contextual Feedback to Superficial Layers of V1. *Current Biology*, *25*(20), 2690–2695. https://doi.org/10.1016/j.cub.2015.08.057

Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*, *48*(2), 507–517. https://doi.org/10.1016/j.neuropsychologia.2009.10.011

Mudrik, L., Shalgi, S., Lamy, D., & Deouell, L. Y. (2014). Synchronous contextual irregularities affect early scene processing: Replication and extension. *Neuropsychologia*, *56*, 447–458. https://doi.org/10.1016/j.neuropsychologia.2014.02.020

Munneke, J., Brentari, V., & Peelen, M. (2013). The influence of scene context on object recognition is independent of attentional focus. *Frontiers in Psychology*, *4*, 552. https://doi.org/10.3389/fpsyg.2013.00552

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian Reconstruction of Natural Images from Human Brain Activity. *Neuron*, *63*(6), 902–915. https://doi.org/10.1016/j.neuron.2009.09.006

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M. P., Poldrack, R. A., Poline, J.-B., Proal, E., Thirion, B., Van Essen, D. C., White, T., & Yeo, B. T. T. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nature Neuroscience*, *20*(3), 299–303. https://doi.org/10.1038/nn.4500

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527. https://doi.org/10.1016/j.tics.2007.09.009

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*, 519–526. https://doi.org/10.3758/BF03197524

Papale, P., Wang, F., Morgan, A. T., Chen, X., Gilhuis, A., Petro, L. S., Muckli, L., Roelfsema, P. R., & Self, M. W. (2022). Feedback brings scene information to the representation of occluded image regions in area V1 of monkeys and humans. *bioRxiv*. https://doi.org/10.1101/2022.11.21.517305

Pfurtscheller, G., Stancák, A., & Neuper, Ch. (1996). Event-related synchronization (ERS) in the alpha band — an electrophysiological correlate of cortical idling: A review. *International Journal of Psychophysiology*, *24*(1), 39–46. https://doi.org/10.1016/S0167-8760(96)00066-9

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the

horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115–126. https://doi.org/10.1038/nrn.2016.167

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate–Inanimate Distinction. *Journal of Cognitive Neuroscience*, *28*(5), 680–692. https://doi.org/10.1162/jocn_a_00924

Ragni, F., Lingnau, A., & Turella, L. (2021). Decoding category and familiarity information during visual imagery. *NeuroImage*, *241*, 118428. https://doi.org/10.1016/j.neuroimage.2021.118428

Ramkumar, P., Jas, M., Pannasch, S., Hari, R., & Parkkonen, L. (2013). Feature-specific information processing precedes concerted activation in human visual cortex. *Journal of Neuroscience*, *33*(18), 7691–7699. https://doi.org/10.1523/JNEUROSCI.3905-12.2013

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), Article 1. https://doi.org/10.1038/4580

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), Article 11. https://doi.org/10.1038/14819

Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research*, *179*(1), 3–20. https://doi.org/10.1016/0006-8993(79)90485-2

Romei, V., Brodbeck, V., Michel, C., Amedi, A., Pascual-Leone, A., & Thut, G. (2008). Spontaneous fluctuations in posterior α-band EEG activity reflect variability in excitability of human visual areas. *Cerebral Cortex*, *18*(9), 2010–2018. https://doi.org/10.1093/cercor/bhm229

Romei, V., Gross, J., & Thut, G. (2010). On the Role of Prestimulus Alpha Rhythms over Occipito-Parietal Areas in Visual Input Regulation: Correlation or Causation? *Journal of Neuroscience*, *30*(25), 8692–8697. https://doi.org/10.1523/JNEUROSCI.0160-10.2010

Rosch, E. (1978). Principles of Categorization. In E. Rosch & Barbara Lloyd (Eds.), *Cognition and categorization* (Cognition and Categorization, pp. 27–48). Lawrence Erlbaum Associates. https://doi.org/10.1016/b978-1-4832-1446-7.50028-5

Sauvé, G., Harmand, M., Vanni, L., & Brodeur, M. B. (2017). The probability of object–scene co-occurrence influences object identification processes. *Experimental*

Brain Research, *235*(7), 2167–2179. https://doi.org/10.1007/s00221-017-4955-y

Schendan, H. E., & Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia*, *40*(7), 931–945. https://doi.org/10.1016/S0028-3932(01)00176-2

Schendan, H. E., & Kutas, M. (2003). Time Course of Processes and Representations Supporting Visual Object Identification and Memory. *Journal of Cognitive Neuroscience*, *15*(1), 111–135. https://doi.org/10.1162/089892903321107864

Schendan, H. E., & Kutas, M. (2007). Neurophysiological Evidence for the Time Course of Activation of Global Shape, Part, and Local Contour Representations during Visual Object Categorization and Memory. *Journal of Cognitive Neuroscience*, *19*(5), 734–749. https://doi.org/10.1162/jocn.2007.19.5.734

Schendan, H. E., & Maher, S. M. (2009). Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *Neuroimage*, *44*(4), 1423–1438. https://doi.org/10.1016/j.neuroimage.2008.09.061

Self, M. W., van Kerkoerle, T., Goebel, R., & Roelfsema, P. R. (2019). Benchmarking laminar fMRI: Neuronal spiking and synaptic activity during top-down and bottom-up processing in the different layers of cortex. *NeuroImage*, *197*, 806–817. https://doi.org/10.1016/j.neuroimage.2017.06.045

Smith, F. W., & Muckli, L. (2010). Nonstimulated early visual areas carry information about surrounding context. *Proceedings of the National Academy of Sciences*, *107*(46), 20099–20103. https://doi.org/10.1073/pnas.1000233107

Stecher, R., & Kaiser, D. (2023). *Imaginary scenes are represented in cortical alpha activity* (p. 2023.10.23.563249). bioRxiv. https://doi.org/10.1101/2023.10.23.563249

Stephan, K. E., Marshall, J. C., Penny, W. D., Friston, K. J., & Fink, G. R. (2007). Interhemispheric Integration of Visual Processing during Task-Driven Lateralization. *Journal of Neuroscience*, *27*(13), 3512–3522. https://doi.org/10.1523/JNEUROSCI.4766-06.2007

Swettenham, J. B., Muthukumaraswamy, S. D., & Singh, K. D. (2009). Spectral Properties of Induced and Evoked Gamma Oscillations in Human Early Visual Cortex to Moving and Stationary Stimuli. *Journal of Neurophysiology*, *102*(2), 1241–1253. https://doi.org/10.1152/jn.91044.2008

Tallon-Baudry, C., Bertrand, O., Tallon-Baudry, C., Bertrand, O., Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, *3*(4), 151–162. https://doi.org/10.1016/S1364-6613(99)01299-1

Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., Rosen, B. R., & Belliveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience*, *15*(4), 3215–3230. https://doi.org/10.1523/JNEUROSCI.15-04-03215.1995

Truman, A., & Mudrik, L. (2018). Are incongruent objects harder to identify? The functional significance of the N300 component. *Neuropsychologia*, *117*, 222–232. https://doi.org/10.1016/j.neuropsychologia.2018.06.004

van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., & Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proceedings of the National Academy of Sciences*, *111*(40), 14332–14341. https://doi.org/10.1073/pnas.1402773111

Vetter, P., Bola, Ł., Reich, L., Bennett, M., Muckli, L., & Amedi, A. (2020). Decoding Natural Sounds in Early "Visual" Cortex of Congenitally Blind Individuals. *Current Biology*, *30*(15), 3039-3044.e2. https://doi.org/10.1016/j.cub.2020.05.071

Vetter, P., Smith, F. W., & Muckli, L. (2014). Decoding Sound and Imagery Content in Early Visual Cortex. *Current Biology*, *24*(11), 1256–1262. https://doi.org/10.1016/j.cub.2014.04.020

Võ, M. L., Boettcher, S. E., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*, *29*, 205–210. https://doi.org/10.1016/j.copsyc.2019.03.009

Võ, M. L., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, *9*(3), 24. https://doi.org/10.1167/9.3.24

Võ, M. L., & Henderson, J. M. (2011). Object–scene inconsistencies do not capture gaze: Evidence from the flash-preview moving-window paradigm. *Attention, Perception, & Psychophysics*, *73*(6), 1742–1753. https://doi.org/10.3758/s13414-011-0150-6

Võ, M. L., & Wolfe, J. M. (2013). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychological Science*, *24*(9), 1816–1823. https://doi.org/10.1177/0956797613476955

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. https://doi.org/10.1111/nyas.14321

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain. *Journal of Neuroscience*, *29*(34), 10573–10581. https://doi.org/10.1523/JNEUROSCI.0559-09.2009

Watson, J. D., Myers, R., Frackowiak, R. S., Hajnal, J. V., Woods, R. P., Mazziotta, J. C., Shipp, S., & Zeki, S. (1993). Area V5 of the human brain: Evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cerebral Cortex*, *3*(2), 79–94. https://doi.org/10.1093/cercor/3.2.79

Williams, M. A., Baker, C. I., Op de Beeck, H. P., Mok Shim, W., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience*, *11*(12), Article 12. https://doi.org/10.1038/nn.2218

Winlove, C. I. P., Milton, F., Ranson, J., Fulford, J., MacKisack, M., Macpherson, F., & Zeman, A. (2018). The neural correlates of visual imagery: A co-ordinate-based meta-analysis. *Cortex*, *105*, 4–25. https://doi.org/10.1016/j.cortex.2017.12.014

Wischnewski, M., & Peelen, M. V. (2021). Causal neural mechanisms of context-based object recognition. *eLife*, *10*, e69736. https://doi.org/10.7554/eLife.69736

Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences*, *109*(9), 3593–3598. https://doi.org/10.1073/pnas.1120118109

Xie, S., Kaiser, D., & Cichy, R. M. (2020). Visual imagery and perception share neural representations in the alpha frequency band. *Current Biology*, *30*(13), 2621-2627.e5. https://doi.org/10.1016/j.cub.2020.04.074

5 Appendix

5.1 Original publication of Study 1

Chen, L., Cichy, R. M.*, & Kaiser, D.* (2022). Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations. Cerebral Cortex, 32(16), 3553–3567. https://doi.org/10.1093/cercor/bhab433

* The authors contributed equally.

Copyright

OXFORD

Cerebral Cortex, 2022, **32**, 3553–3567
https://doi.org/10.1093/cercor/bhab433
Advance access publication date: 10 December 2021
Original Article

Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations

Lixiang Chen [1], Radoslaw Martin Cichy[1,†], Daniel Kaiser [2,3,†]

[1]Department of Education and Psychology, Freie Universität Berlin, Berlin 14195, Germany,
[2]Mathematical Institute, Department of Mathematics and Computer Science, Physics, Geography, Justus-Liebig-Universität Gießen, Gießen 35392, Germany,
[3]Center for Mind, Brain and Behavior (CMBB), Philipps-Universität Marburg and Justus-Liebig-Universität Gießen, Marburg 35032, Germany
*Address correspondence to Lixiang Chen, Department of Education and Psychology, Freie Universität Berlin, Berlin 14195, Germany. Email: lixiang.chen@fu-berlin.de
†R.M.C. and D.K. have contributed equally to this work

During natural vision, objects rarely appear in isolation, but often within a semantically related scene context. Previous studies reported that semantic consistency between objects and scenes facilitates object perception and that scene-object consistency is reflected in changes in the N300 and N400 components in EEG recordings. Here, we investigate whether these N300/400 differences are indicative of changes in the cortical representation of objects. In two experiments, we recorded EEG signals, while participants viewed semantically consistent or inconsistent objects within a scene; in Experiment 1, these objects were task-irrelevant, while in Experiment 2, they were directly relevant for behavior. In both experiments, we found reliable and comparable N300/400 differences between consistent and inconsistent scene-object combinations. To probe the quality of object representations, we performed multivariate classification analyses, in which we decoded the category of the objects contained in the scene. In Experiment 1, in which the objects were not task-relevant, object category could be decoded from ~100 ms after the object presentation, but no difference in decoding performance was found between consistent and inconsistent objects. In contrast, when the objects were task-relevant in Experiment 2, we found enhanced decoding of semantically consistent, compared with semantically inconsistent, objects. These results show that differences in N300/400 components related to scene-object consistency do not index changes in cortical object representations but rather reflect a generic marker of semantic violations. Furthermore, our findings suggest that facilitatory effects between objects and scenes are task-dependent rather than automatic.

Key words: multivariate pattern analysis; N300; object representation; scene-object consistency.

Introduction

In the real world, objects rarely appear in isolation, but practically always within a particular scene context (Bar 2004; Wolfe et al. 2011; Kaiser et al. 2019; Võ et al. 2019). Objects are often semantically related to the scene they appear in: for instance, microwaves usually appear in the kitchen, but practically never in the bathroom. Several behavioral studies have shown that such semantic relations between objects and scenes affect object identification. Early studies using line drawings of scenes and objects found that objects were detected faster and more accurately when they were in a consistent setting than in an inconsistent setting (Palmer 1975; Biederman et al. 1982; Boyce et al. 1989; Boyce and Pollatsek 1992). Similar results were recently reported for scene photographs (Davenport and Potter 2004; Davenport 2007; Munneke et al. 2013). In line with such findings, eye-tracking studies have shown that inconsistent objects are fixated longer and more often than consistent objects (Võ and Henderson 2009, 2011; Cornelissen and Võ 2017), suggesting that objects are perceived more swiftly within a consistent than within an inconsistent scene. Interestingly, such behavioral facilitation effects are also observed when, instead of the object, the scene is task-relevant: Davenport and Potter (2004) and Davenport (2007) reported that scenes were identified more accurately if they contained a consistent foreground object compared with an inconsistent one. These effects suggest that objects and scenes are processed in a highly interactive manner.

To characterize the neural basis of these semantic consistency effects, EEG studies have used paradigms in which objects appear within consistent or inconsistent scenes, either simultaneously or sequentially (Ganis and Kutas 2003; Mudrik et al. 2010; Võ and Wolfe 2013; Draschkow et al. 2018; Coco et al. 2020). For example, Võ and Wolfe (2013) adopted a sequential design, in which participants first viewed a scene image, followed by a location cue, where then appeared a consistent (e.g., a computer mouse on an office table) or an inconsistent

Received: August 25, 2021. **Revised:** November 2, 2021. **Accepted:** November 3, 2021

object (e.g., a soap on an office table). They found objects in an inconsistent scene evoked more negative responses than consistent objects in the N300 (~250–350 ms) and N400 (~350–600 ms) windows. Several other studies (Mudrik et al. 2010, 2014; Truman and Mudrik 2018) using a simultaneous design, in which the scene and object were presented simultaneously, reported similar N300 and/or N400 modulations. Critically, the earlier N300 effects are often considered to reflect differences in perceptual processing between typically and atypically positioned objects (Schendan and Maher 2009; Mudrik et al. 2010; Kumar et al. 2021). On this view, consistency-related differences in EEG waveforms arise as a consequence of differences in the visual analysis of objects and scenes, rather than due to a postperceptual signaling of (in)consistency.

If differences in the N300 waveform indeed index changes in perceptual processing, the N300 ERP effect should be accompanied by differences in the neural representation of the objects. In this study, we put this prediction to the test. Across two experiments, we compared differences in the N300/400 EEG components to multivariate decoding of objects contained in consistent and inconsistent scenes. In both experiments, participants completed a sequential semantic consistency paradigm, in which scenes from eight different categories were consistently or inconsistently combined with objects from 16 categories. We then examined the influence of scene-object consistency on EEG signals, both when the objects were task-irrelevant (Experiment 1) and when participants performed a recognition task on the objects (Experiment 2). In both experiments, we replicated previously reported ERP effects, with greater N300 and N400 components for inconsistent scene-object combinations, compared with consistent combinations. To probe the quality of object and scene representations, we performed multivariate classification analyses, in which we decoded between the object and scene categories separately for each condition. In Experiment 1, in which the objects were not task-relevant, object category could be decoded from ~100 ms after the object presentation, but no difference in decoding performance was found between consistent and inconsistent objects. In Experiment 2, in which the objects were directly task-relevant, we found enhanced decoding of semantically consistent, compared with semantically inconsistent, objects. In both experiments, we found no differences in scene category decoding between semantically consistent and inconsistent conditions. Altogether, these results show that differences in N300/400 components related to scene-object consistency do not necessarily index changes in cortical object representations, but rather reflect a generic marker of semantic violations. Furthermore, they suggest that facilitation effects between objects and scenes are task-dependent rather than automatic.

Materials and Methods

All materials and methods were identical for the two experiments, unless stated otherwise.

Participants. Thirty-two participants (16 males, mean age 26.23 years, $SD = 2.05$ years), with normal or corrected-to-normal vision, took part in Experiment 1. Another 32 participants (14 males, mean age 26.97 years, $SD = 1.67$ years) took part in Experiment 2. Participants were paid volunteers or participated for partial course credits. All participants provided written, informed consent prior to participating in the experiment. The experiments were approved by the ethical committee of the Department of Education and Psychology at Freie Universität Berlin and were conducted in accordance with the Declaration of Helsinki.

Stimuli. The stimulus set comprised scene images from eight categories: beach, bathroom, office, kitchen, gym, street, supermarket, and prairie. The scenes were grouped into four pairs (beach and bathroom, office and kitchen, gym and street, and supermarket and prairie). We chose four objects for each scene pair, two of which were semantically consistent with one scene and two of which were semantically consistent with the other scene. To create semantically inconsistent scene-object combinations, we simply exchanged the objects between the scenes. All combinations of scenes and objects can be found in Table 1. For example, consider the office and kitchen pair: we first chose a computer and a printer as consistent objects for the office and chose a rice cooker and a microwave as consistent objects for the kitchen. We in turn chose the rice cooker and microwave as inconsistent objects for the office and the computer and printer as inconsistent objects for the kitchen. We pasted the objects into the scene images using Adobe Photoshop. The object locations were the same across the consistent and inconsistent objects, and they were always in line with the typical position of the consistent object (e.g., a computer was positioned on an office desk in the same way as a rice cooker). We used three exemplars for each scene category and three exemplars for each object, yielding 288 unique stimuli. During the experiments, the scenes could also be shown without objects (see below). Figure 1A shows some examples of the stimuli.

Paradigm. Participants were seated 53 cm from an LCD monitor with a size of 34 × 27 cm and a refresh rate of 60 Hz. The images were presented on the screen at a visual angle: horizontal 20°, vertical 15.6°. We adopted a sequential scene-object congruity paradigm similar to Võ and Wolfe (2013). Image presentation and recording of subjects' behavioral responses were controlled using MATLAB and the Psychophysics Toolbox (Brainard 1997). Each trial began with a fixation cross "+" shown for a random interval between 1.3 and 1.5 s, after which a scene image (without the critical object) was presented for 500 ms. Next, a red dot cue was presented at a single location in the scene for 500 ms, indicating where the critical object would appear. Participants were instructed

Table 1. Combinations of scenes and objects for the consistent and inconsistent conditions. Note that scenes were grouped into pairs; for each pair, four objects were used as consistent and inconsistent objects

Scenes	Consistent objects	Inconsistent objects
Bathroom	Toilet, washing machine	Parasol, deck chair
Beach	Parasol, deck chair	Toilet, washing machine
Office	Computer, printer	Microwave, rice cooker
Kitchen	Microwave, rice cooker	Computer, printer
Gym	Treadmill, spinning bike	Scooter, bus stop sign
Road	Scooter, bus stop sign	Treadmill, spinning bike
Supermarket	Shopping cart, shop assistant	Ostrich, zebra
Prairie	Ostrich, zebra	Shopping cart, shop assistant

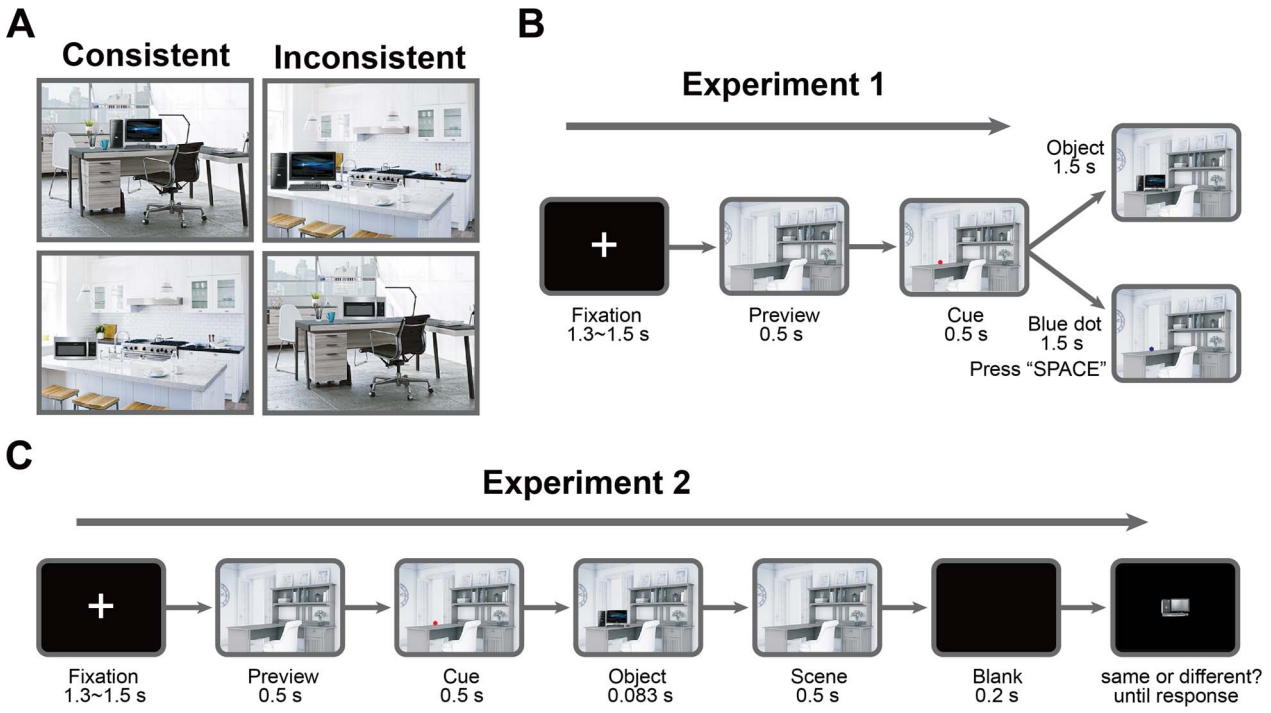

Fig. 1. Experimental design. (A) Examples of consistent and inconsistent scene-object combinations. (B) Trial sequence in Experiment 1. After a fixation interval, a scene without the critical object was presented. Next, a red dot cue was presented, and participants were asked to move their eyes to this location. After that, the critical object appeared at the cued location in the scene. On target trials, the red cue turned blue, and participants were instructed to press spacebar. (C) Trial sequence in Experiment 2. Here, the objects were displayed briefly at the location of the cue. In a subsequent recognition test, an object of the same category appeared on the screen. Participants were asked to determine whether this object exemplar was the one that appeared earlier in the scene.

to move their eyes to the dot as quickly as possible. To avoid eye movement artifacts during the subsequent object presentation, we told the participants not to move their eyes away from the dot location until the beginning of the next trial. After that, a semantically consistent or inconsistent object appeared at the location of the red dot.

In Experiment 1, the objects were not task-relevant. The object simply remained visible together with the scene for 1500 ms before the next trial started. Participants performed an unrelated attention control task: to ensure that they attended the cued location, we added task trials (10% of trials) during which no object appeared after the cue. Instead, the color of the dot changed from red to blue. Participants were instructed to press the spacebar when they detected the change (detection rate in target trials: 97.8%, SE = 0.63%). An example trial for Experiment 1 is shown in Figure 1B.

In Experiment 2, the objects were directly task-relevant. The consistent or inconsistent object appeared only very briefly (83 ms) at the location of the red dot. The scene image remained on the screen for another 500 ms after the object disappeared. After a 200-ms blank internal, participants were asked to perform an object recognition test. During the test, an object was shown on the screen, which was either the same object exemplar they had just seen or a different exemplar of the same category. Test objects were presented in grayscale to increase task difficulty. Participants were asked to determine whether this object exemplar was

the one that appeared earlier in the scene. If it was, the participants should press G button, otherwise press H button. The next trial began as soon as participants made a choice. An example trial for Experiment 2 is shown in Figure 1C.

Both experiments included three runs and all 288 unique stimulus images were presented once in random order within each run. Across runs, there were 27 repetitions for each specific scene-object combination.

EEG Recording and Preprocessing

EEG signals were recorded using an EASYCAP 64-electrode system and a Brainvision actiCHamp amplifier in both Experiments. Electrodes were arranged according to the 10–10 system. EEG data were recorded with a sample rate of 1000 Hz and filtered online between 0.03 and 100 Hz. All electrodes were referenced online to the Fz electrode and rereferenced offline to the average of data from all channels. Offline data preprocessing was performed using FieldTrip (Oostenveld et al. 2011). EEG data were segmented into epochs from −100 to 800 ms relative to the onset of the critical object and baseline corrected by subtracting the mean signal prior to the object onset. To track the temporal representations of scenes, EEG data were segmented into epochs from −1100 to 800 ms relative to the onset of the object and baseline corrected by subtracting the mean signal prior to the scene presentation (−100 to 0 ms relative to the scene onset). Channels and trials containing excessive noise were removed by visual inspection. On average, we removed 1.50 ± 0.51 channels in Experiment 1 and 1.53 ± 0.57 channels in Experiment 2. These channels were not interpolated in further ERP and decoding analyses. Blinks and eye movement artifacts were removed using independent component analysis and visual inspection of the resulting components. The epoched data were downsampled to 200 Hz. As filtering is often recommended for univariate analyses but discouraged for multivariate analyses (Grootswagers et al. 2017; van Driel et al. 2021), the EEG data were not filtered for the decoding analyses. For the ERP analyses, the preprocessed data were additionally band-pass filtered at 0.1–30 Hz. This additional filtering was performed after the other preprocessing steps to equate the preprocessing pipeline between the ERP and decoding analyses. Performing the filtering before epoching yielded highly similar ERP results (see Supplementary Figs 6 and 7).

ERP Analyses

To replicate semantic consistency ERP effect reported in previous scene studies (e.g., Võ and Wolfe 2013; Mudrik et al. 2014), we performed ERP analyses using FieldTrip. In accordance with Võ and Wolfe (2013), we chose nine electrodes (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, and CP2) located in the mid-central region for further ERP analysis. This a-priori electrode selection was corroborated in a topographical analysis of the scene-consistency effect (see Supplementary Fig. 5). We first averaged the evoked responses across these electrodes and then averaged these mean responses separately for the consistent and inconsistent conditions and each participant.

Decoding Analyses

We performed two complementary multivariate decoding analyses to track temporal representations of objects and scenes across time. First, to track representations of objects and investigate how consistent or inconsistent scene contexts affect objects processing, we performed decoding analyses between two consistent and inconsistent objects separately within each scene at each time point from −100 to 800 ms relative to the onset of the object. For example, we performed classification analyses to either differentiate printers (consistent) from computers (consistent) in office scenes or to differentiate printers (inconsistent) from computers (inconsistent) in kitchen scenes, at each time point. Second, to track the impact of consistent or inconsistent objects on scenes representations, we performed decoding analyses to discriminate between every two scene categories separately for consistent and inconsistent conditions at each time point from −100 to 1800 ms relative to the onset of the scene (−1100 to 800 ms relative to the onset of the object). For example, we performed classification analyses to differentiate office scenes containing a printer or computer (consistent) from kitchen scenes containing a microwave or rice cooker (consistent), or to differentiate office scenes containing a microwave or rice cooker (inconsistent) from kitchen scenes containing a printer or computer (inconsistent). In both analyses, we used all available trials, including those in participants responded incorrectly. For each decoding analysis, we adopted two approaches: standard timeseries decoding (Boring et al. 2020; Kaiser and Nyga 2020), using data from a sliding time window, and cumulative decoding (Ramkumar et al. 2013; Kaiser et al. 2020a), using aggregated data from all elapsed time points. The two approaches are detailed in the following paragraphs.

Timeseries decoding. Timeseries decoding analyses were performed using Matlab and CoSMoMVPA (Oosterhof et al. 2016). To increase the power of our timeseries decoding, the analysis was performed on a sliding time window (50-ms width), with a 5-ms resolution. This approach thus not only utilizes data from current time point, but the data from five time points before and after the current time point. For each sliding window position, we concatenated the response patterns across all time points within the time window and then unfolded the whole pattern into a vector. For a comparison with alternative timeseries decoding approach, which use individual time point or average data across the sliding windows, see Supplementary Figures 2 and 3.

Considering excessive data dimensionality may harm classification, we adopted principal component analysis (PCA) to reduce the dimensionality of the data (Grootswagers et al. 2017; Kaiser et al. 2020a;

Kaiser and Nyga 2020). For each classification, a PCA was performed on all data from the training set, and the PCA solution was projected onto data from the testing set (Experiment 1, mean 34.8 PCs for object decoding, mean 52.1 PCs for scene decoding; Experiment 2, mean 41.2 PCs for object decoding, mean 69.9 PCs for scene decoding). For each PCA, we retained the set of components explaining 99% of the variance in the training set data.

The classification was performed separately for each time point from −100 to 800 ms (from −1100 to 800 ms for scene decoding), using LDA classifiers with 10-fold cross-validation. Specifically, the EEG data from all epochs were first allocated to 10 folds randomly. LDA classifiers were then trained on data from 9 folds and then tested on data from the left-out fold. The amount of data in the training set was always balanced across conditions. For each object decoding analysis, the training set included up to 48 trials, and the testing set included up to six trials; for each scene decoding analysis, the training set included up to 96 trials and the testing set included up to 12 trials. The classification was done repeatedly until every fold was left out once. For each time point, the accuracies were averaged across the 10 repetitions.

Cumulative decoding. We also performed cumulative decoding analyses, which takes into account the data of all time points before the current time point in the epoch for classifications (Ramkumar et al. 2013; Kaiser et al. 2020a). For example, for the first time point in the epoch, the classifier was trained and tested on response patterns at this time point in the epoch; at the second time point in the epoch, the classifier was trained and tested on response patterns at the first and second time points in the epoch; and at the last time point in the epoch, the classifier was trained and tested on response patterns at all time points in the epoch. For each time point, we concatenated the response patterns across all time points up to the current one and then unfolded the whole pattern into a vector that was subsequently used for decoding.

The cumulative decoding approach uses increasingly large amounts of data that span multiple time points. This allows classifiers to capitalize on complex spatiotemporal response patterns that emerge across the trial, which may provide additional sensitivity for detecting effects that are not only visible across electrode space but that are also transported by variations in the time domain. On the flip side, this renders the interpretation of the results less straightforward: one can only conclude that spatiotemporal response patterns up to the current point allow for classification, but not which features enable this classification.

As for the timeseries decoding, LDA classifiers with 10-fold cross-validation were used for classifications and PCA was adopted to reduce the dimensionality of the data for each classification step across time (Experiment 1, mean 39.2 PCs for object decoding, mean 73.5 PCs for scene decoding; Experiment 2, mean 43.9 PCs for object decoding, mean 86.9 PCs for scene decoding).

Statistics

For the behavioral responses in Experiment 2, we used paired *t*-tests to compare participants' accuracy and response times when they were asked to recognize consistent and inconsistent objects.

For ERP analyses, we used paired *t*-tests to compare the averaged EEG responses evoked by consistent and inconsistent scene-object combinations, at each time point.

For decoding analyses, we used one-sample *t*-tests to compare decoding accuracies against chance level (similar results were obtained using a permutation-based testing approach, see Supplementary Fig. 1) and paired *t*-tests to compare decoding accuracies between the consistent and inconsistent conditions, at each time point.

Differences in ERP and decoding effects between experiments were assessed using independent *t*-tests. Direct differences between the ERP and decoding effects obtained in the two experiments were assessed in a mixed-effects ANOVA with the factors consistency effect (ERP vs. decoding) and experiment (Experiment 1 vs. Experiment 2).

Multiple-comparison corrections were performed using FDR ($P < 0.05$), and only clusters of at least five consecutive significant time points (i.e., 25 ms) were considered. We also calculated Bayes factors (Rouder et al. 2009) for all analyses.

Results

Experiment 1

ERP Signals Indexing Scene-Object Consistency

To track the influence of scene-object consistency on EEG responses, we first analyzed EEG waveforms in mid-central electrodes. In this analysis, we found more negative responses evoked by inconsistent scene-object combinations than consistent combinations, which emerged at 170–330 ms (peak: $t(31) = 4.884$, $BF_{10} = 765.26$) and 355–470 ms (peak: $t(31) = 3.429$, $BF_{10} = 20.06$) (Fig. 2). These results demonstrate larger N300 and N400 components evoked by inconsistent scenes, which is in line with previous findings (Mudrik et al. 2010, 2014; Võ and Wolfe 2013).

Tracking Object Representations in Consistent and Inconsistent Scenes

Having established reliable ERP differences between consistent and inconsistent scene-object combinations, we were next interested if these differences were accompanied by differences in how well the consistent and inconsistent objects were represented. We performed timeseries and cumulative decoding analyses between two consistent or inconsistent objects separately within each scene at each time point from −100 to 800 ms relative to the onset of the object. In both analyses, we found highly similar decoding performances for

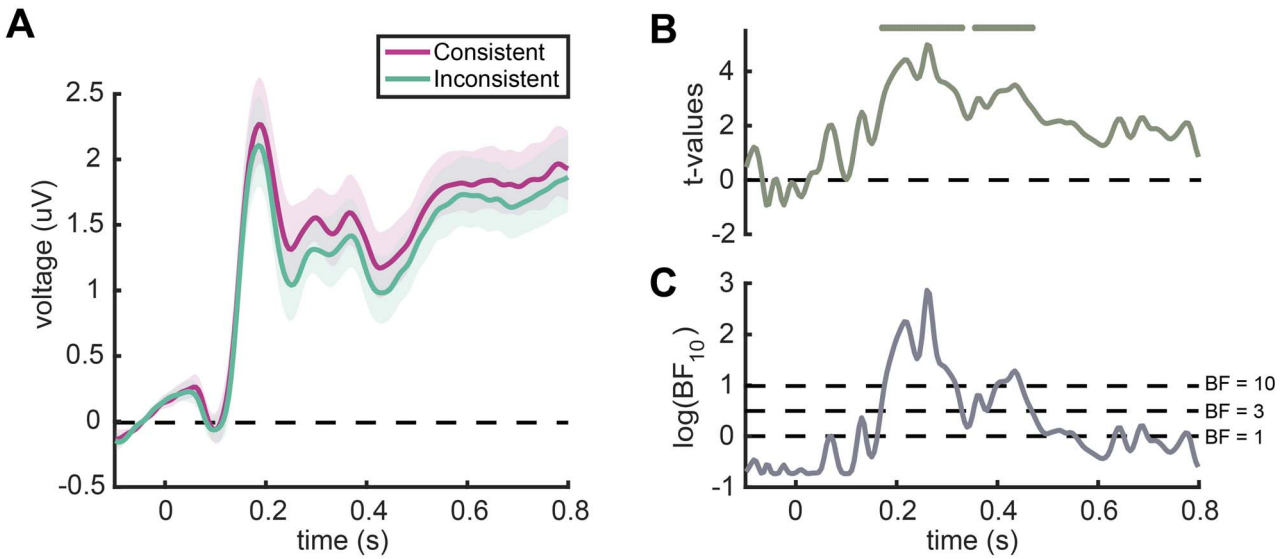

Fig. 2. Event-related potentials (ERPs) in Experiment 1. (*A*) ERPs recorded from the mid-central region for consistent and inconsistent scene-object combinations. Error margins represent standard errors. (*B*) *t*-values for the comparisons between consistent and inconsistent conditions. Line markers denote significant differences between conditions ($P < 0.05$, FDR-corrected). (*C*) Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions. For display purposes, the BF_{10} values were log-transformed. Dotted lines show low ($BF_{10} = 1$), moderate ($BF_{10} = 3$), and high ($BF_{10} = 10$) evidence for a difference between conditions. In line with previous reports, these results show that scene-object consistency is represented in evoked responses 170–330 ms and 355–470 ms after the object onset.

both consistent and inconsistent objects. Specifically, there was significant decoding between consistent objects, which emerged at 65–790 ms in the timeseries decoding (Fig. 3A) and between 80 and 800 ms in the cumulative decoding (Fig. 3D), and there was significant decoding between inconsistent objects in both the timeseries decoding (60–645 ms; Fig. 3A) and cumulative decoding (90–800 ms; Fig. 3D). No significant differences in decoding accuracy were found between consistent and inconsistent objects. Hence, despite the reliable ERP differences between consistent and inconsistent scene-object stimuli, there was no evidence for an automatic facilitation from the scene to the semantically consistent object.

Tracking the Representation of Scenes with Consistent and Inconsistent Objects

Although scene-object consistency does not automatically facilitate the representation of the objects, there may still be an opposite cross-facilitation effect where the consistent object enhances scene representations. To test this possibility, we performed decoding analyses to discriminate between every two categories separately for the consistent and inconsistent conditions from −100 to 1800 ms relative to the onset of the scene. We found significant decoding between scenes with a consistent object in both the timeseries decoding (45–1800 ms) and cumulative decoding (80–1800 ms) analyses. Significant decoding between scenes that contained inconsistent objects was also found in both the timeseries decoding (40–1800 ms) and cumulative decoding (80–1800 ms). These results are consistent with previous findings (Lowe et al. 2018; Kaiser et al. 2020b), which suggest that scene category can be decoded within 100 ms (Fig. 4). However, no significant differences were found between these scenes with consistent and inconsistent objects. Such differences were also not observed when we corrected for multiple comparisons solely between 0 and 800 ms relative to the onset of the object. These results suggest that scene category can be decoded in a temporally sustained way, but semantically consistent objects have no facilitatory effect on scene representations.

Experiment 2

In Experiment 1, we did not find differences in object and scene representations between the consistent and inconsistent object-scene combinations, despite robust ERP differences between the two conditions. However, the objects and scenes were both not task-relevant in Experiment 1—although participants spatially attended the object location, the objects' features were not important for solving the task. Under such conditions, object representations may not benefit from semantically consistent context to the same extent as when object features are critical for solving the task. In Experiment 2, we therefore made the objects task-relevant.

Behavioral Object Recognition in Semantically Consistent and Inconsistent Scenes

In Experiment 2, participants performed a recognition task, in which they were asked to report whether a test object was identical to the one they had previously seen in the scene (Fig. 1C). In line with previous findings (Davenport and Potter 2004; Davenport 2007; Munneke et al. 2013), we found that objects were recognized more accurately when they were embedded in consistent

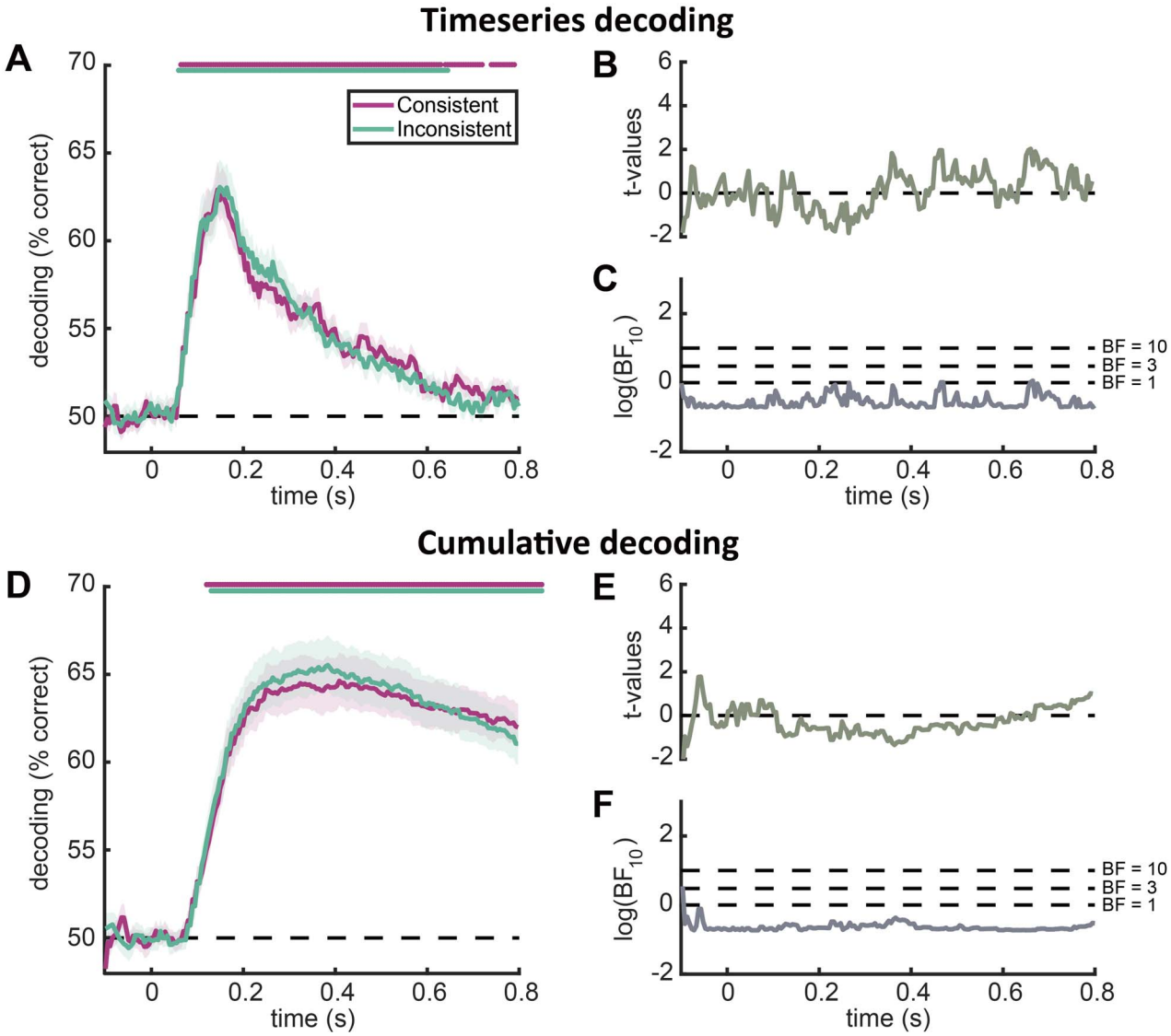

Fig. 3. Decoding for the consistent or inconsistent objects within each scene in Experiment 1. (A) Timeseries decoding results, separately for consistent and inconsistent objects. Line markers denote significant above-chance decoding ($P < 0.05$, FDR-corrected). (*B*) *t*-values for the comparisons between consistent and inconsistent conditions. (*C*) Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions. For display purposes, the BF_{10} values were log-transformed. Dotted lines show low ($BF_{10} = 1$), moderate ($BF_{10} = 3$), and high ($BF_{10} = 10$) evidence for a difference between conditions. (*D*) Cumulative decoding results, separately for consistent and inconsistent objects. Line markers denote significant above-chance decoding ($P < 0.05$, FDR-corrected). (*E*, *F*) *t*-values and Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions, as in (*B*, *C*). These results show robust decoding for consistently and inconsistently placed objects. However, despite the reliable differences between consistent and inconsistent scene-object combinations in ERP signals, object decoding was highly similar between the consistent and inconsistent conditions.

scenes than in inconsistent scenes (mean accuracy: consistent = 82.36%, inconsistent = 79.30%; $t(31) = 2.598$, $P = 0.011$). These results suggest that semantically consistent scenes can enhance the recognition of objects. There was no difference in response times between two conditions (mean response time: consistent = 723.8 ms, inconsistent = 742.4 ms; $t(31) = -0.648$, $P = 0.519$).

ERP Signals Indexing Scene-Object Consistency

Inconsistent scene-object combinations evoked more negative responses in mid-central electrodes than consistent combinations at 240–335 ms (peak: $t(31) = 4.385$, $BF_{10} = 210.85$), 360–500 ms (peak: $t(31) = 4.291$, $BF_{10} = 165.94$), and 570–590 ms (peak: $t(31) = 2.986$, $BF_{10} = 7.36$) (Fig. 5). The results suggest larger N300 and N400 components evoked by semantically inconsistent scene-object combinations, replicating the findings from Experiment 1.

Tracking Object Representations in Consistent and Inconsistent Scenes

To test whether semantically consistent scenes facilitate object representations differently from semantically inconsistent scenes when the objects are task-relevant, we performed both timeseries and cumulative decoding analyses, where we classified two consistent or inconsistent objects within each scene at each time point from −100 to 800 ms relative to the onset of the

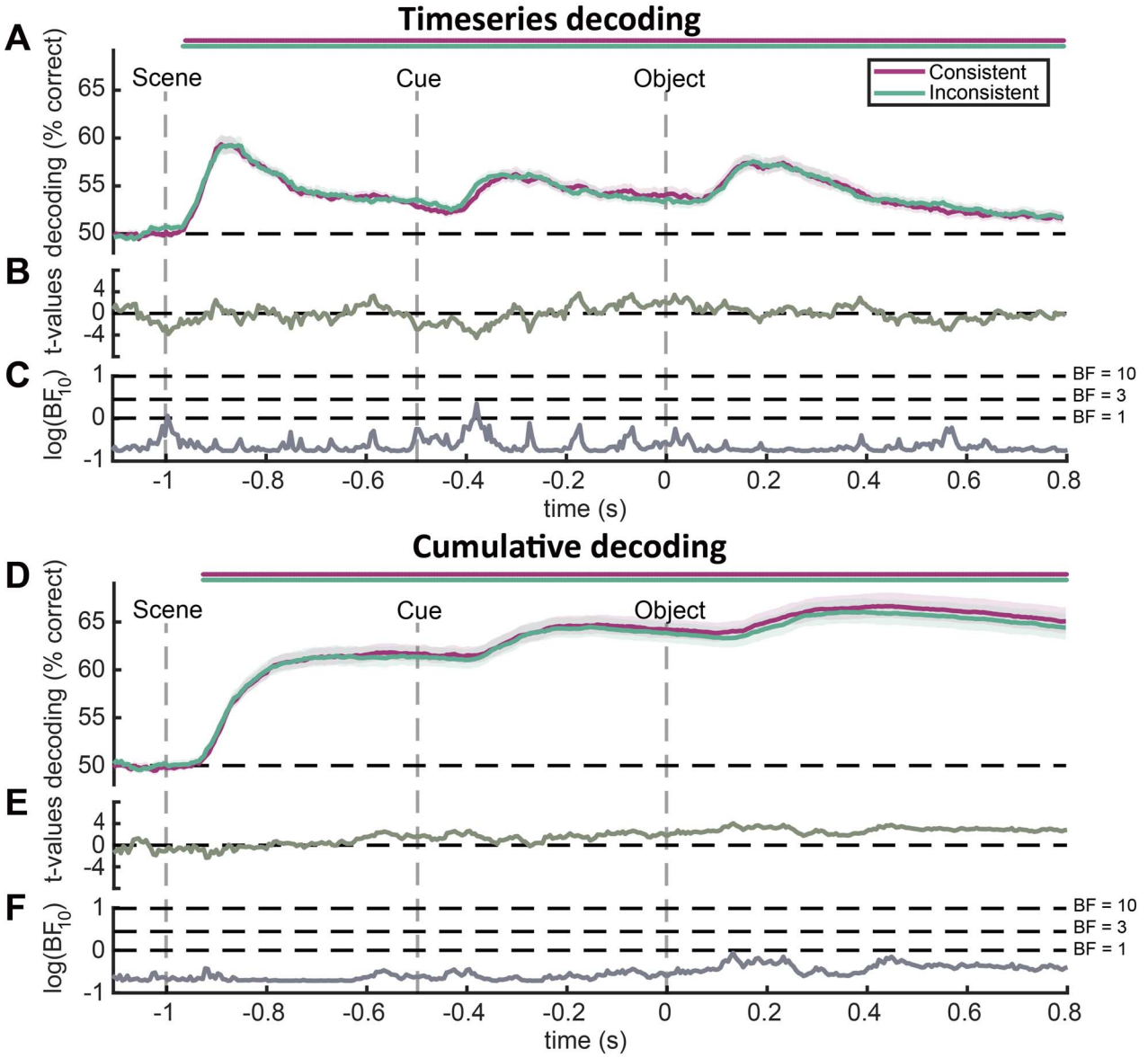

Fig. 4. Decoding between scenes with consistent or inconsistent objects in Experiment 1. (*A*) Timeseries decoding results, separately for scene with consistent and inconsistent objects. Line markers denote significant above-chance decoding ($P < 0.05$, FDR-corrected). (*B*) *t*-values for the comparisons between consistent and inconsistent conditions. (*C*) Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions. For display purposes, the BF_{10} values were log-transformed. Dotted lines show low ($BF_{10} = 1$), moderate ($BF_{10} = 3$), and high ($BF_{10} = 10$) evidence for a difference between conditions. (*D*) Cumulative decoding results, separately for scene with consistent and inconsistent objects. Line markers denote significant above-chance decoding ($P < 0.05$, FDR-corrected). (*E*, *F*) *t*-values and Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions, as in (*B*, *C*). These results show that scene category can be well decoded across time, but the consistency of embedded objects has no facilitatory effects on scene representations.

object. We found significant decoding for both consistent objects (timeseries decoding: 60–800 ms; cumulative decoding: 70–800 ms) and inconsistent objects (timeseries decoding: 70–760 ms; cumulative decoding: 90–800 ms). Critically, we found that the consistent objects were decoded more accurately than inconsistent objects in both the timeseries decoding (310–410 and 545–680 ms) and cumulative decoding analyses (160–190, 255–295, and 370–800 ms) (Fig. 6). Additional electrode-space searchlight analyses suggest that these enhanced representations primarily emerge in posterior electrodes over the right hemisphere (see Supplementary Fig. 4). These results suggest that scene-object consistency can facilitate cortical object representations—but only when the objects are task-relevant. Our data show that such effects arise at least from ~300 ms, although the more sensitive cumulative decoding suggests that such effects may be seen much earlier, even within the first 200 ms of processing. As the current evidence for such early effects is only moderately strong, the exact timing of such effects needs to be confirmed in future studies.

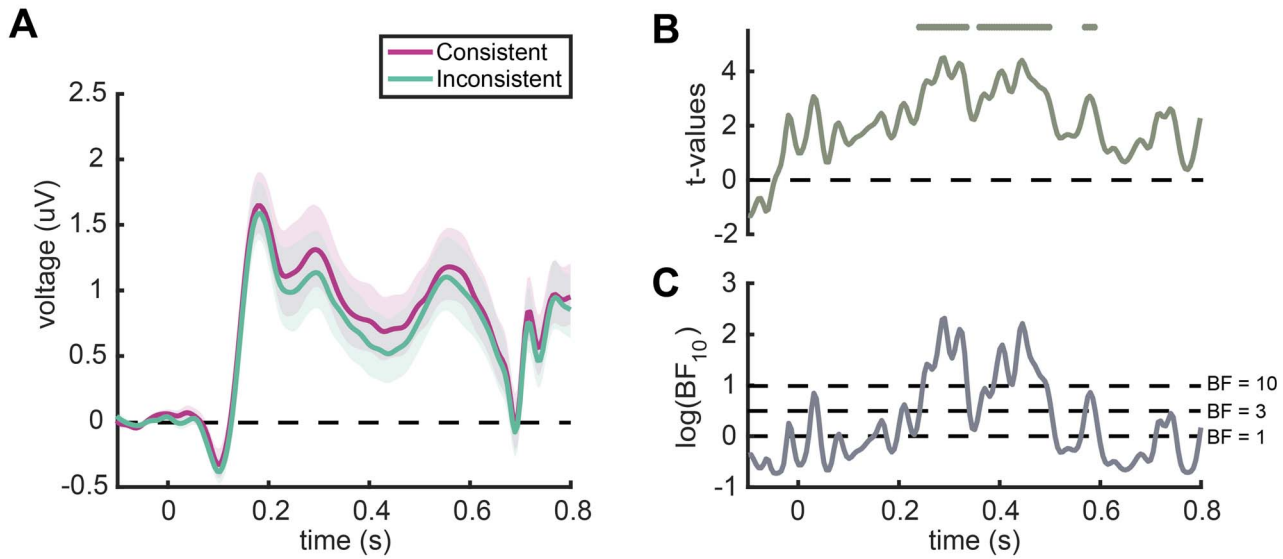

Fig. 5. ERPs in Experiment 2. (*A*) ERPs recorded from the mid-central region for consistent and inconsistent scene-object combinations. Error margins represent standard errors. (*B*) *t*-values for the comparisons between consistent and inconsistent conditions. Line markers denote significant differences between conditions ($P < 0.05$, FDR-corrected). (*C*) Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions. For display purposes, the BF_{10} values were log-transformed. Dotted lines show low ($BF_{10} = 1$), moderate ($BF_{10} = 3$), and high ($BF_{10} = 10$) evidence for a difference between conditions. Similar to Experiment 1, inconsistent scene-object combinations evoked more negative responses at 240–335, 360–500, and 570–590 ms after the object onset relative to consistent combinations.

Tracking the Representation of Scenes with Consistent and Inconsistent Objects

As in Experiment 1, we also tested whether semantically consistent objects can facilitate scene representations. We performed timeseries and cumulative decoding analyses to discriminate between every two scene categories separately for the consistent and inconsistent conditions from −100 to 1800 ms relative to the onset of the scene. We found very similar results as the Experiment 1, with significant decoding for both consistent scenes (timeseries decoding: 55–1800 ms; cumulative decoding: 85–1800 ms) and inconsistent scenes (timeseries decoding: 60–1800 ms; cumulative decoding: 85–1800 ms), but no difference in decoding performance between consistent and inconsistent conditions (Fig. 7). As in Experiment 1, such differences were also not observed when we corrected for multiple comparisons solely between 0 and 800 ms relative to the onset of the object. These results suggest that facilitation effects between scenes and objects are not mutual but that they likely depend on behavioral goals: once the objects were task relevant, we found a facilitation effect originating from semantically consistent scenes.

Comparison across Experiments

The pattern of results across our experiments revealed reliable ERP effects that are independent of task-relevance, but multivariate decoding demonstrated that representational facilitation effects can only be observed when the objects are task relevant. To statistically quantify this pattern, we directly compared the ERP and decoding results between two experiments. For each participant, we computed the difference between the consistent and inconsistent conditions and then compared these differences across experiments using independent-samples *t*-tests. For the ERP results, we found no statistical differences across the experiments (all $P > 0.05$, FDR-corrected; Fig. 8A), suggesting that N300/400 effects emerge independently of the task-relevance of the objects. On the flipside, object representations benefitted more strongly from semantically consistent context when the objects were directly task-relevant: In the timeseries decoding, differences between the two experiments emerged at 215–245, 295–335, and 610–635 ms (max $t(62) = 3.40$, $P = 0.001$), suggesting that during these time points, task-relevance enhances the effect of semantically consistent scene context (Fig. 8B). Clear evidence for this effect was also found in the cumulative decoding, where the effect of semantically consistent scene context was stronger in Experiment 2 between 160–190 and 255–800 ms (Fig. 8C). There were no significant differences in scene decoding between the two experiments.

Finally, we asked whether the difference in results across experiments was statistically different for the ERP and the object decoding analyses. To answer this question, we performed a 2×2 mixed ANOVA with the factors task-relevance (task-irrelevant vs. task-relevant objects, i.e., Experiment 1 vs. Experiment 2) and neural measure (ERP vs. object decoding). For this analysis, we first needed to make the ERP and decoding effects comparable. To achieve this, we first calculated the ERP difference between consistent and inconsistent conditions at each time point for each participant and standardized the difference values by dividing them by the standard deviation of ERP differences within the group at each

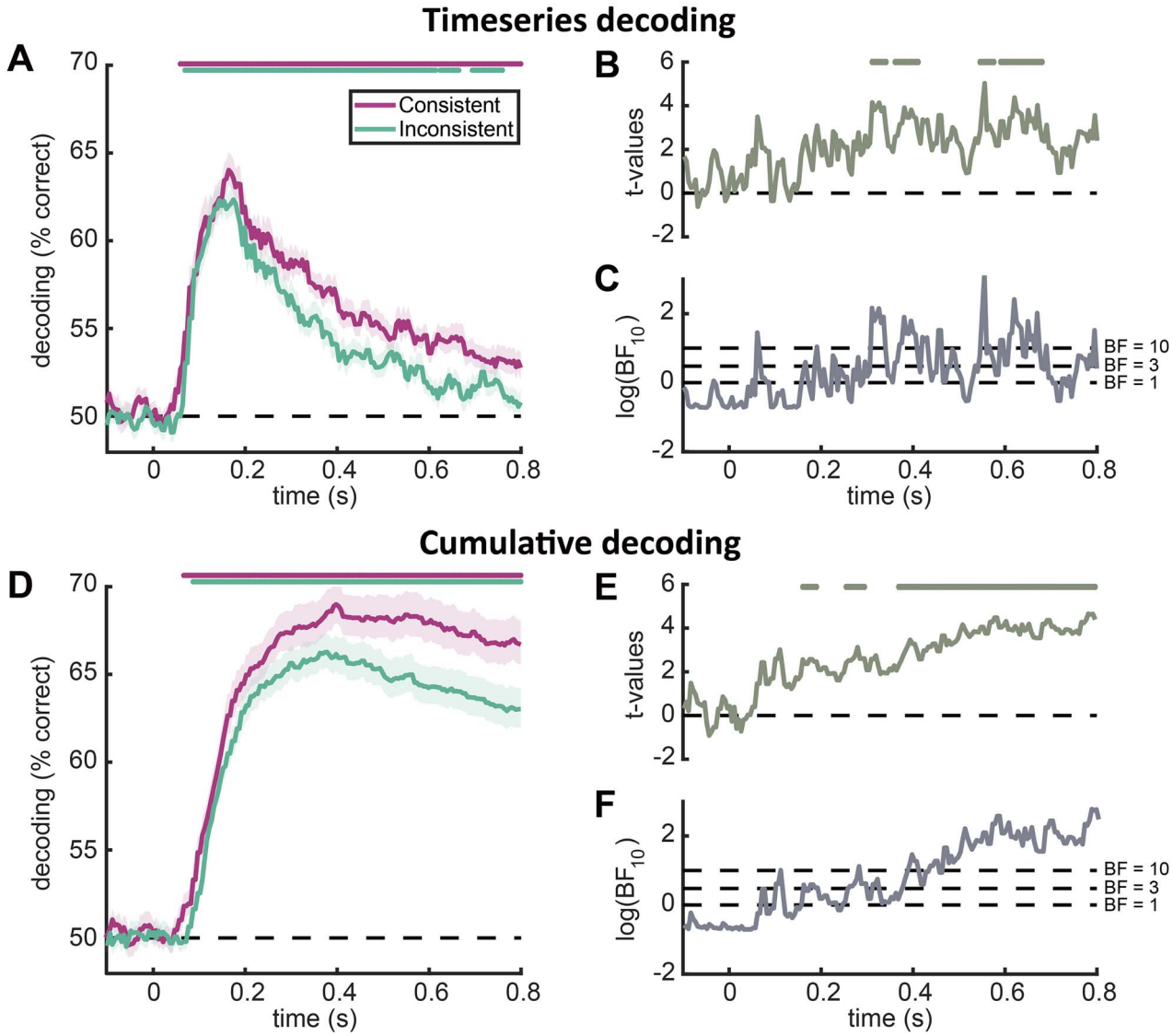

Fig. 6. Decoding for the consistent or inconsistent objects within each scene in Experiment 2. (*A*) Timeseries decoding results, separately for consistent and inconsistent objects. Line markers denote significant above-chance decoding ($P < 0.05$, FDR-corrected). (*B*) *t*-values for the comparisons between consistent and inconsistent conditions. Line markers denote significant differences between the consistent and inconsistent conditions ($P < 0.05$, FDR-corrected). (*C*) Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions. For display purposes, the BF_{10} values were log-transformed. Dotted lines show low ($BF_{10} = 1$), moderate ($BF_{10} = 3$), and high ($BF_{10} = 10$) evidence for a difference between conditions. (*D*) Cumulative decoding results, separately for consistent and inconsistent objects. Line markers denote significant above-chance decoding ($P < 0.05$, FDR-corrected). (*E*, *F*) *t*-values and Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions, as in (*B*, *C*). These results are markedly different from Experiment 1: Scenes can indeed facilitate the cortical representations of consistent objects when the objects are task-relevant.

time point. The effects for timeseries and cumulative object decoding were calculated and standardized in the same way. Next, to obtain a more reliable estimate of the ERP and decoding effects, we used the time span in which the ERP effects emerged in both experiments (between 170 and 590 ms after object onset) to average the ERP and decoding effects across time. Finally, we performed 2 × 2 mixed ANOVAs to test the interaction effect, separately for the timeseries and cumulative decoding results. The expected interaction between task-irrelevance/relevance and ERP/decoding failed to reach significance when looking at the timeseries decoding but revealed a trend ($F(1, 62) = 3.009$, $P = 0.088$; Fig. 8D). When comparing between ERPs and cumulative decoding, the interaction reached significance ($F(1, 62) = 4.296$, $P = 0.042$; Fig. 8E). This result indicates that the effect of task-relevance is significantly larger in the cumulative object decoding than it is in the ERP analysis. This corroborates the notion that the N300/400 effects are dissociable from changes in object representation, as indexed by our decoding analyses.

Altogether, this shows that N300/400 ERP differences emerge independently of task relevance, suggesting that they do not index changes in object representations. In contrast, multivariate decoding reveals that changes in object representations are modulated by task-relevance: when the objects are critical for behavior, semantically

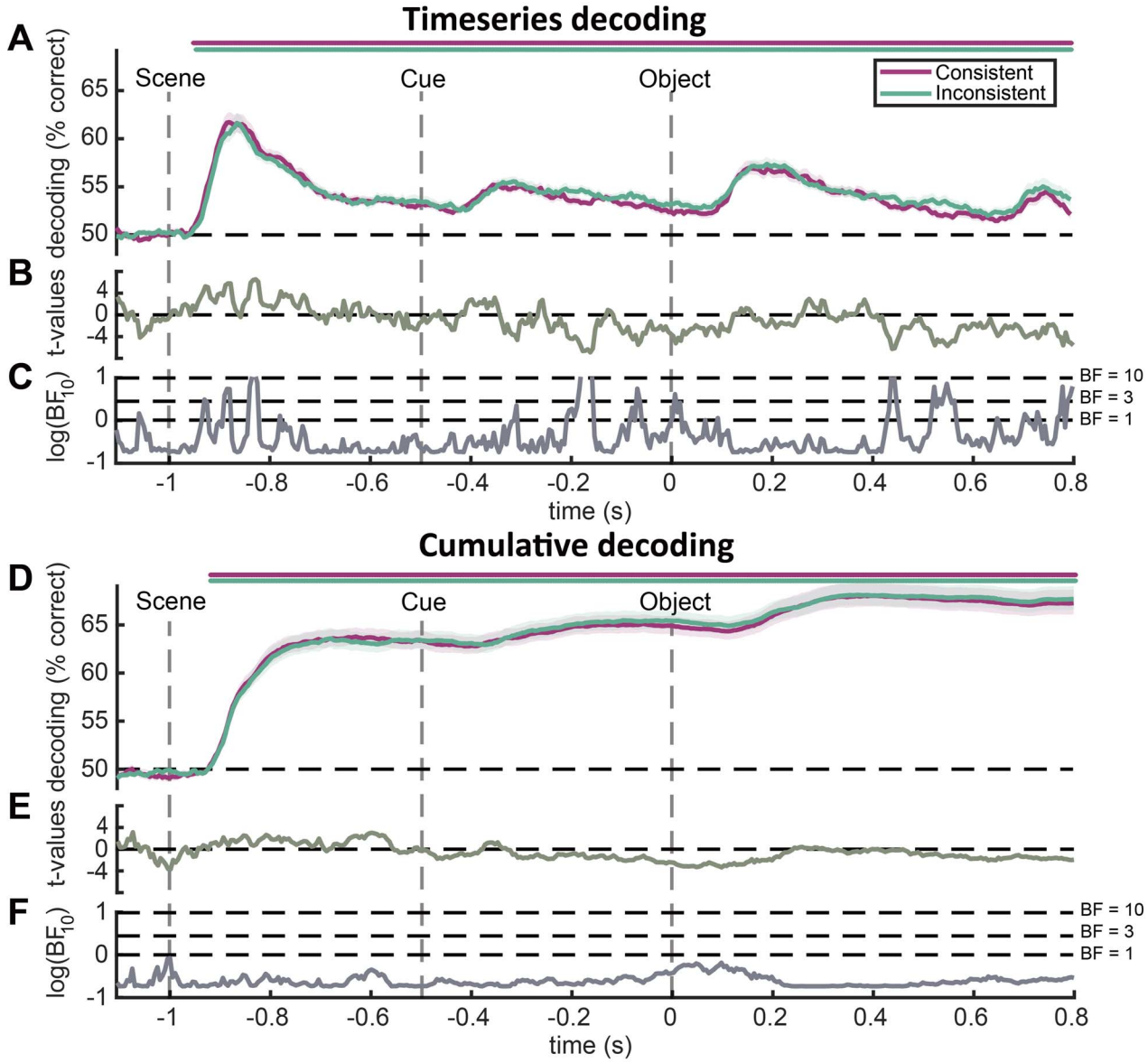

Fig. 7. Decoding between scenes with consistent or inconsistent objects in Experiment 2. (*A*) Timeseries decoding results, separately for scene with consistent and inconsistent objects. Line markers denote significant above-chance decoding ($P < 0.05$, FDR-corrected). (*B*) *t*-values for the comparisons between consistent and inconsistent conditions. (*C*) Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions. For display purposes, the BF_{10} values were log-transformed. Dotted lines show low ($BF_{10} = 1$), moderate ($BF_{10} = 3$), and high ($BF_{10} = 10$) evidence for a difference between conditions. (*D*) Cumulative decoding results, separately for scene with consistent and inconsistent objects. Line markers denote significant above-chance decoding ($P < 0.05$, FDR-corrected). (*E*, *F*) *t*-values and Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions, as in (*B*, *C*). The results show that consistent embedded objects do not automatically facilitate the representation of scenes.

consistent scenes more strongly enhance their cortical representation.

Discussion

In this study, we used EEG to investigate how scene-object consistency affects the quality of object and scene representations. In two experiments, we replicated previous scene-object consistency ERP effects (Mudrik et al. 2010, 2014; Võ and Wolfe 2013), showing that inconsistent scene-object combinations evoked more negative responses in the N300 and N400 windows than consistent combinations. Critically, multivariate decoding analyses revealed whether these scene-object consistency effects in ERPs were accompanied by changes in the quality of cortical object and scene representations. We found that task-irrelevant consistent and inconsistent objects were decoded equally well in Experiment 1, despite pronounced ERP differences in the N300/400 range. When the objects were task-relevant in Experiment 2, we observed a comparable N300/400 ERP effect, which now was accompanied by enhanced object decoding. Across both experiments, we found no significant differences in scene category decoding between consistent and inconsistent conditions. These results suggest that the N300/400 ERP effects are not necessarily

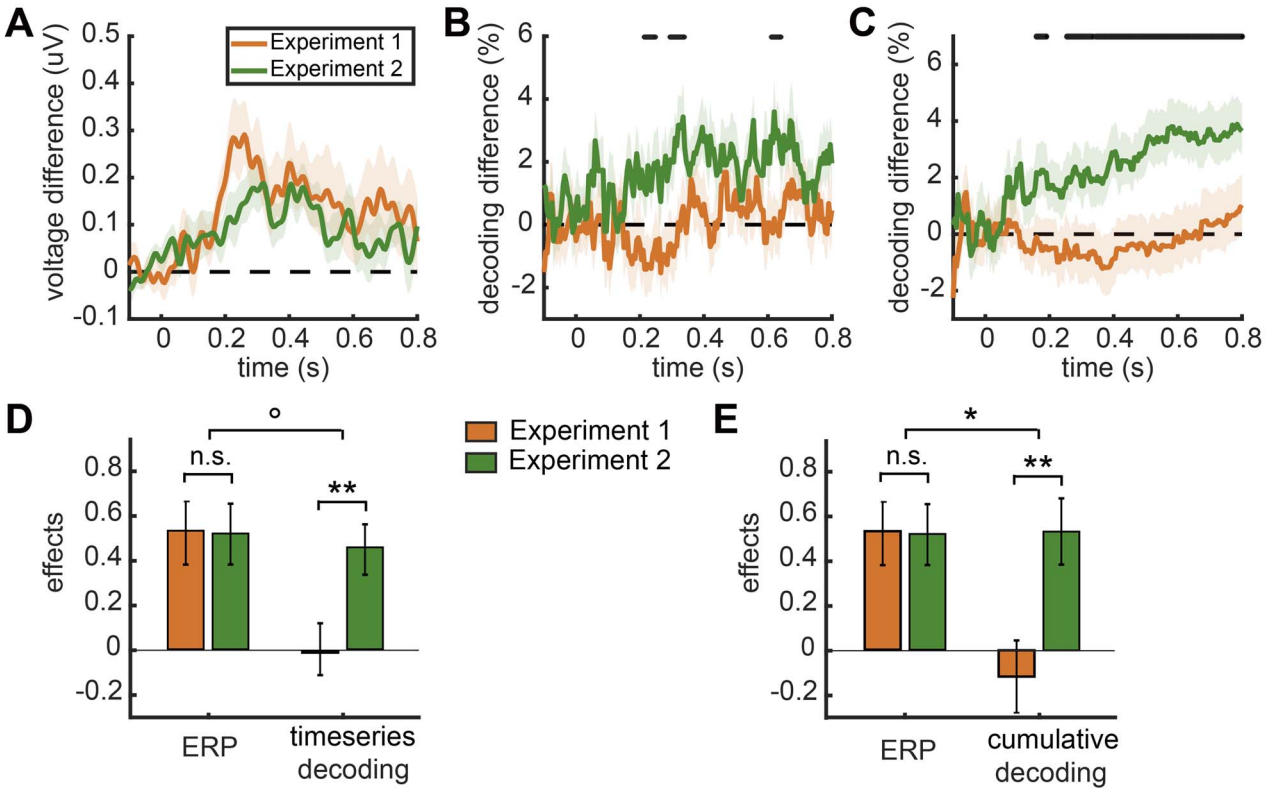

Fig. 8. Comparisons of ERP and decoding effects between experiments. (*A*) Differences in ERP effects (consistent—inconsistent) between experiments. (*B*) Differences in decoding effects (consistent—inconsistent) for object in timeseries decoding analyses between experiments. Line markers denote significant differences between two experiments ($P < 0.05$). (*C*) Differences in decoding effects for object in cumulative decoding analyses between experiments. Line markers denote significant differences between two experiments ($P < 0.05$). These results show that the N300/400 effects emerge independently of the task relevance of the objects, but facilitations from scenes to the representations of objects are task-dependent. (*D*) Standardized consistency effects in ERPs and timeseries object decoding in both experiments, averaged across the time span during which ERP effects were observed in the two experiments (170–590 ms). (*E*) Standardized effects in ERPs and cumulative object decoding in both experiments, averaged across 170–590 ms. The difference between ERP and decoding effects in the two experiments was assessed using a mixed effects ANOVA. Error bars represent standard error of the mean. ° represents $P < 0.1$; * represents $P < 0.05$; ** represents $P < 0.01$.

indicative of enhanced object or scene representations. Furthermore, they suggest that facilitations between objects and scenes are task-dependent rather than automatic.

N300 Effects Do Not Index Changes in Perceptual Processing

The N300 effects found in the study replicated previous findings in studies of scene-object consistency (Võ and Wolfe 2013; Mudrik et al. 2014; Draschkow et al. 2018; Truman and Mudrik 2018; Coco et al. 2020). Particularly, the early N300 effects were often interpreted as reflecting differences in perceptual processing (Schendan and Maher 2009; Mudrik et al. 2010; Dyck and Brodeur 2015; Sauvé et al. 2017; Kumar et al. 2021). Such findings are often explained through models of contextual facilitation (Bar and Ullman 1996; Bar 2004), which propose that object representations are refined by more readily available information about the consistent context. Specifically, when a scene is presented, gist-consistent schemas are rapidly activated through nonselective processing channels (Wolfe et al. 2011). By comparing this rapidly available scene gist to incoming visual information, perceptual uncertainty in object recognition is reduced. However, if the object does not match the scene gist, its identification should be impeded. It was argued that this mismatch between inconsistent objects and the preactivated schemas elicits a larger N300 amplitude, signifying a prediction error that occurs during perceptual object analysis (Kumar et al. 2021).

Our data challenge this interpretation. We show that enhanced N300 amplitudes are observed independently of changes in object decoding. We found reliable N300 differences between consistent and inconsistent objects, which were highly similar for task-relevant and task-irrelevant objects. In contrast, object information, as measured by our multivariate object decoding analyses, was similar for consistent and inconsistent objects when they were not task-relevant; only when they were task-relevant, we found that scene-object consistency facilitated object representations. It is worth noting that both task-relevant and task-irrelevant objects within the scenes could be decoded reliably and with high accuracy in both experiments, which is in line with previous reports (Kaiser et al. 2016); our results therefore cannot be attributed to a failure to decode the objects in the first place.

The pattern of results obtained in our study is therefore inconsistent with the N300 indexing a change in perceptual representations. Our results are rather consistent with an interpretation that views the N300 as a general marker of inconsistency or a purely attentional response to a violation of expectation. On this view, N300 differences are postperceptual in nature. Contrary to the N300, consistency-related differences in the N400 time window are commonly interpreted as a marker of differences in postperceptual semantic processing (Võ and Wolfe 2013; Truman and Mudrik 2018). In fact, a recent study has shown that N400 effects are qualitatively similar to N300 effects (Draschkow et al. 2018), further supporting the view that N300 differences are not directly indicative of changes in perceptual encoding.

When interpreting the results of our study, two limitations should be considered. First, to render the object-level task sufficiently difficult, we drastically reduced the presentation time of the object in Experiment 2 (from 1500 to 83 ms). There is thus a possibility that the longer timing, rather than the lack of behavioral relevance, in Experiment 1 abolished decoding effects in Experiment 1. Further experiments are needed to establish a clear distinction between these explanations. However, if the ERP effects and the decoding effects were indeed a reflection of the same underlying changes in perceptual representations, they should both be affected by object timing—given that we did not observe any change in ERP effects across the two experiments, but a marked difference in object decoding, it is unlikely that the timing difference concealed an otherwise tight correspondence between the two effects. Furthermore, we observed highly comparable overall decoding in the two experiments, which suggests that longer presentation times do not per se alter object decoding. Second, our study compares univariate ERP analyses that average many trials on a small set of electrodes with multivariate decoding analyses that probe a set of pairwise combinations between conditions across large-scale electrode patterns. Ultimately, these different approaches yield different (unknown) sensitivities, so that a comparison between results obtained with the two approaches within a single experiment can be challenging. However, the different results obtained across our two experiments cannot be attributed to an overall sensitivity difference between methods.

Semantic Consistency Only Facilitates Task-Relevant Representations

Our results suggest that cross-facilitation effects between objects and scenes are not automatic but task-dependent. Consistent objects were only decoded better than inconsistent objects in Experiment 2 where they were directly task-relevant, suggesting that semantically consistent scenes only facilitate object processing when the objects are critical for behavior. Furthermore, decoding between the different scenes was similar for scenes that contained consistent and inconsistent objects. As the scenes were never task-relevant, this supports the view that mutual influences between scene and object representations are only observed when they support ongoing behavior.

Several previous neuroimaging studies reported a cross-facilitation between scene and object processing (Brandman and Peelen 2017, 2019; Kaiser et al. 2021), reporting that scenes enhance the cortical representation of objects (Brandman and Peelen 2017; Kaiser et al. 2021), and objects facilitate the representation of scenes (Brandman and Peelen 2019). In these studies, participants were asked to attend the objects or scenes by memorizing them, completing repetition detection tasks, or categorization tasks. One recent study directly compared cross-facilitation effects between objects and scenes under different task demands (Kaiser et al. 2021). In this study, spatially consistent scene context facilitated object representation more than spatially inconsistent scene context when the objects were task-relevant. When participants instead performed a task on the scene, object representations were comparable for the spatially consistent and inconsistent scene contexts.

These results are in line with the current study, in which semantically consistent scene context only facilitated perceptual object processing when it was beneficial for the task at hand. Our findings therefore support a view where the visual system uses contextual information in a flexible and strategic way: when scene context is beneficial for the current task demands, the visual system harnesses contextual information to enhance object representations. Conversely, if the current task does not benefit from contextual information, no cross-facilitation between object and scene processing is found.

What mechanism governs such situational interactions between the scene and object processing systems? A recent TMS study shines new light on how contextual information from a scene may impact object processing in visual cortex (Wischnewski and Peelen 2021): by virtually lesioning key nodes of the scene and object processing networks, they established that information is first processed in parallel in object- and scene-selective cortices (until ~200 ms of processing), before information from scene-selective brain regions converges with object coding in object-selective regions (after 250 ms of processing). Our results suggest that the information flow from scene-selective to object-selective cortex is gated by behavioral demands: when the task requires perceiving object details, the brain may use scene representations to actively generate predictions about the objects that are likely to appear in the scene (Hochstein and Ahissar 2002; Bar 2004). By feeding back these predictions to object-selective cortex, the perceptual representation of consistent object information is then facilitated, for instance by sharpening the neural response (de Lange et al. 2018). When the task does not require

perceiving object detail, such predictions may simply not be generated—or not generated to the same extent. This showcases how predictive processing is adaptive tailored to situational needs.

Conclusion

In the study, we investigated how scene-object consistency affects scene and object representations. Our results suggest that differences in the N300/400 components related to scene-object consistency do not directly index differences in perceptual representations, but rather reflect a generic marker of semantic violations. Furthermore, they suggest that facilitation effects between objects and scenes are task-dependent rather than automatic. Our findings highlight that there are multiple markers of semantic consistency that reflect different underlying brain mechanisms. How these mechanisms interact to support efficient real-world vision needs to be explored in future studies.

Supplementary Material

Supplementary material can be found at *Cerebral Cortex* online.

Notes

The authors thank the HPC Service of ZEDAT, Freie Universität Berlin, for computing time and support. *Conflict of Interests*: The authors declare no competing interests.

Funding

Deutsche Forschungsgemeinschaft (grants CI241/1-1, CI241/3-1, CI241/7-1, KA4683/2-1 to D.K. and R.M.C.); European Research Council (grant 803370 to R.M.C.); Chinese Scholarship Council (CSC) to L.C.

References

Bar M. 2004. Visual objects in context. *Nat Rev Neurosci*. 5:617–629.

Bar M, Ullman S. 1996. Spatial context in recognition. *Perception*. 25: 343–352.

Biederman I, Mezzanotte RJ, Rabinowitz JC. 1982. Scene perception: detecting and judging objects undergoing relational violations. *Cogn Psychol*. 14:143–177.

Boring MJ, Ridgeway K, Shvartsman M, Jonker TR. 2020. Continuous decoding of cognitive load from electroencephalography reveals task-general and task-specific correlates. *J Neural Eng*. 17: 056016.

Boyce SJ, Pollatsek A. 1992. Identification of objects in scenes: the role of scene background in object naming. *J Exp Psychol Learn Mem Cogn*. 18:531–543.

Boyce SJ, Pollatsek A, Rayner K. 1989. Effect of background information on object identification. *J Exp Psychol Hum Percept Perform*. 15: 556–566.

Brainard DH. 1997. The psychophysics toolbox. *Spat Vis*. 10:433–436.

Brandman T, Peelen MV. 2017. Interaction between scene and object processing revealed by human fMRI and MEG decoding. *J Neurosci*. 37:7700–7710.

Brandman T, Peelen MV. 2019. Signposts in the fog: objects facilitate scene representations in left scene-selective cortex. *J Cogn Neurosci*. 31:390–400.

Coco MI, Nuthmann A, Dimigen O. 2020. Fixation-related brain potentials during semantic integration of object–scene information. *J Cogn Neurosci*. 32:571–589.

Cornelissen TH, Võ ML. 2017. Stuck on semantics: processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Atten Percept Psychophys*. 79:154–168.

Davenport JL. 2007. Consistency effects between objects in scenes. *Mem Cognit*. 35:393–401.

Davenport JL, Potter MC. 2004. Scene consistency in object and background perception. *Psychol Sci*. 15:559–564.

de Lange FP, Heilbron M, Kok P. 2018. How do expectations shape perception? *Trends Cogn Sci*. 22:764–779.

Draschkow D, Heikel E, Võ ML, Fiebach CJ, Sassenhagen J. 2018. No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing. *Neuropsychologia*. 120:9–17.

Dyck M, Brodeur MB. 2015. ERP evidence for the influence of scene context on the recognition of ambiguous and unambiguous objects. *Neuropsychologia*. 72:43–51.

Ganis G, Kutas M. 2003. An electrophysiological study of scene effects on object identification. *Cogn Brain Res*. 16:123–144.

Grootswagers T, Wardle SG, Carlson TA. 2017. 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.

Hochstein S, Ahissar M. 2002. View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*. 36: 791–804.

Kaiser D, Häberle G, Cichy RM. 2020a. Real-world structure facilitates the rapid emergence of scene category information in visual brain signals. *J Neurophysiol*. 124:145–151.

Kaiser D, Häberle G, Cichy RM. 2021. Coherent natural scene structure facilitates the extraction of task-relevant object information in visual cortex. *Neuroimage*. 240:118365.

Kaiser D, Inciuraite G, Cichy RM. 2020b. Rapid contextualization of fragmented scene information in the human visual system. *Neuroimage*. 219:117045.

Kaiser D, Nyga K. 2020. Tracking cortical representations of facial attractiveness using time-resolved representational similarity analysis. *Sci Rep*. 10:1–10.

Kaiser D, Oosterhof NN, Peelen MV. 2016. The neural dynamics of attentional selection in natural scenes. *J Neurosci*. 36: 10522–10528.

Kaiser D, Quek GL, Cichy RM, Peelen MV. 2019. Object vision in a structured world. *Trends Cogn Sci*. 23:672–685.

Kumar M, Federmeier KD, Beck DM. 2021. The N300: an index for predictive coding of complex visual objects and scenes. *Cereb Cortex Commun*. 2:tgab030.

Lowe MX, Rajsic J, Ferber S, Walther DB. 2018. Discriminating scene categories from brain activity within 100 milliseconds. *Cortex*. 106:275–287.

Mudrik L, Lamy D, Deouell LY. 2010. ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*. 48:507–517.

Mudrik L, Shalgi S, Lamy D, Deouell LY. 2014. Synchronous contextual irregularities affect early scene processing: replication and extension. *Neuropsychologia*. 56:447–458.

Munneke J, Brentari V, Peelen M. 2013. The influence of scene context on object recognition is independent of attentional focus. *Front Psychol*. 4:552.

Oostenveld R, Fries P, Maris E, Schoffelen J-M. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci*. 2011:156869.

Oosterhof NN, Connolly AC, Haxby JV. 2016. CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU octave. *Front Neuroinform*. 10:27.

Palmer SE. 1975. The effects of contextual scenes on the identification of objects. *Mem Cognit*. 3:519–526.

Ramkumar P, Jas M, Pannasch S, Hari R, Parkkonen L. 2013. Feature-specific information processing precedes concerted activation in human visual cortex. *J Neurosci*. 33:7691–7699.

Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. 2009. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon Bull Rev*. 16:225–237.

Sauvé G, Harmand M, Vanni L, Brodeur MB. 2017. The probability of object–scene co-occurrence influences object identification processes. *Exp Brain Res*. 235:2167–2179.

Schendan HE, Maher SM. 2009. Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *Neuroimage*. 44:1423–1438.

Truman A, Mudrik L. 2018. Are incongruent objects harder to identify? The functional significance of the N300 component. *Neuropsychologia*. 117:222–232.

van Driel J, Olivers CNL, Fahrenfort JJ. 2021. High-pass filtering artifacts in multivariate classification of neural time series data. *J Neurosci Methods*. 352:109080.

Võ ML, Boettcher SE, Draschkow D. 2019. Reading scenes: how scene grammar guides attention and aids perception in real-world environments. *Curr Opin Psychol*. 29:205–210.

Võ ML, Henderson JM. 2009. Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *J Vis*. 9:24.1–24.2415.

Võ ML, Henderson JM. 2011. Object–scene inconsistencies do not capture gaze: evidence from the flash-preview moving-window paradigm. *Atten Percept Psychophys*. 73:1742–1753.

Võ ML, Wolfe JM. 2013. Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychol Sci*. 24: 1816–1823.

Wischnewski M, Peelen MV. 2021. Causal neural mechanisms of context-based object recognition. *Elife*. 10:e69736.

Wolfe JM, Võ ML, Evans KK, Greene MR. 2011. Visual search in scenes involves selective and nonselective pathways. *Trends Cogn Sci*. 15: 77–84.

Supplementary Information

Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations

Lixiang Chen, Radoslaw Martin Cichy, Daniel Kaiser

Index of figures

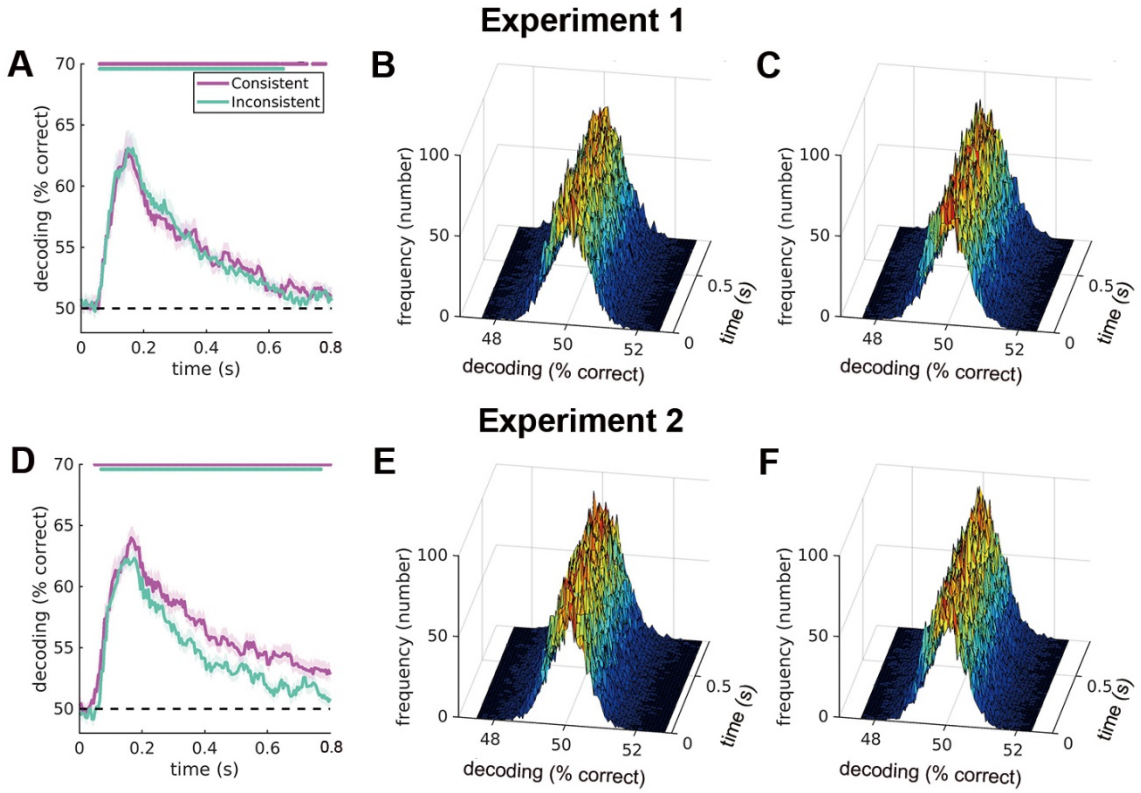

Fig. S1 Permutation-test for timeseries object decoding against chance level. **(A)** Timeseries decoding results in Experiment 1, separately for consistent and inconsistent objects. Line markers denote significant above-chance decoding (permutation-test 1,000 times, FDR-corrected). **(B)** Null distribution for the timeseries decoding of consistent objects in Experiment 1 at each time point obtained from the permutation test. **(C)** Null distribution for the timeseries decoding of inconsistent objects in Experiment 1 at each time point obtained from the permutation test. **(D)** Timeseries decoding results in Experiment 2. **(E)** Null distribution for the timeseries decoding of consistent objects in Experiment 2 at each time point. **(F)** Null distribution for the timeseries decoding of inconsistent objects in Experiment 2 at each time point. These results show the empirical chance levels for the timeseries decoding of both consistent and inconsistent objects in both experiments are centered around 50% and the significant timepoints are virtually identical to our original approach.

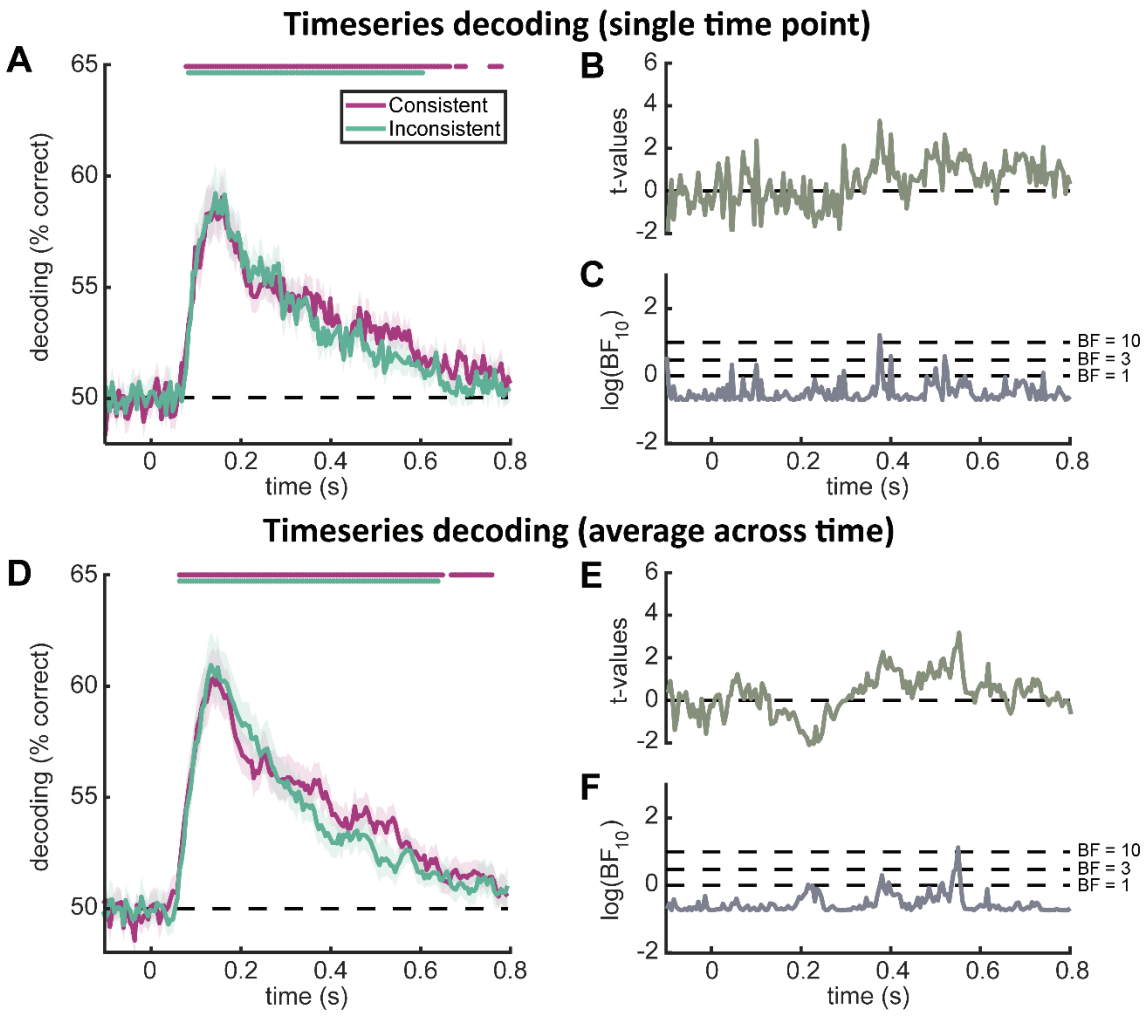

Fig. S2 Decoding of consistent and inconsistent objects in Experiment 1. **(A)** Results of timeseries decoding using the data from a single time point without using sliding windows, separately for consistent and inconsistent objects. Line markers denote significant above-chance decoding ($p < 0.05$, FDR-corrected). **(B)** *t*-values for the comparisons between consistent and inconsistent conditions. **(C)** Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions. For display purposes, the BF_{10} values were log-transformed. Dotted lines show low ($BF_{10} = 1$), moderate ($BF_{10} = 3$), and high ($BF_{10} = 10$) evidence for a difference between conditions. **(D)** Results of timeseries decoding using the data averaged across time within each 50 ms sliding window, separately for consistent and inconsistent objects. **(E, F)** *t*-values and Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions, as in (B, C). These results are very similar to the results of our timeseries decoding using sliding windows.

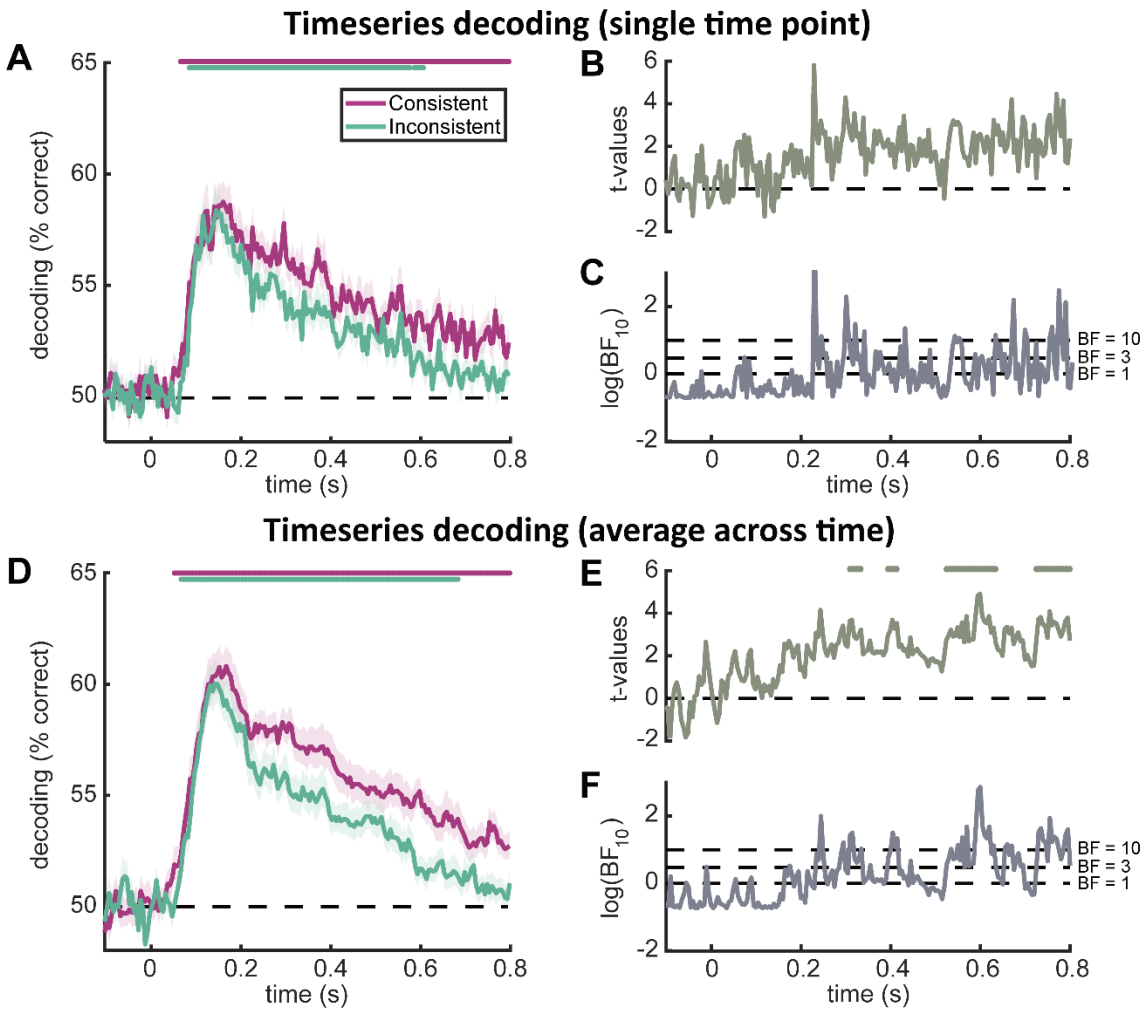

Fig. S3 Decoding of consistent and inconsistent objects in Experiment 2. **(A)** Results of timeseries decoding using the data from a single time point without using sliding windows, separately for consistent and inconsistent objects. Line markers denote significant above-chance decoding (p < 0.05, FDR-corrected). **(B)** t-values for the comparisons between consistent and inconsistent conditions. Line markers denote significant differences between the consistent and inconsistent conditions (p < 0.05, FDR-corrected). **(C)** Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions. For display purposes, the BF_{10} values were log-transformed. Dotted lines show low (BF_{10} = 1), moderate (BF_{10} = 3), and high (BF_{10} = 10) evidence for a difference between conditions. **(D)** Results of timeseries decoding using the data averaged across time within each 50 ms sliding window, separately for consistent and inconsistent objects. **(E, F)** t-values and Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions, as in (B, C). Results for the analysis on averaged time windows are similar to our original results using sliding windows: the consistent objects are decoded better than inconsistent objects when the objects are task relevant. The results on individual time points are also qualitatively similar, but the differences failed to reach statistical significance.

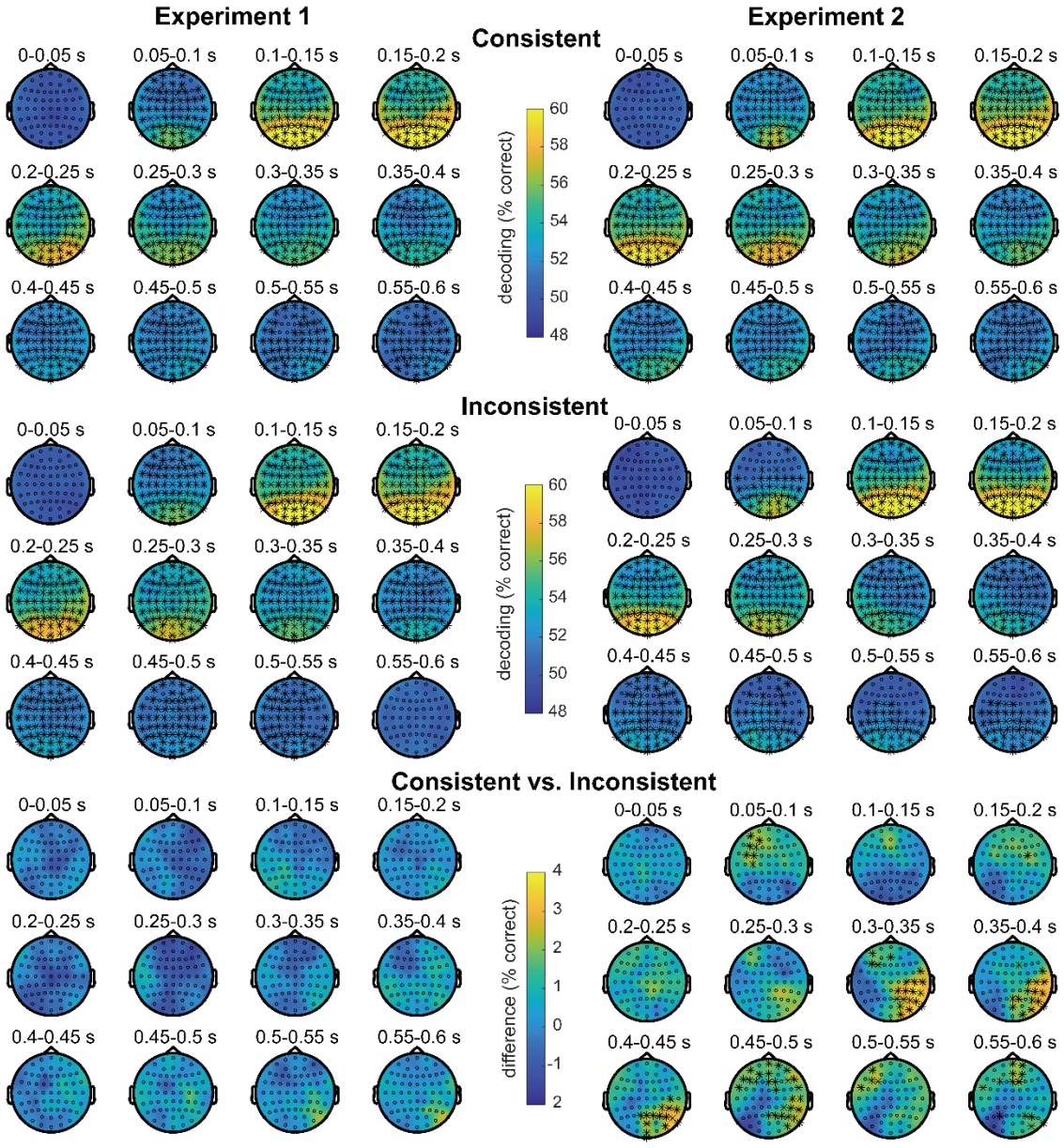

Fig. S4 Topographies of decoding accuracy for consistent and inconsistent objects in Experiments 1 and 2. Here, the timeseries decoding analysis was repeated for a moving spherical neighborhood of 11 electrodes and accuracies were mapped back onto the scalp location of the central electrode in the neighborhood. The results show that both consistent and inconsistent objects could be decoded across the scalp and from 50-100 ms after object onset, with the strongest decoding in posterior sensors located over visual cortex. Differences between consistent and inconsistent object decoding in Experiment 2 emerged primarily in right-posterior electrodes. For this analysis, electrodes removed during preprocessing were interpolated using data from neighboring electrodes. Asterisks (*) indicate the significant results of comparisons against chance level or between conditions, after FDR correction across electrodes.

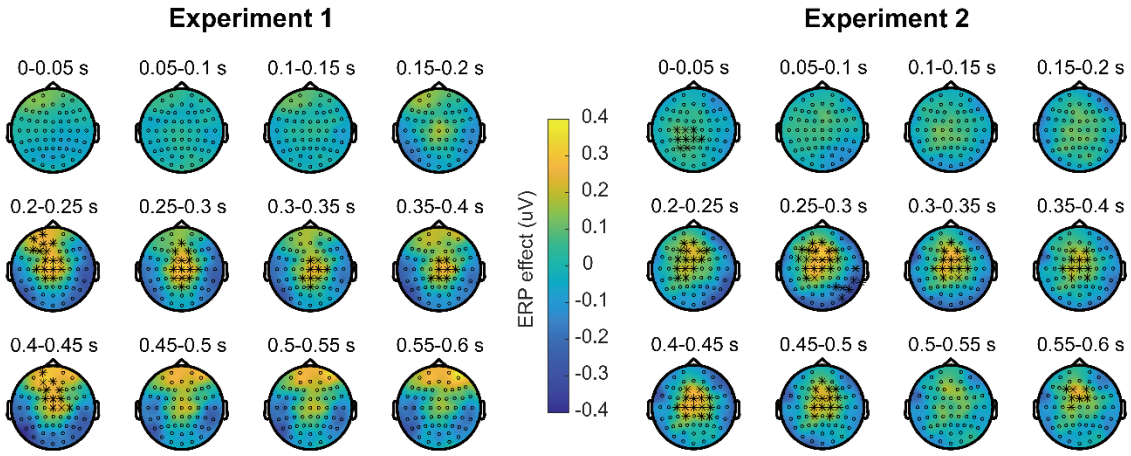

Fig. S5 Topographies of ERP differences between consistent and inconsistent scene-object combinations in Experiments 1 and 2. Corroborating our electrode selection for main ERP analysis, the most significant scene consistency effects emerged in mid-central region. For this analysis, electrodes removed during preprocessing were interpolated using data from neighboring electrodes. Asterisks (*) indicate the significant differences between the consistent and inconsistent conditions, after FDR correction across electrodes.

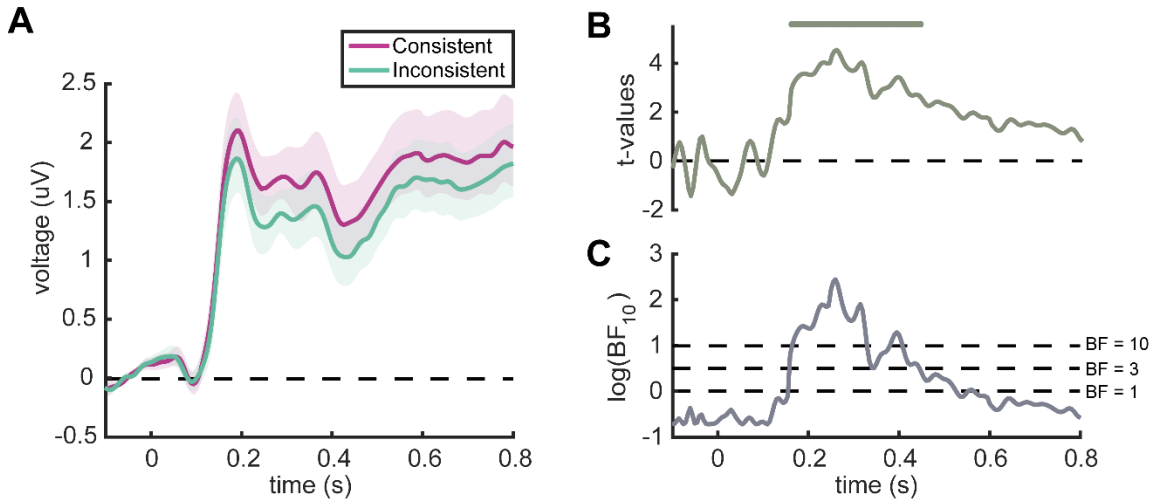

Fig. S6 Event-related potentials (ERPs) in Experiment 1, with filtering performed before epoching. **(A)** ERPs recorded from the mid-central region for consistent and inconsistent scene-object combinations. Error margins represent standard errors. **(B)** *t*-values for the comparisons between consistent and inconsistent conditions. Line markers denote significant differences between conditions ($p < 0.05$, FDR-corrected). **(C)** Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions. For display purposes, the BF_{10} values were log-transformed. Dotted lines show low ($BF_{10} = 1$), moderate ($BF_{10} = 3$), and high ($BF_{10} = 10$) evidence for a difference between conditions. Similar to our original analysis, in which filtering was performed after the other preprocessing steps, inconsistent scene-object combinations evoked more negative responses than consistent combinations at 160-445 ms after object onset.

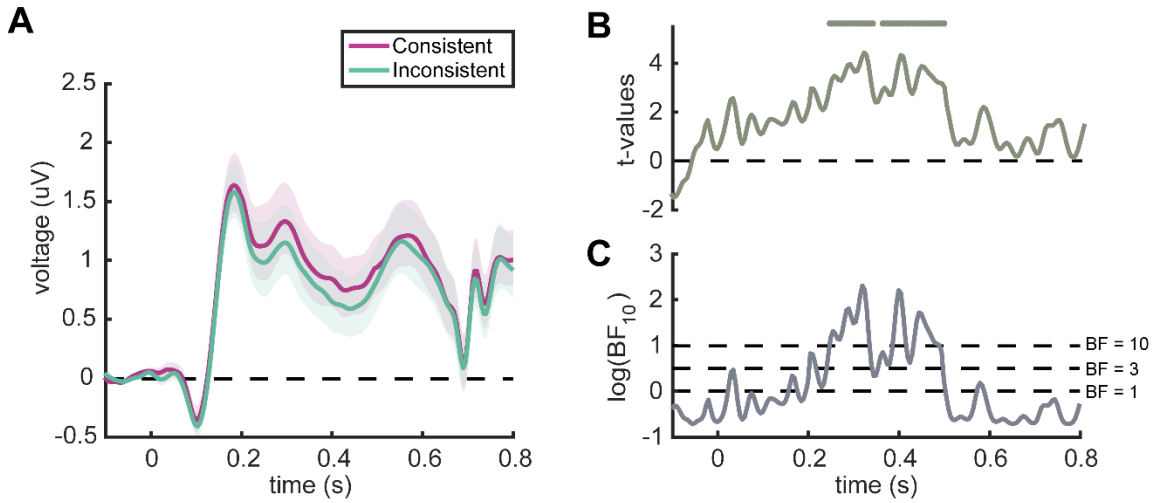

Fig. S7 Event-related potentials (ERPs) in Experiment 2, with filtering performed before epoching. **(A)** ERPs recorded from the mid-central region for consistent and inconsistent scene-object combinations. Error margins represent standard errors. **(B)** *t*-values for the comparisons between consistent and inconsistent conditions. Line markers denote significant differences between conditions ($p < 0.05$, FDR-corrected). **(C)** Bayes factors (BF_{10}) for the comparisons between consistent and inconsistent conditions. For display purposes, the BF_{10} values were log-transformed. Dotted lines show low ($BF_{10} = 1$), moderate ($BF_{10} = 3$), and high ($BF_{10} = 10$) evidence for a difference between conditions. As for Experiment 1, and similar to our original analysis, inconsistent scene-object combinations evoked more negative responses at 245-340 ms and 360-495 ms after object onset, relative to consistent combinations.

5.2 Original publication of Study 2

Chen, L., Cichy, R. M.*, & Kaiser, D.* (2023). Alpha -frequency feedback to early visual cortex orchestrates coherent naturalistic vision. Science Advances, 9(45), eadi2321. https://doi.org/10.1126/sciadv.adi2321

* The authors contributed equally.

Copyright

COGNITIVE NEUROSCIENCE

Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision

Lixiang Chen[1]*, Radoslaw M. Cichy[1]†, Daniel Kaiser[2,3]*†

During naturalistic vision, the brain generates coherent percepts by integrating sensory inputs scattered across the visual field. Here, we asked whether this integration process is mediated by rhythmic cortical feedback. In electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) experiments, we experimentally manipulated integrative processing by changing the spatiotemporal coherence of naturalistic videos presented across visual hemifields. Our EEG data revealed that information about incoherent videos is coded in feedforward-related gamma activity while information about coherent videos is coded in feedback-related alpha activity, indicating that integration is indeed mediated by rhythmic activity. Our fMRI data identified scene-selective cortex and human middle temporal complex (hMT) as likely sources of this feedback. Analytically combining our EEG and fMRI data further revealed that feedback-related representations in the alpha band shape the earliest stages of visual processing in cortex. Together, our findings indicate that the construction of coherent visual experiences relies on cortical feedback rhythms that fully traverse the visual hierarchy.

INTRODUCTION

We consciously experience our visual surroundings as a coherent whole that is phenomenally unified across space (*1*, *2*). In our visual system, however, inputs are initially transformed into a spatially fragmented mosaic of local signals that lack integration. How does the brain integrate this fragmented information across the subsequent visual processing cascade to mediate unified perception?

Classic hierarchical theories of vision posit that integration is solved during feedforward processing (*3*, *4*). On this view, integration is hard wired into the visual system: Local representations of specific features are integrated into more global representations of meaningful visual contents through hierarchical convergence over features distributed across visual space.

More recent theories instead posit that visual integration is achieved through complex interactions between feedforward information flow and dynamic top-down feedback (*5*–*7*). On this view, feedback information flow from downstream adaptively guides the integration of visual information in upstream regions. Such a conceptualization is anatomically plausible, as well as behaviorally adaptive, as higher-order regions can flexibly adjust current whether or not stimuli are integrated through the visual system's abundant top-down connections (*8*–*10*).

However, the proposed interactions between feedforward and feedback information pose a critical challenge: Feedforward and feedback information needs to be multiplexed across the visual hierarchy to avoid unwanted interferences through spurious interactions of these signals. Previous studies propose that neural systems meet this challenge by routing feedforward and feedback information in different neural frequency channels: High-frequency gamma (31 to 70 Hz) rhythms may mediate feedforward propagation, whereas low-frequency alpha (8 to 12 Hz) and beta (13 to 30 Hz) rhythms carry predictive feedback to upstream areas (*11*–*14*).

Here, we set out to test the hypothesis that rhythmic coding acts as a mechanism mediating coherent visual perception. We used a novel experimental paradigm that manipulated the degree to which stimuli could be integrated across space through the spatiotemporal coherence of naturalistic videos shown in the two visual hemifields. Combining electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) measurements, we show that when inputs are integrated into a coherent percept, cortical alpha dynamics carry stimulus-specific feedback from high-level visual cortex to early visual cortex. Our results show that spatial integration of naturalistic visual inputs is mediated by feedback dynamics that traverse the visual hierarchy in low-frequency alpha rhythms.

RESULTS

We experimentally mimicked the spatially distributed nature of naturalistic inputs by presenting eight 3-s naturalistic videos (Fig. 1A) through two circular apertures right and left of fixation (diameter, 6° visual angle; minimal distance to fixation, 2.64°). To assess spatial integration in a controlled way, we varied how the videos were presented through these apertures (Fig. 1B): In the right- or left-only condition, the video was shown only through one of the apertures, providing a baseline for processing inputs from one hemifield, without the need for spatial integration across hemifields. In the coherent condition, the same original video was shown through both apertures. Here, the input had the spatiotemporal statistics of a unified scene expected in the real world and could thus be readily integrated into a coherent unitary percept. In the incoherent condition, by contrast, the videos shown through the two apertures stemmed from two different videos (see fig. S1). Here, the input did not have the spatiotemporal real-world statistics of a unified scene and thus could not be readily integrated. Contrasting brain activity for the coherent and incoherent condition thus reveals

[1]Department of Education and Psychology, Freie Universität Berlin, Berlin 14195, Germany. [2]Mathematical Institute, Department of Mathematics and Computer Science, Physics, Geography, Justus-Liebig-Universität Gießen, Gießen 35392, Germany. [3]Center for Mind, Brain and Behavior (CMBB), Philipps-Universität Marburg and Justus-Liebig-Universität Gießen, Marburg 35032, Germany.
*Corresponding author. Email: lixiang.chen@fu-berlin.de (L.C.); danielkaiser.net@gmail.com (D.K.).
†These authors contributed equally to this work.

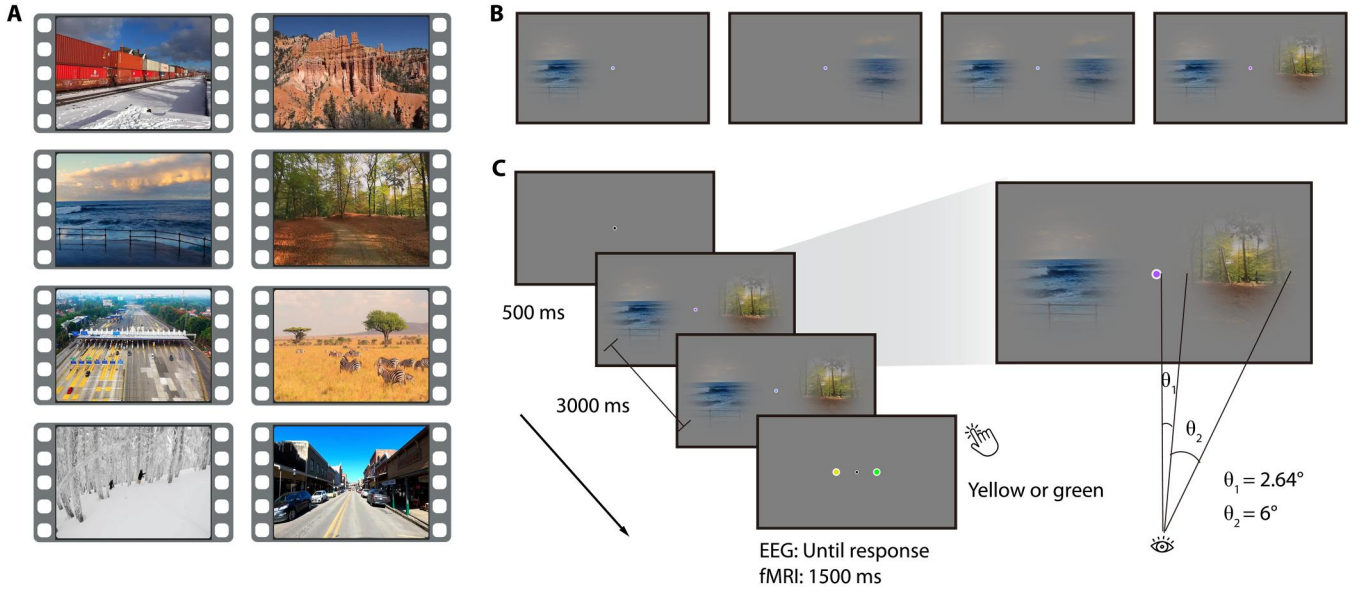

Fig. 1. Stimuli and experimental design. (**A**) Snapshots from the eight videos used. (**B**) In the experiment, videos were either presented through one aperture in the right or left visual field or through both apertures in a coherent or incoherent way. (**C**) During the video presentation, the color of the fixation dot changed periodically (every 200 ms). Participants reported whether a green or yellow fixation dot was included in the sequence.

neural signatures of spatial integration into unified percepts across visual space.

Participants viewed the video stimuli in separate EEG (n = 48) and fMRI (n = 36) recording sessions. Participants performed an unrelated central task (Fig. 1C) to ensure fixation and to allow us to probe integration processes in the absence of explicit task demands.

Harnessing the complementary frequency resolution and spatial resolution of our EEG and fMRI recordings, we then delineated how inputs that either can or cannot be integrated into a coherent percept are represented in rhythmic neural activity and regional activity across the visual hierarchy. Specifically, we decoded between the eight different video stimuli in each of the four conditions from frequency-resolved EEG sensor patterns (*15*, *16*) and from spatially resolved fMRI multivoxel patterns (*17*).

Rhythmic brain dynamics mediate integration across visual space

Our first key analysis determined how the feedforward and feedback information flows involved in the processing and integrating visual information across space are multiplexed in rhythmic codes. We hypothesized that conditions not affording integration lead to neural coding in feedforward-related gamma activity (*11*, *14*), whereas conditions that allow for spatiotemporal integration lead to coding in feedback-related alpha/beta activity (*11*, *14*).

To test this hypothesis, we decoded the video stimuli from spectrally resolved EEG signals, aggregated within the alpha, beta, and gamma frequency bands, during the whole stimulus duration (Fig. 2A; see Materials and Methods for details). Our findings supported our hypothesis. We observed that incoherent video stimuli, as well as single video stimuli, were decodable only from the gamma frequency band [all $t(47) > 3.41$, $P < 0.001$; Fig. 2, B and C]. By stark contrast, coherent video stimuli were decodable only from the alpha-frequency band [$t(47) = 5.43$, $P < 0.001$; Fig. 2C]. Comparing the pattern of decoding performance across frequency bands revealed that incoherent video stimuli were better decodable than coherent stimuli from gamma responses [$t(47) = 3.04$, $P = 0.004$] and coherent stimuli were better decoded than incoherent stimuli from alpha responses [$t(47) = 2.32$, $P = 0.025$; interaction: $F(2, 94) = 7.47$, $P < 0.001$; Fig. 2C]. The observed effects also held when analyzing the data continuously across frequency space rather than aggregated in predefined frequency bands (see fig. S2) and were not found trivially in the evoked broadband responses (see fig. S3). We also analyzed the theta (4 to 7 Hz) and high-gamma bands (71 to 100 Hz) using the decoding analysis. For the theta band, we did not find any significant decoding (see figs. S2 and S4). The results from the high-gamma band were highly similar to the results obtained for the lower-gamma frequency range (see figs. S2 and S4). In addition, we conducted both univariate and decoding analyses on time- and frequency-resolved responses, but neither of these analyses revealed any differences between the coherent and incoherent conditions (see fig. S5), indicating a lack of statistical power for resolving the data both in time and frequency. Together, our results demonstrate the multiplexing of visual information in rhythmic information flows. When no integration across hemifields was required, visual feedforward activity is carried by gamma rhythms. When spatiotemporally coherent inputs allow for integration, integration-related feedback activity is carried by alpha rhythms.

The observation of a frequency-specific channel for feedback information underlying spatial integration immediately poses two questions: (i) Where does this feedback originate from? and (ii) Where is this feedback heading? We used fMRI recordings to answer these two questions in turn.

Scene-selective cortex is the source of integration-related feedback

To reveal the source of the feedback, we evaluated how representations across visual cortex differ between stimuli that can or cannot

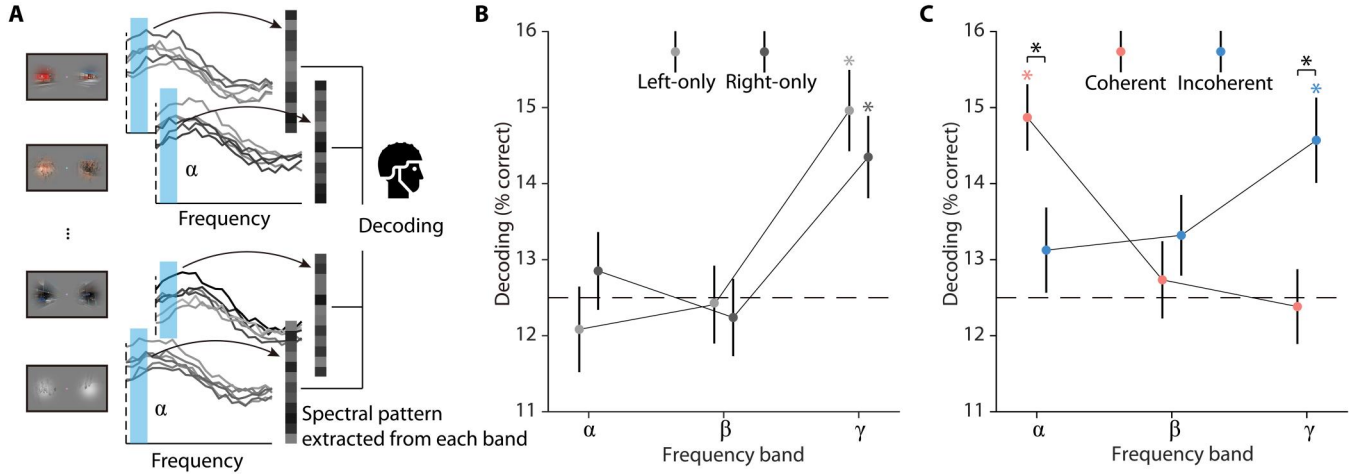

Fig. 2. EEG decoding analysis. (**A**) Frequency-resolved EEG decoding analysis. In each condition, we used eight-way decoding to classify the video stimuli from patterns of spectral EEG power across electrodes, separately for each frequency band (alpha, beta, and gamma). (**B** and **C**) Results of EEG frequency-resolved decoding analysis. The incoherent and single video stimuli were decodable from gamma responses, whereas the coherent stimuli were decodable from alpha responses, suggesting a switch from dominant feedforward processing to the recruitment of cortical feedback. Error bars represent SEs. *$P < 0.05$ (FDR-corrected).

be integrated across space (Fig. 3A). We reasoned that regions capable of exerting integration-related feedback should show stronger representations of spatiotemporally coherent inputs that can be integrated, compared to incoherent inputs that do not. Scene-selective areas in visual cortex are a strong contender for the source of this feedback, as they have been previously linked to the spatial integration of coherent scene information (*18*, *19*).

To test this assertion, we decoded the video stimuli from multivoxel patterns in a set of three early visual cortex regions (V1, V2, and V3), one motion-selective region [human middle temporal complex (hMT)/V5], and three scene-selective regions [the occipital place area (OPA), the medial place area (MPA), and the parahippocampal place area (PPA)].

In a first step, we decoded between the single video stimuli and found information in early visual cortex (V1, V2, and V3) and scene-selective cortex (OPA, MPA, and PPA) only when video stimuli were shown in the hemifield contralateral to the region investigated [all $t(35) > 3.75$, $P < 0.001$; Fig. 3B]. This implies that any stronger decoding for coherent, compared to incoherent, video stimuli can only be driven by the interaction of ipsilateral and contralateral inputs, rather than by the ipsilateral input alone. On this interpretative backdrop, we next decoded coherent and incoherent video stimuli. Both were decodable in each of the seven regions [all $t(35) > 4.43$, $P < 0.001$; Fig. 3C]. Critically, coherent video stimuli were only better decoded than incoherent stimuli in the MPA [$t(35) = 3.61$, $P < 0.001$; Fig. 3C] and PPA [$t(35) = 3.32$, $P = 0.002$; Fig. 3C]. A similar trend was found in hMT [$t(35) = 1.73$, $P = 0.092$; Fig. 3C]. In hMT, MPA, and PPA, coherent video stimuli were also better decoded than contralateral single video stimuli [all $t(35) > 2.99$, $P < 0.005$]. Similar results were found in the whole-brain searchlight-decoding analysis. We found significant decoding for single video stimuli across the visual cortex in the contralateral hemisphere (see fig. S6) as well as significant decoding across the visual cortex for both coherent (Fig. 3D) and incoherent stimuli (Fig. 3E). The differences between coherent and incoherent conditions were only found in locations overlapping—or close to—scene-selective cortex and hMT (Fig. 3F). Given the involvement of motion-selective hMT in integrating visual information, we also tested whether differences in motion coherence (operationalized as motion energy and motion direction) contribute to the integration effects observed here. When assessing differences between videos with high and low motion coherence across hemifields, however, we did not find qualitatively similar effects to our main analyses (see fig. S7), suggesting that motion coherence is not the main driver of the integration effects.

Together, these results show that scene-selective cortex and hMT aggregate spatiotemporally coherent information across hemifields, suggesting these regions as likely generators of feedback signals guiding visual integration.

Integration-related feedback traverses the visual hierarchy

Last, we determined where the feedback-related alpha rhythms are localized in brain space. We were particularly interested in whether integration-related feedback traverses the visual hierarchy up to the earliest stages of visual processing (*20*, *21*). To investigate this, we performed an EEG/fMRI fusion analysis (*22*, *23*) that directly links spectral representations in the EEG with spatial representations in the fMRI. To link representations across modalities, we first computed representational similarities between all video stimuli using pairwise decoding analyses and then correlated the similarities obtained from EEG alpha responses and fMRI activations across the seven visual regions (Fig. 4A). Here, we focused on the crucial comparison of regional representations (fMRI) and alpha-frequency representations (EEG) between the coherent and the incoherent conditions. Our fMRI decoding analyses for the single video stimuli demonstrate that V1 only receives sensory information from the contralateral visual field. As feedforward inputs from the contralateral visual field are identical across both conditions, any stronger correspondence between regional representations and alpha-frequency representations in the coherent condition can unequivocally be attributed to feedback from higher-order systems, which have access to both ipsi- and contralateral input. We found that representations in the alpha band were more strongly related to representations in the coherent condition

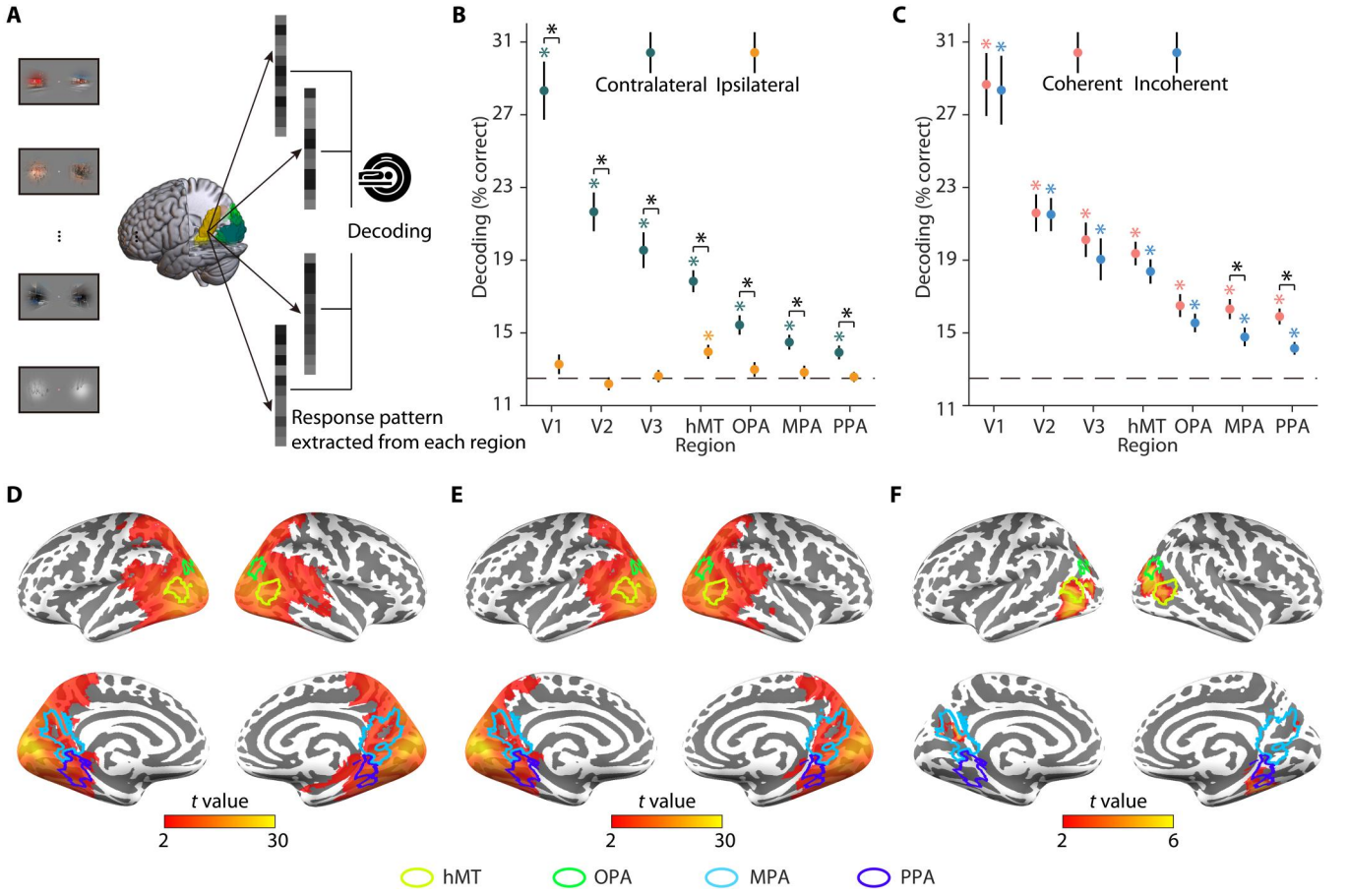

Fig. 3. fMRI decoding analysis. (**A**) fMRI decoding analysis in regions of interest (ROIs). In each condition, we used eight-way decoding to classify the video stimuli on response patterns in each ROI (V1, V2, V3, hMT, OPA, MPA, and PPA). (**B**) Results of fMRI ROI decoding analysis for the right- and left-only conditions. Single video stimuli were decodable in regions contralateral to the stimulation. In the ipsilateral hemisphere, they were only decodable in hMT but not in early visual cortex (V1, V2, and V3) and scene-selective cortex (OPA, MPA, and PPA). (**C**) Results of fMRI ROI decoding analysis for the coherent and incoherent conditions. Video stimuli were decodable in both conditions in each of the seven regions. Coherent stimuli were decoded better than incoherent stimuli in scene-selective cortex (MPA and PPA). (**D**) Results of fMRI searchlight decoding analysis for the coherent condition. Coherent stimuli were decodable across visual cortex. (**E**) Results of fMRI searchlight decoding analysis for the incoherent condition. Incoherent stimuli were decodable across visual cortex. (**F**) Significant differences between coherent and incoherent conditions in fMRI searchlight decoding analysis. Significant differences between the coherent and incoherent conditions were observed in locations overlapping—or close to—scene-selective cortex and hMT. Together, the results suggested that scene-selective cortex and hMT integrate dynamic information across visual hemifields. Error bars represent SEs. *$P < 0.05$ (FDR-corrected).

than in the incoherent condition in V1 [$t(35) = 3.37$, $P = 0.001$; Fig. 4B]. A similar trend emerged in V2 [$t(35) = 2.32$, $P_{uncorrected} = 0.025$; Fig. 4B] and V3 [$t(35) = 2.15$, $P_{uncorrected} = 0.036$; Fig. 4B] but not in hMT [$t(35) = -0.28$, $P = 0.783$; Fig. 4B] and scene-selective cortex [OPA: $t(35) = 0.94$, $P = 0.351$; MPA: $t(35) = 0.005$, $P = 0.996$; PPA: $t(35) = -1.64$, $P = 0.108$; Fig. 4B]. The correspondence between alpha-band representations in the EEG and activity in early visual cortex persisted after we controlled for motion coherence in the fusion analysis (see fig. S8), suggesting that the effect was not solely attributable to coherent patterns of motion. By contrast, no such correspondences were found between beta/gamma EEG responses and regional fMRI activations (see fig. S9). The results of the fusion analysis show that when inputs are spatiotemporally coherent and can be integrated into a unified percept, feedback-related alpha rhythms are found at the earliest stages of visual processing in cortex.

DISCUSSION

Our findings demonstrate that the spatial integration of naturalistic inputs integral to mediating coherent perception is achieved by cortical feedback: Only when spatiotemporally coherent inputs can be integrated into a coherent whole, stimulus-specific information was coded in feedback-related alpha activity. We further show that scene-selective cortex and hMT interactively process information across visual space, highlighting them as likely sources of integration-related feedback. Last, we reveal that integration-related alpha dynamics are linked to representations in early visual cortex, indicating that integration is accompanied by feedback that traverses the whole cortical visual hierarchy from top to bottom. Together, our results promote an active conceptualization of the visual system, where concurrent feedforward and feedback information flows are critical for establishing coherent naturalistic vision.

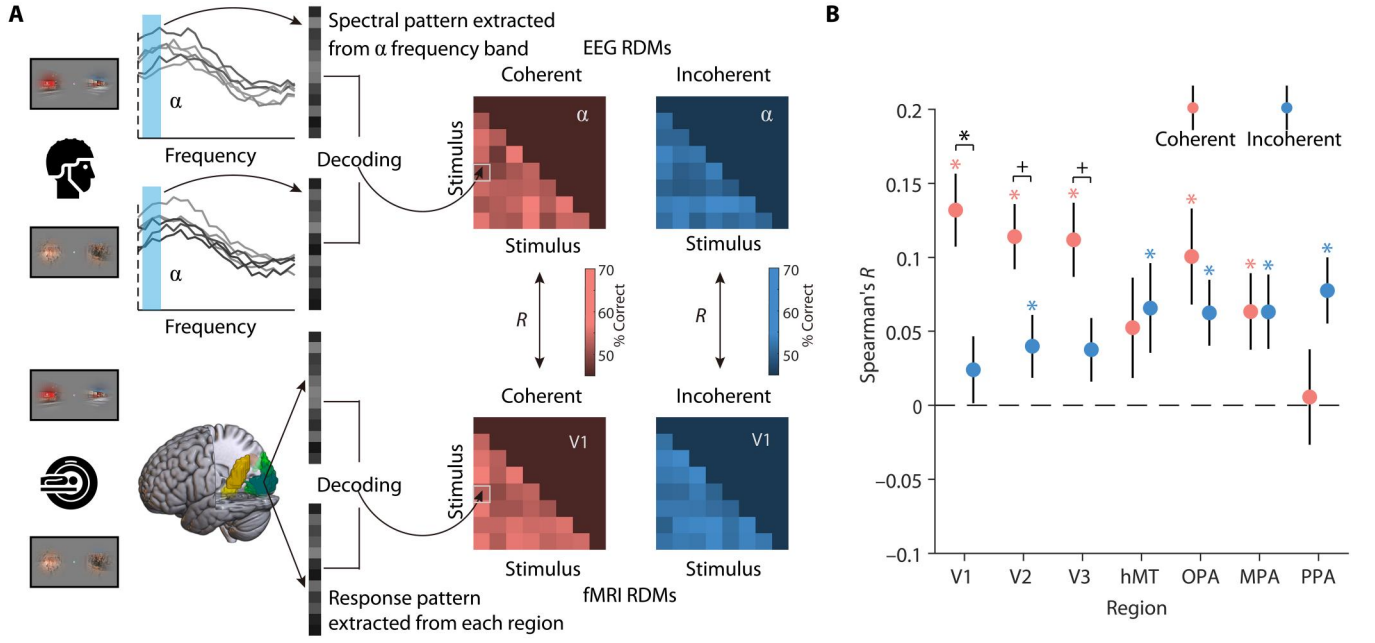

Fig. 4. EEG-fMRI fusion analysis. (**A**) For each condition, EEG representational dissimilarity matrices (RDMs) for each frequency band (alpha, beta, and gamma) and fMRI RDMs for each ROI (V1, V2, V3, hMT, OPA, MPA, and PPA) were first obtained using pairwise decoding analyses. To assess correspondences between spectral and regional representations, we calculated Spearman correlations between the participant-specific EEG RDMs in each frequency band and the group-averaged fMRI RDMs in each region, separately for each condition. (**B**) Results of EEG-fMRI fusion analysis in the alpha band. Representations in the alpha band corresponded more strongly with representations in V1 (with a similar trend in V2 and V3) when the videos were presented coherently, rather than incoherently. No correspondences were found between the beta and gamma bands and regional activity (see fig. S9). Error bars represent SEs. *P < 0.05 (FDR-corrected), +P < 0.05 (uncorrected).

Our finding that feedback reaches all the way to initial stages of visual processing supports the emerging notion that early visual cortex receives various types of stimulus-specific feedback, such as during mental imagery (*24*, *25*), in cross-modal perception (*26*, *27*), and during the interpolation of missing contextual information (*28*, *29*). Further supporting the interpretation of such signals as long-range feedback, recent animal studies have found that contextual signals in V1 are substantially delayed in time, compared to feedforward processing (*30*–*32*). Such feedback processes may use the spatial resolution of V1 as a flexible sketchpad mechanism (*33*, *34*) for recreating detailed feature mappings that are inferred from global context.

Our fMRI data identify scene-selective areas in the anterior ventral temporal cortex (the MPA and the PPA) and motion-selective area hMT as probable sources of the feedback to early visual cortex. These regions exhibited stronger representations for spatiotemporally coherent stimuli placed in the two hemifields. Scene-selective cortex is a logical candidate for a source of the feedback: Scene-selective regions are sensitive to the typical spatial configuration of scene stimuli (*18*, *19*, *35*, *36*), allowing them to create feedback signals that carry information about whether and how stimuli need to be integrated at lower levels of the visual hierarchy. These feedback signals may stem from adaptively comparing contralateral feedforward information with ipsilateral information from interhemispheric connections. In the incoherent condition, the ipsilateral information received from the other hemisphere does not match with typical real-world regularities and may thus not trigger integration. Conversely, when stimuli are coherent, interhemispheric transfer of information may be critical for facilitating integration across visual fields. This idea is consistent with previous studies showing increased interhemispheric connectivity when object or word information needs to be integrated across visual hemifields (*37*, *38*). Motion-selective hMT is also a conceivable candidate for integration-related feedback. The region not only showed enhanced representations for spatiotemporally coherent stimuli but also had representations for both contralateral and ipsilateral stimuli. The hMT's sensitivity to motion (*39*, *40*) and its bilateral visual representation (*41*) make it suited for integrating coherent motion patterns across hemifields. Although speculative at this point, scene-selective and motion-selective cortical areas may jointly generate adaptive feedback signals that combine information about coherent scene content (analyzed in MPA and PPA) and coherent motion patterns (hMT). Future studies need to map out cortico-cortical connectivity during spatial integration to test this idea.

Our results inform theories about the functional role of alpha rhythms in cortex. Alpha is often considered an idling rhythm (*42*, *43*), a neural correlate of active suppression (*44*–*46*), or a correlate of working memory maintenance (*47*, *48*). More recently, alpha rhythms were associated with an active role in cortical feedback processing (*12*, *14*, *49*, *50*). Our results highlight that alpha dynamics not only modulate feedforward processing but also encode stimulus-specific information. Our findings thus invite a different conceptualization of alpha dynamics, where alpha rhythms are critically involved in routing feedback-related information across the visual cortical hierarchy (*15*, *16*, *51*, *52*). An important remaining question is whether the feedback itself traverses in alpha rhythms or whether the feedback initiates upstream representations that

themselves fluctuate in alpha rhythms (*53*, *54*). The absence of a correspondence of alpha-band representations and regional activity in scene-selective cortex may suggest that it is not the feedback itself that is rhythmic, but alpha dynamics in scene-selective cortex may also be weaker—or to some extent initiated for both coherent and incoherent stimuli. More studies are needed to dissociate between the rhythmic nature of cortical feedback and the representations it instills in early visual cortex.

The increased involvement of alpha rhythms in coding the coherent visual stimuli was accompanied by an absence of concurrent representations in the gamma band. A potential reason for the absence of decoding from feedforward-related gamma activity in the coherent condition is that feedforward representations were efficiently suppressed by accurate top-down predictions (*5*). In our experiment, the video stimuli were presented for a relatively long time, without many rapid or unexpected visual events, potentially silencing feedforward propagation in the gamma range. We also did not find a correspondence between gamma dynamics and regional fMRI activity. Despite a general difficulty in linking high-frequency EEG activity and fMRI signals, another reason may be that, unlike the alpha dynamics, the gamma dynamics were relatively broadband and did not reflect a distinct neural rhythm (see fig. S2).

Our findings can be linked to theories of predictive processing that view neural information processing as a dynamic exchange of sensory feedforward signals and predictions stemming from higher-order areas of cortex (*5*, *6*, *51*). On this view, feedback signals arising during stimulus integration are conceptualized as predictions about sensory input derived from spatially and temporally coherent contralateral input. In our paradigm, feedback signals can be conceptualized as predictions for the contralateral input generated from the spatiotemporally coherent ipsilateral input. A challenge for predictive coding theories is that it requires a strict separation of the feedforward sensory input and the predictive feedback. Our results indicate a compelling solution through multiplexing of feedforward and feedback information in dedicated frequency-specific channels (*11*, *13*, *14*, *49*) in human cortex. It will be interesting to see whether similar frequency-specific correlates of predictive processing are unveiled in other brain systems in the future.

Our findings further pave the way toward researching integration processes under various task demands. In our experiments, we engaged participants in an unrelated fixation task, capitalizing on automatically triggered integration processes for spatiotemporally coherent stimuli. Such automatic integration is well in line with phenomenological experience: The coherent video stimuli, but not the incoherent stimuli, strongly appear as coherent visual events that happen behind an occlude. Future studies should investigate how integration effects vary when participants are required to engage with the stimuli that need to be integrated. It will be particularly interesting to see whether tasks that require more global or local analysis are related to different degrees of integration and different rhythmic codes in the brain. These future studies could also set out to determine the critical features that enable integration and what is integrated. As our incoherent stimuli were designed to be incoherent along many different dimensions (e.g., low- and mid-level visual features, categorical content, and motion patterns), a comprehensive mapping of how these dimensions independently contribute to integration is needed. Future studies could thereby delineate how integration phenomenologically depends on the coherence of these candidate visual features.

More generally, our results have general implications for understanding and modeling feedforward and feedback information flows in neural systems. Processes like stimulus integration that are classically conceptualized as solvable in pure feedforward cascades may be more dynamic than previously thought. Discoveries like ours arbitrate between competing theories that either stress the power of feedforward hierarchies (*3*, *4*) or emphasize the critical role of feedback processes (*5*, *6*). They further motivate approaches to computational modeling that capture the visual system's abundant feedback connectivity (*55*, *56*).

Together, our results reveal feedback in the alpha-frequency range across the visual hierarchy as a key mechanism for integrating spatiotemporally coherent information in naturalistic vision. This strongly supports an active conceptualization of the visual system, where top-down projections are critical for the construction of coherent and unified visual percepts from fragmented sensory information.

MATERIALS AND METHODS

Participants

Forty-eight healthy adults (gender: 12 males/36 females, age: 21.1 ± 3.8 years) participated in the EEG experiment, and another 36 (gender: 17 males/19 females, age: 27.3 ± 2.5 years) participated in the fMRI experiment. Sample size resulted from convenience sampling, with the goal of exceeding $n = 34$ in both experiments (i.e., exceeding 80% power for detecting a medium effect size of $d = 0.5$ in a two-sided t test). All participants had normal or corrected-to-normal vision and had no history of neurological/psychiatric disorders. They all signed informed consent before the experiment, and they were compensated for their time with partial course credit or cash. The EEG protocol was approved by the ethical committee of the Department of Psychology at the University of York, and the fMRI protocol was approved by the ethical committee of the Department of Psychology at Freie Universität Berlin. All experimental protocols were in accordance with the Declaration of Helsinki.

Stimuli and design

The stimuli and design were identical for the EEG and fMRI experiments unless stated otherwise. Eight short video clips (3 s each; Fig. 1A) depicting everyday situations (e.g., a train driving past the observer; a view of red mountains; a view of waves crashing on the coast; first-person perspective of walking along a forest path; an aerial view of a motorway toll station; a group of zebras grazing on a prairie; first-person perspective of skiing in the forest; and first-person perspective of walking along a street) were used in the experiments. During the experiment, these original videos were presented on the screen through circular apertures right and left of central fixation. We manipulated four experimental conditions (right-only, left-only, coherent, and incoherent) by presenting the original videos in different ways (Fig. 1B). In the right- or left-only condition, we presented the videos through right or left aperture only. We also showed two matching segments from the same video in the coherent condition, while we showed segments from two different videos in the incoherent condition, through both apertures. In the incoherent condition, the eight original videos were yoked into eight fixed pairs (see fig. S1), and each video was always only shown with its paired video. Thus, there

were a total of 32 unique video stimuli (8 for each condition). The diameter of each aperture was 6° visual angle, and the shortest distance between the stimulation and central fixation was 2.64° visual angle. The borders of the apertures were slightly smoothed. The central fixation dot subtended 0.44° visual angle. We selected our videos to be diverse in content (e.g., natural versus manmade) and motion (e.g., natural versus camera motion). This was done to maximize the contrast between the coherent and incoherent video stimuli. For assessing motion, we quantified the motion energy for each video stimulus using Motion Energy Analysis software (https://psync.ch/mea/). We did not find any significant between-condition differences [comparison on the means of motion energy: right- versus left-only, $t(7) = 0.36$, $P = 0.728$, coherent versus incoherent, $t(7) = 0.03$, $P = 0.976$; comparison on the SDs of motion energy: right- versus left-only, $t(7) = 1.08$, $P = 0.316$, coherent versus incoherent, $t(7) = 1.13$, $P = 0.294$]. Although this suggests that there was no difference in overall motion between conditions, there are many other candidates for critical differences between conditions, which will have to be evaluated in future studies.

The experiments were controlled through MATLAB and the Psychophysics Toolbox (*57*, *58*). In each trial, a fixation dot was first shown for 500 ms, after which a unique video stimulus was displayed for 3000 ms. During the video stimulus playback, the color of the fixation changed periodically (every 200 ms) and turned either green or yellow at a single random point in the sequence (but never the first or last point). After every trial, a response screen prompted participants to report whether a green or yellow fixation dot was included in the sequence. Participants were instructed to keep central fixation during the sequence so they would be able to solve this task accurately. In both experiments, participants performed the color discrimination with high accuracy (EEG: 93.28 ± 1.65% correct; fMRI: 91.44 ± 1.37% correct), indicating that they indeed focused their attention on the central task. There were no significant differences in behavioral accuracy and response time (RT) between the coherent and incoherent conditions in both the EEG and fMRI experiments [accuracy-EEG, $t(47) = 1.07$, $P = 0.29$; accuracy-fMRI, $t(35) = 0.20$, $P = 0.85$; RT-fMRI, $t(35) = 0.41$, $P = 0.69$]. Note that RTs were not recorded in the EEG experiment. The mean accuracy and RT for each condition in both experiments are listed in table S1. In the EEG experiment, the next trial started once the participant′s response was received. In the fMRI experiment, the response screen stayed on the screen for 1500 ms, irrespective of participants′ RT. An example trial is shown in Fig. 1C.

In the EEG experiment, each of the 32 unique stimuli was presented 20 times, resulting in a total of 640 trials, which were presented in random order. In the fMRI experiment, participants performed 10 identical runs. In each run, each unique stimulus was presented twice, in random order. Across the 10 runs, this also resulted in a total of 640 trials. The extensive repetition of the incoherent combinations may lead to some learning of the inconsistent stimuli in our experiment (*59*). However, such learning would, if anything, lead to an underestimation of the effects: Learning of the incoherent combination would ultimately also lead to an integration of the incoherent stimuli and thus create similar—albeit weaker—neural signatures of integration as found in the coherent condition.

To make sure that our intuition about the coherence of video stimuli in the coherent and incoherent conditions is valid, we conducted an additional behavioral experiment on 10 participants (gender: 4 males/6 females, age: 24.8 ± 2.5 years). In the experiment, we presented each of the coherent and incoherent video stimuli once. After each trial, we asked the participants to rate the degree of unified perception of the stimulus on a 1 to 5 scale. We found that the rating of coherent stimuli was higher than the rating of incoherent stimuli (mean ratings for coherent stimuli: 4.1 to 4.6, mean ratings for incoherent stimuli: 1.2 to 1.6; $t(9) = 36.66$, $P < 0.001$), showing that the coherent stimuli were indeed rated as more coherent than the incoherent ones.

To assess the general fixation stability for our paradigm, we collected additional eye-tracking data from six new participants (see fig. S10 for details). We calculated the mean and SD of the horizontal and vertical eye movement across time (0 to 3 s) in each trial and then averaged the mean and SD values across trials separately for each condition. For all participants, we found means of eye movement lower than 0.3°, and SDs of eye movement lower than 0.2°, indicating stable central fixation (see fig. S10A). In addition, participants did not disengage from fixating after the target color was presented (see fig. S10B).

EEG recording and preprocessing

EEG signals were recorded using an ANT waveguard 64-channel system and a TSMi REFA amplifier, with a sample rate of 1000 Hz. The electrodes were arranged according to the standard 10-10 system. EEG data preprocessing was performed using FieldTrip (*60*). The data were first band-stop filtered to remove 50-Hz line noise and then band-pass filtered between 1 and 100 Hz. The filtered data were epoched from −500 to 4000 ms relative to the onset of the stimulus, re-referenced to the average over the entire head, downsampled to 250 Hz, and baseline corrected by subtracting the mean prestimulus signal for each trial. After that, noisy channels and trials were removed by visual inspection, and the removed channels (2.71 ± 0.19 channels) were interpolated by the mean signals of their neighboring channels. Blinks and eye movement artifacts were removed using independent component analysis and visual inspection of the resulting components.

EEG power spectrum analysis

Spectral analysis was performed using FieldTrip. Power spectra were estimated between 8 and 70 Hz (from alpha to gamma range), from 0 to 3000 ms (i.e., the period of stimulus presentation) on the preprocessed EEG data, separately for each trial and each channel. A single taper with a Hanning window was used for the alpha band (8 to 12 Hz, in steps of 1 Hz) and the beta band (13 to 30 Hz, in steps of 2 Hz), and the discrete prolate spheroidal sequences multitaper method with ±8 Hz smoothing was used for the gamma band (31 to 70 Hz, in steps of 2 Hz).

EEG decoding analysis

To investigate whether the dynamic integration of information across the visual field is mediated by oscillatory activity, we performed multivariate decoding analysis using CoSMoMVPA (*61*) and the Library for Support Vector Machines (LIBSVM) (*62*). In this analysis, we decoded between the eight video stimuli using patterns of spectral power across channels, separately for each frequency band (alpha, beta, and gamma) and each condition. Specifically, for each frequency band, we extracted the power of the frequencies included in that band (e.g., 8 to 12 Hz for the alpha band) across all

channels from the power spectra and then used the resulting patterns across channels and frequencies to classify the eight video stimuli in each condition. For all classifications, we used linear support vector machine (SVM) classifiers to discriminate the eight stimuli in a 10-fold cross-validation scheme. For each classification, the data were allocated to 10 folds randomly, and then an SVM classifier was trained on data from 9 folds and tested on data from the left-out fold. The classification was done repeatedly until every fold was left out once, and accuracies were averaged across these repetitions. The amount of data in the training set was always balanced across stimuli. For each classification, a maximum of 144 trials (some trials were removed during preprocessing) were included in the training set (18 trials for each stimulus) and 16 trials were used for testing (2 trials for each stimulus). Before classification, principal components analysis (PCA) was applied to reduce the dimensionality of the data (*63*). Specifically, for each classification, PCA was performed on the training data, and the PCA solution was projected onto the testing data. For each PCA, we selected the set of components that explained 99% of the variance of the training data. As a result, we obtained decoding accuracies for each frequency band and each condition, which indicated how well the video stimuli were represented in frequency-specific neural activity. We first used a one-sample *t* test to investigate whether the video stimuli could be decoded in each condition and each frequency band. We also performed a 2-condition (coherent and incoherent) × 3-frequency (alpha, beta, and gamma) two-way analysis of variance (ANOVA) and post hoc paired *t* tests [false discovery rate (FDR)–corrected across frequencies; $P_{corrected} < 0.05$] to compare the decoding differences between coherent and incoherent conditions separately for each frequency band. The comparisons of right- and left-only conditions were conducted using the same approaches. To track where the effects appeared across a continuous frequency space, we also decoded between the eight stimuli at each frequency from 8 to 70 Hz using a sliding window approach with a five-frequency resolution (see fig. S2).

fMRI recording and processing

MRI data were acquired using a 3T Siemens Prisma scanner (Siemens, Erlangen, Germany) equipped with a 64-channel head coil. T2*-weighted BOLD images were obtained using a multiband gradient-echo echo-planar imaging (EPI) sequence with the following parameters: multiband factor = 3, repetition time (TR) = 1500 ms, echo time (TE) = 33 ms, field of view = 204 mm by 204 mm, voxel size = 2.5 mm by 2.5 mm by 2.5 mm, 70° flip angle, 57 slices, and 10% interslice gap. Field maps were also obtained with a double-echo gradient echo field map sequence (TR = 545 ms, TE1/TE2 = 4.92 ms/7.38 ms) to correct for distortion in EPI. In addition, a high-resolution 3D T1-weighted image was collected for each participant (magnetization-prepared rapid gradient-echo, TR = 1900 ms, TE = 2.52 ms, TI = 900 ms, 256 × 256 matrix, 1-mm by 1-mm by 1-mm voxel, 176 slices).

MRI data were preprocessed using MATLAB and SPM12 (www.fil.ion.ucl.ac.uk/spm/). Functional data were first corrected for geometric distortion with the SPM FieldMap toolbox (*64*) and realigned for motion correction. In addition, individual participants′ structural images were coregistered to the mean realigned functional image, and transformation parameters to Montreal Neurological Institute (MNI) standard space (as well as inverse transformation parameters) were estimated.

The GLMsingle Toolbox (*65*) was used to estimate the fMRI responses to the stimulus in each trial based on realigned fMRI data. To improve the accuracy of trialwise beta estimations, a three-stage procedure was used, including identifying an optimal hemodynamic response function (HRF) for each voxel from a library of 20 HRFs, denoising data-driven nuisance components identified by cross-validated PCA, and applying fractional ridge regression to regularize the beta estimation on a single-voxel basis. The resulting single-trial betas were used for further decoding analyses.

fMRI regions of interest definition

fMRI analyses were focused on seven regions of interest (ROIs). We defined three scene-selective areas—OPA [also termed transverse occipital sulcus (*66*, *67*)], MPA [also termed retrosplenial cortex (*68*, *69*)], and PPA (*70*)—from a group functional atlas (*71*) and three early visual areas—V1, V2, and V3, as well as motion-selective hMT/V5—from a probabilistic functional atlas (*72*). All ROIs were defined in MNI space and separately for each hemisphere and then transformed into individual-participant space using the inverse normalization parameters estimated during preprocessing.

fMRI ROI decoding analysis

To investigate how the video stimuli were processed in different visual regions, we performed multivariate decoding analysis using CoSMoMVPA and LIBSVM. For each ROI, we used the beta values across all voxels included in the region to decode between the eight video stimuli, separately for each condition. Leave-one-run-out cross-validation and PCA were used to conduct SVM classifications. For each classification, there were 144 trials (18 for each stimulus) in the training set and 16 trials (2 for each stimulus) in the testing set. For each participant, we obtained a 4-condition × 14-ROI (7 ROIs by two hemispheres) decoding matrix. Results were averaged across hemispheres, as we consistently found no significant interhemispheric differences (condition × hemisphere and condition × region × hemisphere interaction effects) in a 2-condition (coherent and incoherent) × 7-region (V1, V2, V3, hMT, OPA, MPA, and PPA) × 2-hemisphere (left and right) three-way ANOVA test. We first tested whether the video stimuli were decodable in each condition and each region using one-sample *t* tests (FDR-corrected across regions; $P_{corrected} < 0.05$). To further investigate the integration effect, we used paired *t* tests to compare the decoding difference between coherent and incoherent conditions in different regions. For the right- and left-only conditions, we averaged the decoding results in a contralateral versus ipsilateral fashion (e.g., left stimulus, right brain region was averaged with right stimulus, left brain region to obtain the contralateral decoding performance).

fMRI searchlight decoding analysis

To further investigate the whole-brain representation of video stimuli, we performed searchlight decoding analyses using CoSMoMVPA and LIBSVM. The single-trial beta maps in the native space were first transformed into the MNI space using the normalization parameters estimated during preprocessing. For the searchlight analysis, we defined a sphere with a radius of five voxels around a given voxel and then used the beta values of the voxels within this sphere to classify the eight video stimuli in each condition. For the left- and right-only conditions, the decoding analysis was performed separately for each hemisphere. The decoding parameters were identical to the ROI-decoding analysis. The resulting

searchlight maps were subsequently smoothed with a Gaussian kernel (full width at half maximum = 6 mm). To investigate how well the video stimuli in each condition were represented across the whole brain, we used one-sample *t* tests to compare decoding accuracies against chance separately for each condition [Gaussian random field (GRF) correction, voxel-level $P < 0.005$, cluster-extent $P < 0.05$]. To investigate the integration effect in the whole brain, we used paired *t* tests to compare the differences in decoding accuracy between coherent and incoherent conditions and performed multiple comparisons correction within the voxels showing significant decoding for either coherent or incoherent stimuli (GRF correction, voxel-level $P < 0.005$, cluster-extent $P < 0.05$).

EEG-fMRI fusion with representational similarity analysis

To investigate the relationship between the frequency-specific effects obtained in the EEG and the spatial mapping obtained in the fMRI, we performed EEG-fMRI fusion analysis (*22*, *23*). This analysis can be used to compare neural representations of stimuli as characterized by EEG and fMRI data to reveal how the representations correspond across brain space and spectral signatures. Specifically, we first calculated representational dissimilarity matrices (RDMs) using pairwise decoding analysis for EEG and fMRI data, respectively. For the EEG power spectra, in each frequency band, we decoded between each pair of eight video stimuli using the oscillatory power of the frequencies included in the frequency band, separately for each condition; for the fMRI data, in each ROI, we classified each pair of eight stimuli using the response patterns of the region, separately for each condition. Decoding parameters were otherwise identical to the eight-way decoding analyses (see above). In each condition, we obtained a participant-specific EEG RDM (8 stimuli × 8 stimuli) in each frequency band and a participant-specific fMRI RDM (8 stimuli × 8 stimuli) in each ROI. Next, we calculated the similarity between EEG and fMRI RDMs for each condition; this was done by correlating all lower off-diagonal entries between the EEG and fMRI RDMs (the diagonal was always left out). To increase the signal-to-noise ratio, we first averaged fMRI RDMs across participants and then calculated the Spearman correlation between the averaged fMRI RDM for each ROI with the participant-specific EEG RDM for each frequency. As a result, we obtained a 4-condition × 3-frequency × 14-ROI fusion matrix for each EEG participant. For the coherent and incoherent conditions, the results were averaged across hemispheres, as no condition × hemisphere, no condition × region × hemisphere, and no condition × frequency × hemisphere interaction effects were found in a 2-condition (coherent and incoherent) × 7-region (V1, V2, V3, hMT, OPA, MPA, and PPA) × 2-hemisphere (left and right) × 3-frequency (alpha, beta, and gamma) four-way ANOVA test. We first used one-sample *t* tests to test the fusion effect in each condition (FDR-corrected across regions; $P_{corrected} < 0.05$) and each frequency-region combination and then used a 2-condition × 3-frequency × 7-region three-way ANOVA to compare the frequency-region correspondence between coherent and incoherent conditions. As we found a significant condition × frequency × region interaction effect, we further performed a 2-condition × 7-region ANOVA and paired *t* tests (FDR-corrected across regions; $P_{corrected} < 0.05$) to compare frequency-region correspondence between coherent and incoherent conditions separately for each frequency. For the right- and left-only conditions, we averaged the fusion results across two conditions separately for contralateral and ipsilateral presentations and then compared contralateral and ipsilateral presentations using the same approaches we used for the comparisons of coherent and incoherent conditions (see fig. S9).

Supplementary Materials

This PDF file includes:

Figs. S1 to S10

Table S1

REFERENCES AND NOTES

1. N. Block, Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* **30**, 481–499 (2007).
2. M. A. Cohen, D. C. Dennett, N. Kanwisher, What is the bandwidth of perceptual experience? *Trends Cogn. Sci.* **20**, 324–335 (2016).
3. M. Riesenhuber, T. Poggio, Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**, 1019–1025 (1999).
4. J. J. DiCarlo, D. D. Cox, Untangling invariant object recognition. *Trends Cogn. Sci.* **11**, 333–341 (2007).
5. R. P. N. Rao, D. H. Ballard, Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
6. K. Friston, A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**, 815–836 (2005).
7. A. M. Bastos, W. M. Usrey, R. A. Adams, G. R. Mangun, P. Fries, K. J. Friston, Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711 (2012).
8. P. A. Salin, J. Bullier, Corticocortical connections in the visual system: Structure and function. *Physiol. Rev.* **75**, 107–154 (1995).
9. V. A. Lamme, H. Supèr, H. Spekreijse, Feedforward, horizontal, and feedback processing in the visual cortex. *Curr. Opin. Neurobiol.* **8**, 529–535 (1998).
10. N. T. Markov, M. Ercsey-Ravasz, D. C. Van Essen, K. Knoblauch, Z. Toroczkai, H. Kennedy, Cortical high-density counterstream architectures. *Science* **342**, 1238406 (2013).
11. T. van Kerkoerle, M. W. Self, B. Dagnino, M.-A. Gariel-Mathis, J. Poort, C. van der Togt, P. R. Roelfsema, Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14332–14341 (2014).
12. A. M. Bastos, J. Vezoli, C. A. Bosman, J.-M. Schoffelen, R. Oostenveld, J. R. Dowdall, P. De Weerd, H. Kennedy, P. Fries, Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* **85**, 390–401 (2015).
13. P. Fries, Rhythms for cognition: Communication through coherence. *Neuron* **88**, 220–235 (2015).
14. G. Michalareas, J. Vezoli, S. van Pelt, J.-M. Schoffelen, H. Kennedy, P. Fries, Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* **89**, 384–397 (2016).
15. S. Xie, D. Kaiser, R. M. Cichy, Visual imagery and perception share neural representations in the alpha frequency band. *Curr. Biol.* **30**, 2621–2627.e5 (2020).
16. D. Kaiser, Spectral brain signatures of aesthetic natural perception in the α and β frequency bands. *J. Neurophysiol.* **128**, 1501–1505 (2022).
17. J. D. Haynes, A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives. *Neuron* **87**, 257–270 (2015).
18. D. J. Mannion, D. J. Kersten, C. A. Olman, Regions of mid-level human visual cortex sensitive to the global coherence of local image patches. *J. Cogn. Neurosci.* **26**, 1764–1774 (2014).
19. D. Kaiser, R. M. Cichy, Parts and wholes in scene processing. *J. Cogn. Neurosci.* **34**, 4–15 (2021).
20. S. Clavagnier, A. Falchier, H. Kennedy, Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. *Cogn. Affect. Behav. Neurosci.* **4**, 117–126 (2004).
21. L. Muckli, L. S. Petro, Network interactions: Non-geniculate input to V1. *Curr. Opin. Neurobiol.* **23**, 195–201 (2013).
22. R. M. Cichy, D. Pantazis, A. Oliva, Resolving human object recognition in space and time. *Nat. Neurosci.* **17**, 455–462 (2014).
23. R. M. Cichy, A. Oliva, A M/EEG-fMRI fusion primer: Resolving human brain responses in space and time. *Neuron* **107**, 772–781 (2020).
24. C. I. P. Winlove, F. Milton, J. Ranson, J. Fulford, M. MacKisack, F. Macpherson, A. Zeman, The neural correlates of visual imagery: A co-ordinate-based meta-analysis. *Cortex* **105**, 4–25 (2018).

25. F. Ragni, A. Lingnau, L. Turella, Decoding category and familiarity information during visual imagery. *Neuroimage* **241**, 118428 (2021).
26. P. Vetter, F. W. Smith, L. Muckli, Decoding sound and imagery content in early visual cortex. *Curr. Biol.* **24**, 1256–1262 (2014).
27. P. Vetter, Ł. Bola, L. Reich, M. Bennett, L. Muckli, A. Amedi, Decoding natural sounds in early "visual" cortex of congenitally blind individuals. *Curr. Biol.* **30**, 3039–3044.e2 (2020).
28. F. W. Smith, L. Muckli, Nonstimulated early visual areas carry information about surrounding context. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20099–20103 (2010).
29. L. Muckli, F. De Martino, L. Vizioli, L. S. Petro, F. W. Smith, K. Ugurbil, R. Goebel, E. Yacoub, Contextual feedback to superficial layers of V1. *Curr. Biol.* **25**, 2690–2695 (2015).
30. A. J. Keller, M. M. Roth, M. Scanziani, Feedback generates a second receptive field in neurons of the visual cortex. *Nature* **582**, 545–549 (2020).
31. L. Kirchberger, S. Mukherjee, M. W. Self, P. R. Roelfsema, Contextual drive of neuronal responses in mouse V1 in the absence of feedforward input. *Sci. Adv.* **9**, eadd2498 (2023).
32. P. Papale, F. Wang, A. T. Morgan, X. Chen, A. Gilhuis, L. S. Petro, L. Muckli, P. R. Roelfsema, M. W. Self, Feedback brings scene information to the representation of occluded image regions in area V1 of monkeys and humans. bioRxiv 2022.11.21.517305 [**Preprint**]. 22 November 2022. https://doi.org/10.1101/2022.11.21.517305.
33. S. Dehaene, L. Cohen, Cultural recycling of cortical maps. *Neuron* **56**, 384–398 (2007).
34. M. A. Williams, C. I. Baker, H. P. Op de Beeck, W. Mok Shim, S. Dang, C. Triantafyllou, N. Kanwisher, Feedback of visual object information to foveal retinotopic cortex. *Nat. Neurosci.* **11**, 1439–1445 (2008).
35. M. Bilalić, T. Lindig, L. Turella, Parsing rooms: The role of the PPA and RSC in perceiving object relations and spatial layout. *Brain Struct. Funct.* **224**, 2505–2524 (2019).
36. D. Kaiser, G. Häberle, R. M. Cichy, Cortical sensitivity to natural scene structure. *Hum. Brain Mapp.* **41**, 1286–1295 (2020).
37. K. E. Stephan, J. C. Marshall, W. D. Penny, K. J. Friston, G. R. Fink, Interhemispheric integration of visual processing during task-driven lateralization. *J. Neurosci.* **27**, 3512–3522 (2007).
38. T. Mima, T. Oluwatimilehin, T. Hiraoka, M. Hallett, Transient interhemispheric neuronal synchrony correlates with object recognition. *J. Neurosci.* **21**, 3942–3948 (2001).
39. J. D. Watson, R. Myers, R. S. Frackowiak, J. V. Hajnal, R. P. Woods, J. C. Mazziotta, S. Shipp, S. Zeki, Area V5 of the human brain: Evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cereb. Cortex* **3**, 79–94 (1993).
40. R. B. Tootell, J. B. Reppas, K. K. Kwong, R. Malach, R. T. Born, T. J. Brady, B. R. Rosen, J. W. Belliveau, Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* **15**, 3215–3230 (1995).
41. D. Cohen, E. Goddard, K. T. Mullen, Reevaluating hMT+ and hV4 functional specialization for motion and static contrast using fMRI-guided repetitive transcranial magnetic stimulation. *J. Vis.* **19**, 11 (2019).
42. G. Pfurtscheller, A. Stancák, C. Neuper, Event-related synchronization (ERS) in the alpha band — an electrophysiological correlate of cortical idling: A review. *Int. J. Psychophysiol.* **24**, 39–46 (1996).
43. V. Romei, V. Brodbeck, C. Michel, A. Amedi, A. Pascual-Leone, G. Thut, Spontaneous fluctuations in posterior α-band EEG activity reflect variability in excitability of human visual areas. *Cereb. Cortex* **18**, 2010–2018 (2008).
44. O. Jensen, A. Mazaheri, Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Front. Hum. Neurosci.* **4**, 186 (2010).
45. S. Haegens, V. Nácher, R. Luna, R. Romo, O. Jensen, α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19377–19382 (2011).
46. M. S. Clayton, N. Yeung, R. Cohen Kadosh, The roles of cortical oscillations in sustained attention. *Trends Cogn. Sci.* **19**, 188–195 (2015).
47. D. Jokisch, O. Jensen, Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream. *J. Neurosci.* **27**, 3244–3251 (2007).
48. I. E. J. de Vries, H. A. Slagter, C. N. L. Olivers, Oscillatory control over representational states in working memory. *Trends Cogn. Sci.* **24**, 150–162 (2020).
49. A. M. Bastos, M. Lundqvist, A. S. Waite, N. Kopell, E. K. Miller, Layer and rhythm specificity for predictive routing. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 31459–31469 (2020).
50. M. S. Clayton, N. Yeung, R. Cohen Kadosh, The many characters of visual alpha oscillations. *Eur. J. Neurosci.* **48**, 2498–2508 (2018).
51. A. Clark, Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).
52. Y. Hu, Q. Yu, Spatiotemporal dynamics of self-generated imagery reveal a reverse cortical hierarchy from cue-induced imagery. bioRxiv 2023.01.25.525474 [**Preprint**]. 25 January 2023. https://doi.org/10.1101/2023.01.25.525474.
53. A. Alamia, R. VanRullen, Alpha oscillations and traveling waves: Signatures of predictive coding? *PLOS Biol.* **17**, e3000487 (2019).
54. D. Lozano-Soldevilla, R. VanRullen, The hidden spatial dimension of alpha: 10-Hz perceptual echoes propagate as periodic traveling waves in the human brain. *Cell Rep.* **26**, 374–380.e4 (2019).
55. G. Kreiman, T. Serre, Beyond the feedforward sweep: Feedback computations in the visual cortex. *Ann. N. Y. Acad. Sci.* **1464**, 222–241 (2020).
56. G. W. Lindsay, Convolutional neural networks as a model of the visual system: Past, present, and future. *J. Cogn. Neurosci.* **33**, 2017–2031 (2021).
57. D. H. Brainard, The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
58. D. G. Pelli, The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
59. H. E. M. den Ouden, J. Daunizeau, J. Roiser, K. J. Friston, K. E. Stephan, Striatal prediction error modulates cortical coupling. *J. Neurosci.* **30**, 3210–3219 (2010).
60. R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
61. N. N. Oosterhof, A. C. Connolly, J. V. Haxby, CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Front. Neuroinform.* **10**, 27 (2016).
62. C.-C. Chang, C.-J. Lin, LIBSVM. *ACM Trans. Intell. Syst. Technol.* **2**, 1–27 (2011).
63. L. Chen, R. M. Cichy, D. Kaiser, Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations. *Cereb. Cortex* **32**, 3553–3567 (2022).
64. C. Hutton, A. Bork, O. Josephs, R. Deichmann, J. Ashburner, R. Turner, Image distortion correction in fMRI: A quantitative evaluation. *Neuroimage* **16**, 217–240 (2002).
65. J. S. Prince, I. Charest, J. W. Kurzawski, J. A. Pyles, M. J. Tarr, K. N. Kay, Improving the accuracy of single-trial fMRI response estimates using GLMsingle. *eLife* **11**, e77599 (2022).
66. K. Grill-Spector, The neural basis of object perception. *Curr. Opin. Neurobiol.* **13**, 159–166 (2003).
67. D. D. Dilks, J. B. Julian, A. M. Paunov, N. Kanwisher, The occipital place area is causally and selectively involved in scene perception. *J. Neurosci.* **33**, 1331–1336 (2013).
68. E. Maguire, The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scand. J. Psychol.* **42**, 225–238 (2001).
69. R. A. Epstein, C. I. Baker, Scene Perception in the Human Brain. *Annu. Rev. Vis. Sci.* **5**, 373–397 (2019).
70. R. Epstein, A. Harris, D. Stanley, N. Kanwisher, The parahippocampal place area. *Neuron* **23**, 115–125 (1999).
71. J. B. Julian, E. Fedorenko, J. Webster, N. Kanwisher, An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* **60**, 2357–2364 (2012).
72. M. Rosenke, R. van Hoof, J. van den Hurk, K. Grill-Spector, R. Goebel, A probabilistic functional atlas of human occipito-temporal visual cortex. *Cereb. Cortex* **31**, 603–619 (2021).

Acknowledgments: We thank D. Marinova and A. Carter for help in EEG data acquisition. We would also thank the HPC Service of ZEDAT, Freie Universität Berlin, for computing time.
Funding: L.C. is supported by a PhD stipend from the China Scholarship Council (CSC). R.M.C. is supported by the Deutsche Forschungsgemeinschaft (DFG; CI241/1-1, CI241/3-1, and CI241/7-1) and by a European Research Council (ERC) starting grant (ERC-2018-STG 803370). D.K. is supported by the DFG (SFB/TRR135 – INST162/567-1, project number 222641018), an ERC starting grant (PEP, ERC-2022-STG 101076057), and "The Adaptive Mind" funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. **Author contributions:** Conceptualization: L.C. and D.K. Methodology: L.C. and D.K. Software: L.C. and D.K. Formal analysis: L.C. Investigation: L.C. and D.K. Resources: L.C. and D.K. Data curation: L.C. Writing—original draft: L.C. and D.K. Writing—review and editing: L.C., R.M.C., and D.K. Visualization: L.C. Supervision: R.M.C. and D.K. Project administration: R.M.C. and D.K. Funding acquisition: R.M.C. and D.K. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Raw data used in the analyses are available at the following Zenodo repository: https://doi.org/10.5281/zenodo.8369131. Processed data and code used in the analyses are available at the following Zenodo repository: https://doi.org/10.5281/zenodo.8369136.

Submitted 12 April 2023
Accepted 12 October 2023
Published 10 November 2023
10.1126/sciadv.adi2321

ScienceAdvances

Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision

Lixiang Chen, Radoslaw M. Cichy, and Daniel Kaiser

Sci. Adv. **9** (45), eadi2321. DOI: 10.1126/sciadv.adi2321

Science Advances
AAAS

Supplementary Materials for

Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision

Lixiang Chen *et al*.

Corresponding author: Lixiang Chen, lixiang.chen@fu-berlin.de; Daniel Kaiser, danielkaiser.net@gmail.com

Sci. Adv. **9**, eadi2321 (2023)
DOI: 10.1126/sciadv.adi2321

This PDF file includes:

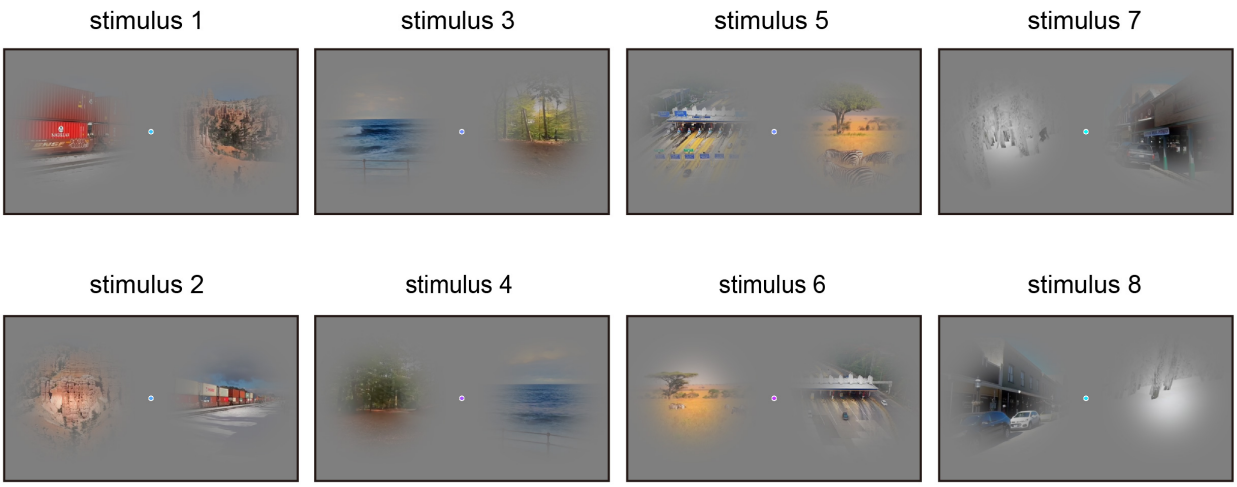

Fig. S1.
Still images from the incoherent video stimuli.

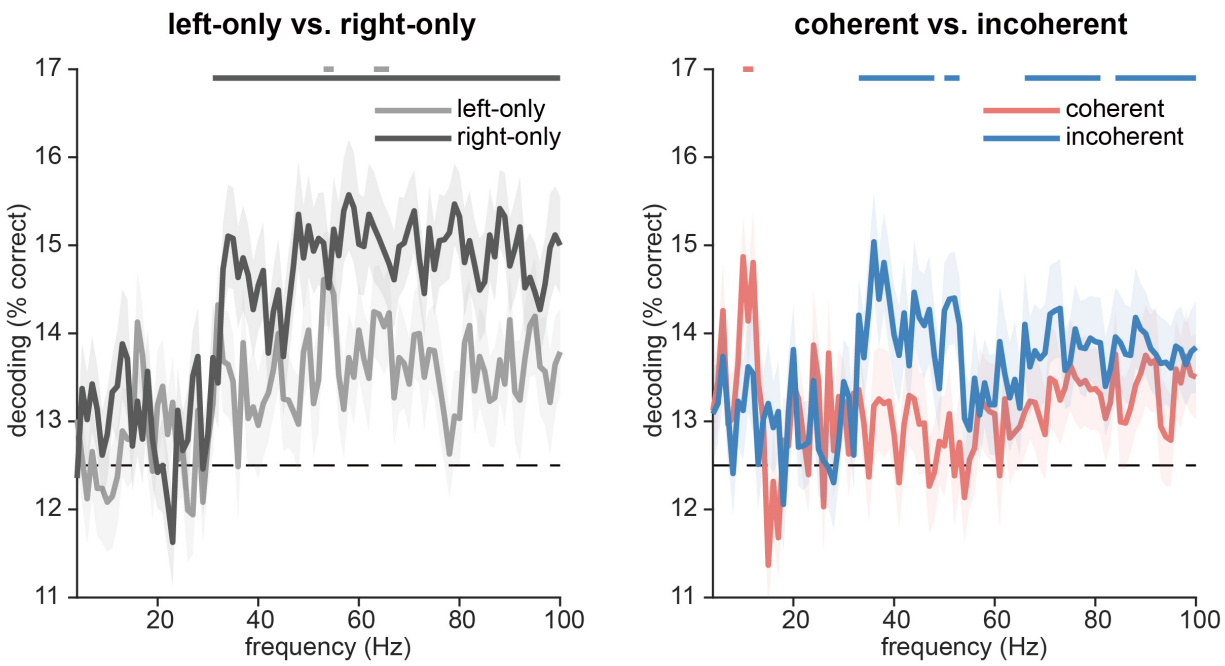

Fig. S2.

EEG frequency-resolved decoding analysis on spectral power patterns using sliding windows. We decoded between the eight video stimuli at each frequency from 4 to 100 Hz using a sliding window approach with a 5-frequency resolution, separately for each condition. The incoherent and single video stimuli were decodable from the γ frequency band, whereas coherent stimuli were decodable from the α frequency band. Line markers denote significant above-chance decoding ($p < 0.05$; FDR-corrected).

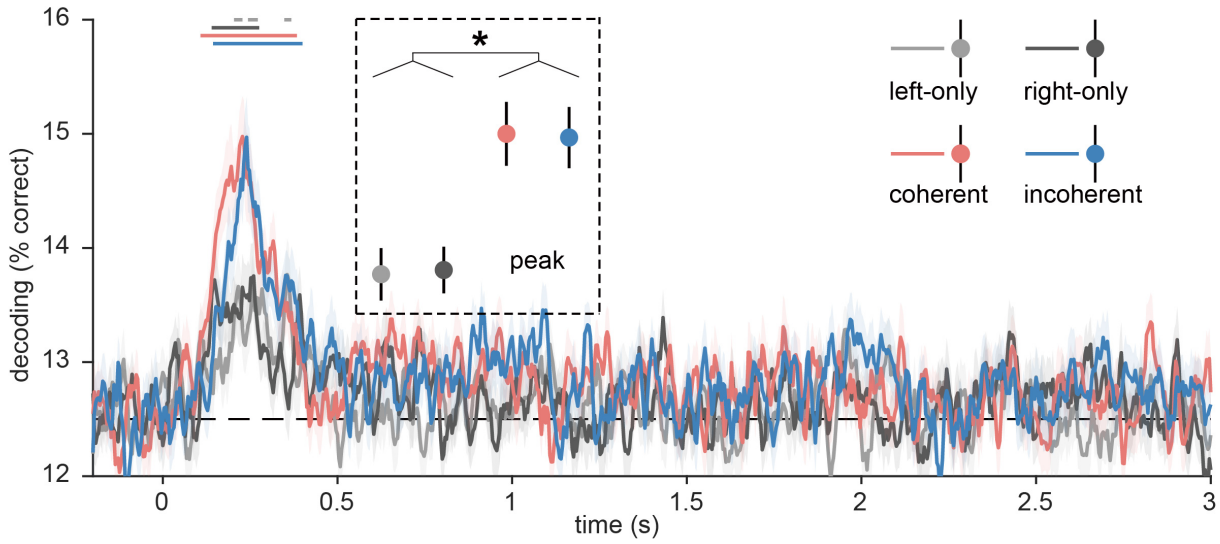

Fig. S3.

EEG time-resolved decoding on evoked response patterns. We performed decoding analysis on time-resolved broadband responses across channels to discriminate the eight video stimuli at each time from -200 ms to 3,000 ms relative to the onset of the stimulus, separately for each condition. The obtained decoding timeseries for each condition were smoothed by the moving average algorithm (6 time points). We extracted the peak decoding accuracy for each condition and then compared the decoding difference between conditions using paired t-tests. The results revealed a sustained representation of the video stimuli across the first 500 ms of processing, with stronger peak responses to two video conditions (coherent/incoherent) than to single video conditions (right-/left-only), but no differences between the coherent and incoherent conditions. Decoding onsets did not differ between the coherent and incoherent video stimuli (permutation test, $p = 0.176$). Error bars represent standard errors. Line markers denote significant above-chance decoding ($p < 0.05$; FDR-corrected). *: $p < 0.05$.

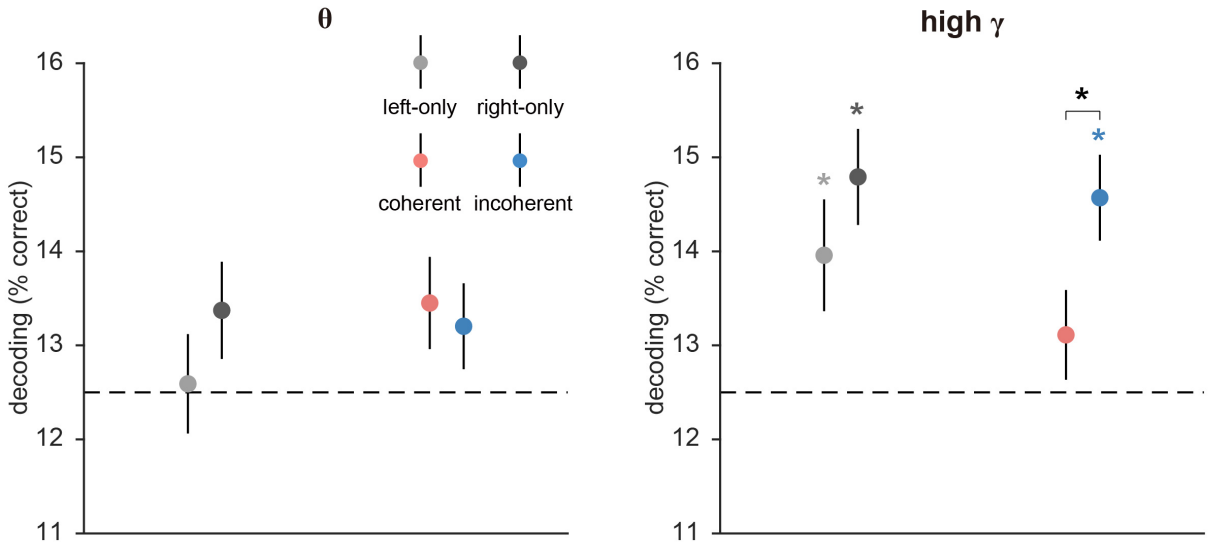

Fig. S4.

EEG decoding analysis on spectral power patterns separately for the theta (4–7 Hz) and high-gamma (71–100 Hz) frequency bands. For each frequency band (theta and high-gamma), we extracted the power of the frequencies included in that band across all channels from the power spectra, and then used the resulting patterns across channels and frequencies to classify the eight video stimuli in each condition. In the theta band, we did not find any significant above-chance decoding. In the high-gamma band, we found significant above-chance decoding for both single video stimuli and incoherent stimuli, but not for coherent stimuli. As in the 31-70 Hz gamma range, the incoherent stimuli were also decoded better than coherent stimuli. *: $p < 0.05$.

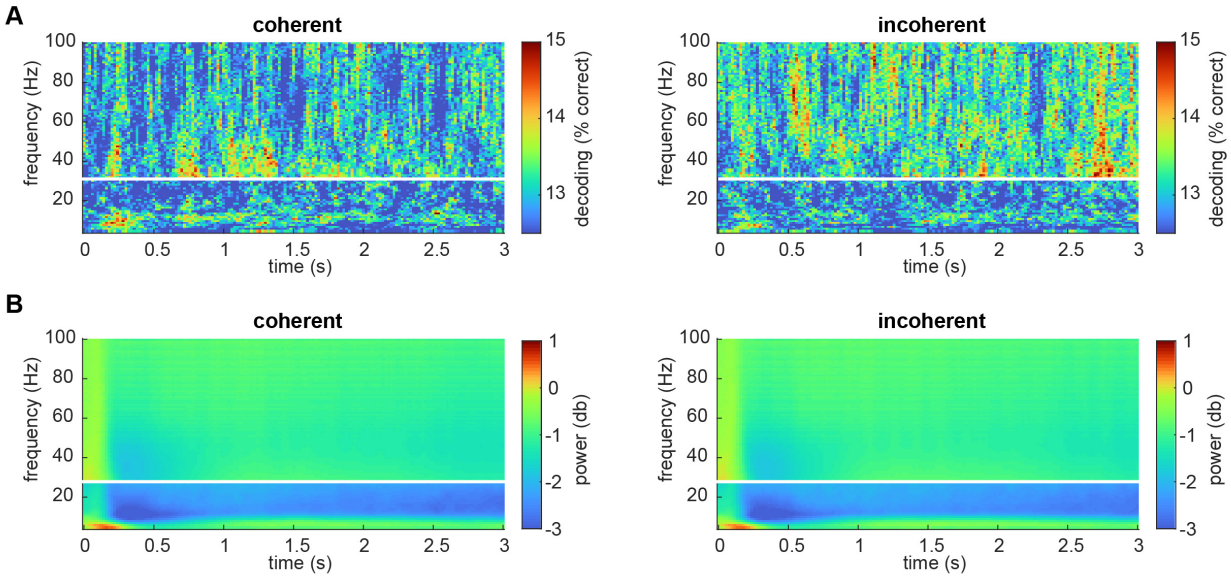

Fig. S5.

EEG time-frequency analysis. We first performed a time-frequency analysis to estimate the power at each frequency (4–100 Hz) and each time point (0–3 s) separately for each channel. As in the powerspectrum analysis (see Materials and Methods for details), we used single tapers for the low frequencies (4–30 Hz), and multitapers for high frequencies (31–100 Hz). **A)** We performed the decoding analysis at each time-frequency combination. We did not find significant differences in decoding performance between the coherent and incoherent conditions ($p < 0.05$; FDR-corrected; top two panels). The results suggested that we had insufficient statistical power for concurrently resolving the data across time and frequency, given the signal-to-noise ratio of our data. **B)** We transformed the power values to dB relative to the baseline to obtain the event-related spectral perturbation (ERSP). No significant differences were found in ERSP between the coherent and incoherent conditions ($p < 0.05$; FDR-corrected; bottom two panels). This suggests that the shift in representation from gamma to alpha dynamics is not accompanied by large-scale changes in the univariate spectral power over visual cortex.

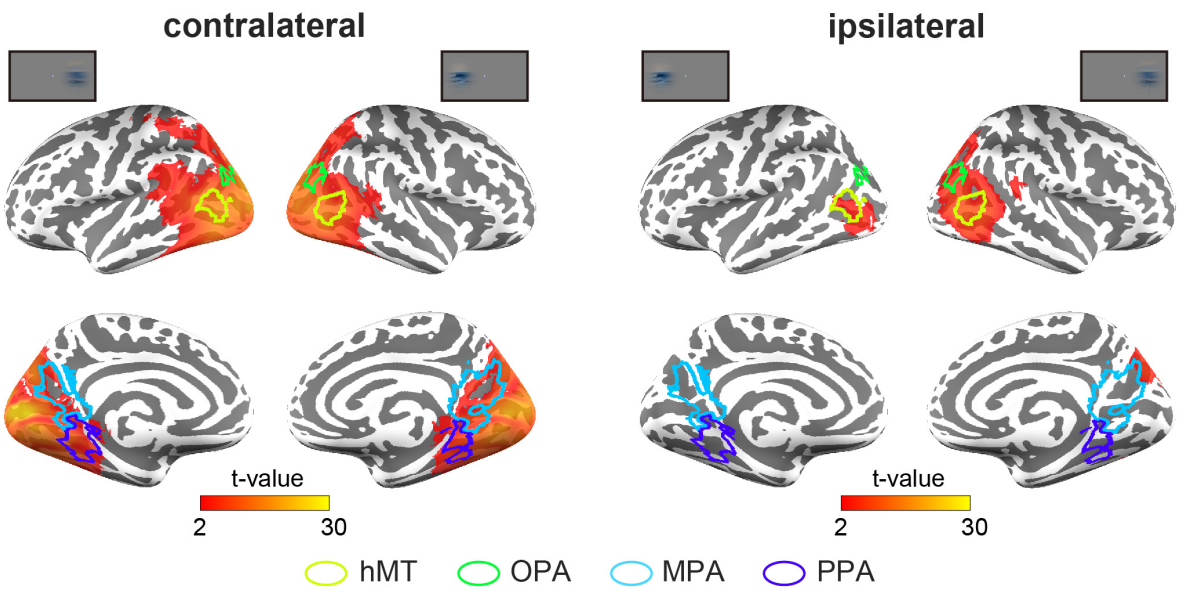

Fig. S6.

fMRI searchlight decoding analysis for right- and left-only conditions. Similar to the ROI decoding results, single video stimuli were decodable across the visual cortex in the contralateral hemisphere. In the ipsilateral hemisphere, the stimuli were primarily decodable in the parietal lobe including hMT. Multiple comparison correction was performed using GRF (voxel-level $p < 0.005$, cluster-extent $p < 0.05$).

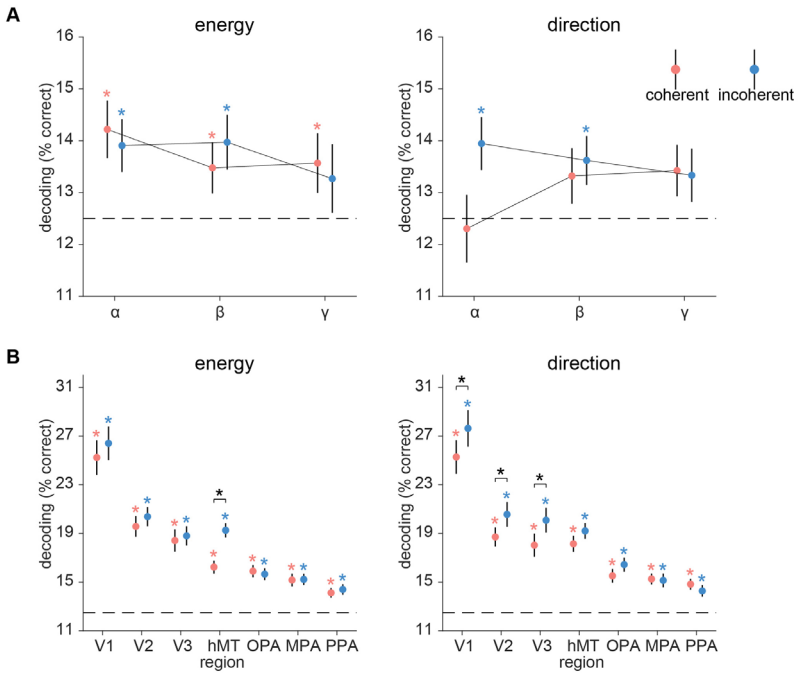

Fig. S7.

EEG and fMRI decoding analyses (grouping stimuli based on motion coherence). For each of the coherent and incoherent video stimuli, we first quantified inter-frame motion energy using the Motion Energy Analysis (MEA) software (https://psync.ch/mea/) separately for the left and right apertures, and then estimated its motion energy coherence by calculating the Pearson correlation between the left and right motion energy time-series. We split the stimulus set into two halves based on the values of motion energy coherence. Eight stimuli that are more coherent in motion energy were grouped into the motion coherent group and the other 8 stimuli were grouped into the motion incoherent group. In addition, we also grouped the stimuli based on motion direction coherence. For each stimulus, we estimated inter-frame optical flow using the Computer Vision Toolbox implemented in MATLAB and estimated its motion direction coherence by calculating the Pearson correlation between the mean motion direction time-series of the left and right apertures. Equivalent as described above, based on the values of motion direction coherence we split half the stimuli into a motion direction coherent and a motion direction incoherent group. **A)** We performed EEG frequency-resolved decoding analysis between the 8 stimuli in each frequency band (alpha, beta, gamma), separately for the newly formed coherent and incoherent groups. For the motion energy grouping, we found significant decoding for coherent stimuli in the alpha, beta, and gamma frequency bands, and significant decoding for incoherent stimuli in the alpha and beta frequency bands. For the motion direction grouping, the incoherent stimuli were decodable in the alpha and beta bands. **B)** We performed fMRI ROI decoding analysis to classify the stimuli in each ROI (V1, V2, V3, hMT, OPA, MPA, PPA), separately for the newly formed coherent and incoherent groups. For the motion energy grouping, both coherent and incoherent stimuli were decodable in all seven ROIs, and the incoherent stimuli were decoded better than coherent stimuli in the hMT. For the motion direction grouping, both coherent and incoherent stimuli were decodable in all the ROIs, and the incoherent stimuli were more decodable than coherent stimuli in V1, V2, and V3. *: $p < 0.05$ (FDR-corrected).

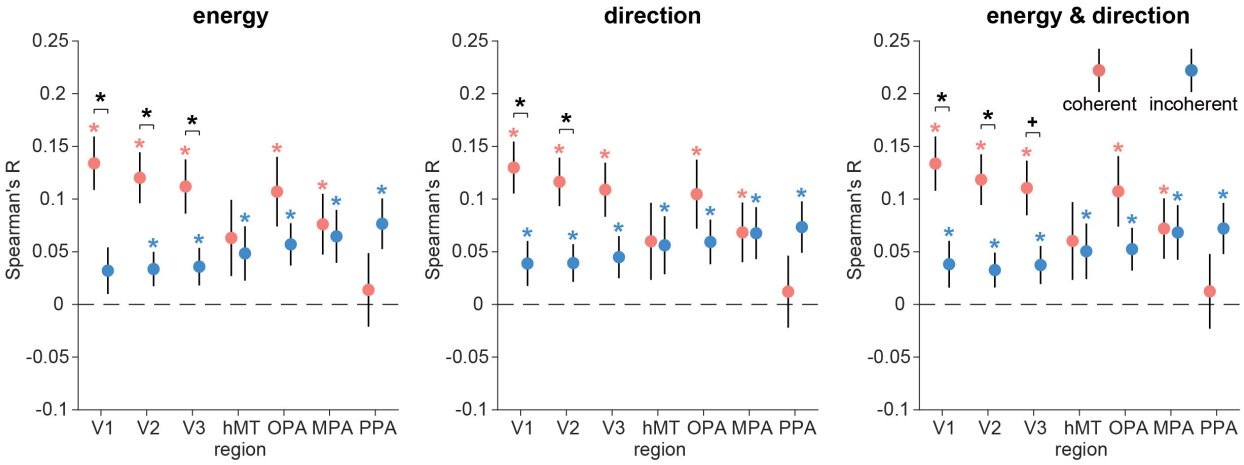

Fig. S8.

EEG-fMRI fusion analysis while controlling for motion coherence. For each of the coherent and incoherent video stimuli, we first quantified the inter-frame motion energy using the Motion Energy Analysis (MEA) software (https://psync.ch/mea/) separately for the left and right apertures, and then estimated its motion energy coherence by calculating the Pearson correlation between the left and right motion energy time-series. To construct the motion representational dissimilarity matrices (RDMs), we used the absolute difference in motion energy coherence as the distance between stimuli separately for coherent and incoherent conditions. In addition, we also constructed a motion RDM based on motion direction coherence. Specifically, for each stimulus, we estimated the inter-frame optical flow using the Computer Vision Toolbox (in MATLAB) and estimated its motion direction coherence by calculating the Pearson correlation between the mean motion direction time-series of the left and right apertures. Next, we constructed a motion direction RDM based on the absolute difference in motion direction coherence between stimuli separately for the coherent and incoherent conditions. In the EEG-fMRI fusion analysis, we calculated partial correlations between the participant-specific EEG RDMs in each frequency band (α, β, γ) and the group-averaged fMRI RDMs in each region (V1, V2, V3, hMT, OPA, MPA, PPA) that control for the motion energy RDM, motion direction RDM, or both motion RDMs, separately for the coherent and incoherent conditions. Similar to the main fusion results, we found that representations in the alpha band corresponded more strongly with representations in the early visual cortex when the videos were presented coherently, rather than incoherently, in all three analyses. Error bars represent standard errors. *: $p < 0.05$ (FDR-corrected), $^{+}$: $p < 0.05$ (uncorrected).

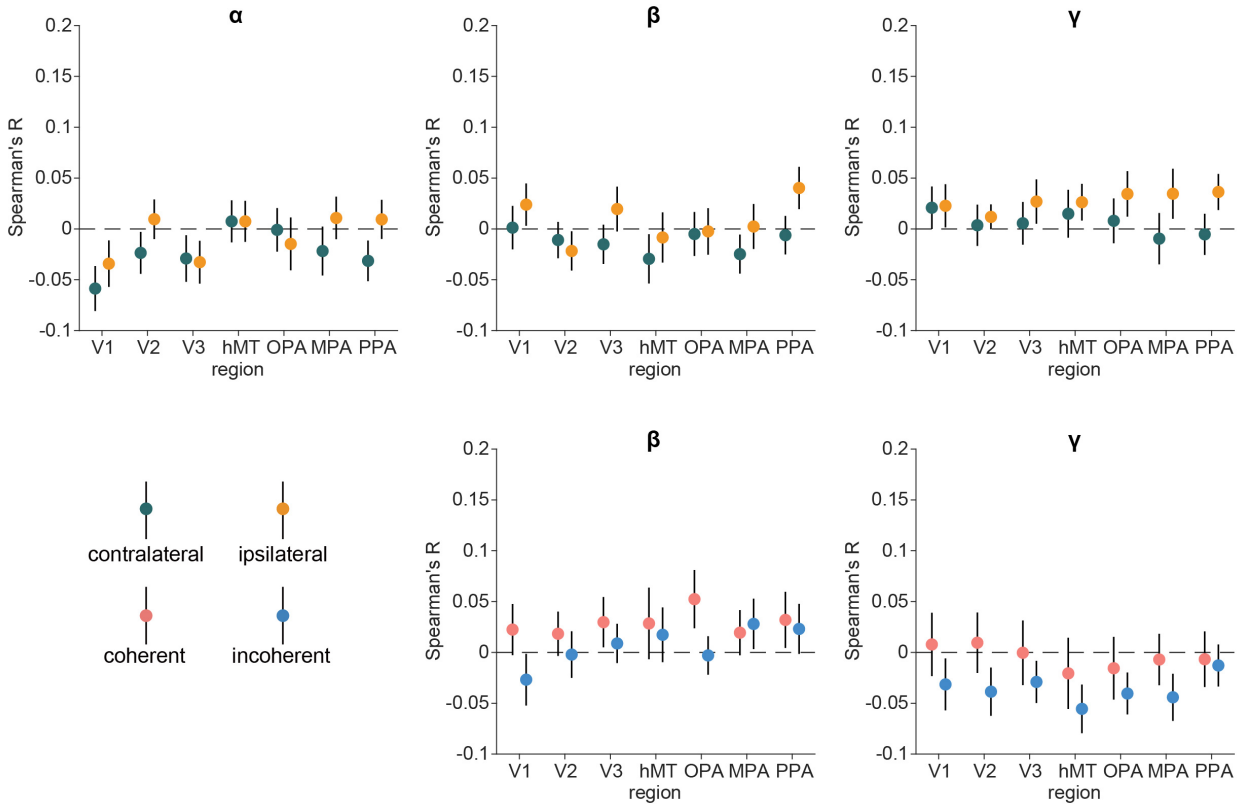

Fig. S9.

EEG-fMRI fusion analysis separately for each frequency band. For each condition, EEG representational dissimilarity matrices (RDMs) for each frequency band (α, β, γ) and fMRI RDMs for each region of interest (V1, V2, V3, hMT, OPA, MPA, PPA) were first obtained using pair-wise decoding analyses. To assess correspondences between spectral and regional representations, we calculated *Spearman*-correlations between the participant-specific EEG RDMs in each frequency band and the group-averaged fMRI RDMs in each region, separately for each condition. For the right- and left-only conditions, there was no significant correspondence between EEG responses in each frequency band and fMRI activations in each region, either for contralateral or ipsilateral presentations. For the coherent and incoherent conditions, there were no significant correspondences between β/γ responses and fMRI activations. Error bars represent standard errors.

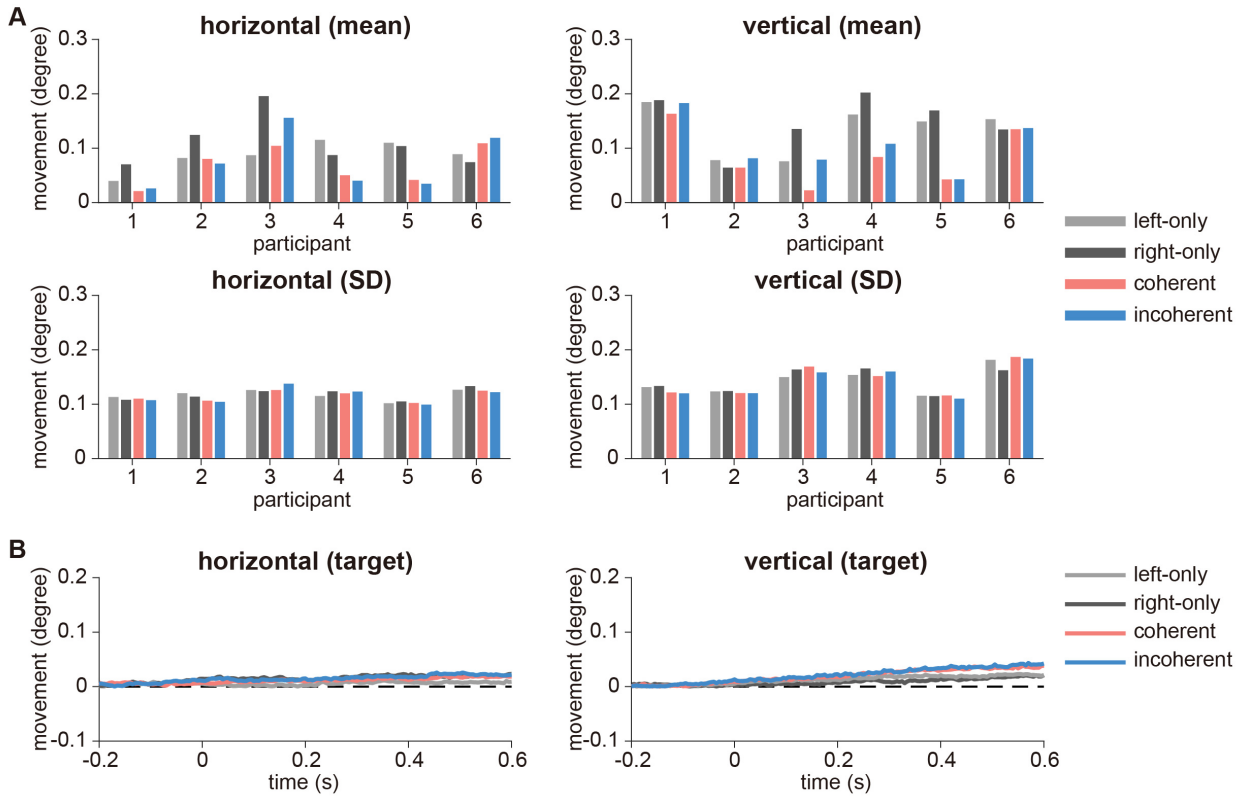

Fig. S10.

Eye-tracking data. Six participants (gender: 2 M/4 F, age: 26.5 ± 1.2 years) took part in the eye-tracking experiment using the same paradigm and the same stimuli as in the EEG and fMRI experiments. Eye movements were recorded monocularly (left eye) with an Eyelink 1000 Tower Mount (SR Research Ltd., Mississauga, Ontario, Canada) using the Psychophysics and Eyelink Toolbox extensions at 1000 Hz. Eye tracking data were segmented into epochs from -0.5 to 3.5 s relative to the onset of the stimulus, downsampled to 200 Hz, and baseline corrected. The data were then transformed from their original screen coordinate units (pixels) to visual angle units (degrees). **A)** To check fixation patterns during video presentation, we calculated the mean and standard deviation (SD) of the horizontal and vertical eye movement across time (0–3 s) in each trial and then averaged the mean and SD values across trials separately for each condition. For all participants, we found means of eye movement lower than 0.3 degrees (top two panels), and SDs of eye movement lower than 0.2 degrees (middle two panels), indicating stable central fixation. **B)** To determine whether eye movements occurred once the fixation color was detected, we extracted eye tracking data from -200 to 600 ms relative to the onset of the target color. We found no significant eye movement deviations after the target was presented, indicating that participants did not disengage from fixation after the target was presented.

Table S1.

Behavioral accuracy and response time for the color discrimination task in both EEG and fMRI experiments. Notes: Means ± standard errors (SE); N.A., not available.

	Left-only	Right-only	Coherent	Incoherent
Accuracy (EEG)	93.27 ± 1.70%	93.27 ± 1.58%	93.13 ± 1.66%	93.45 ± 1.71%
Accuracy (fMRI)	91.58 ± 1.18%	91.09 ± 1.19%	91.58 ± 1.24%	91.51 ± 1.21%
Response Time (EEG)	N.A.	N.A.	N.A.	N.A.
Response Time (fMRI)	521.9 ± 19.3 ms	521.7 ± 18.6 ms	525.2 ± 19.4 ms	526.8 ± 18.5 ms

5.3 Original publication of Study 3

Chen, L., Cichy, R. M., & Kaiser, D. (2024). Coherent categorical information triggers integration-related alpha dynamics. Journal of Neurophysiology, 131(4), 619–625. https://doi.org/10.1152/jn.00450.2023

Copyright

american physiological society

JNP JOURNAL OF NEUROPHYSIOLOGY

Original publication of Study 3

J Neurophysiol 131: 619–625, 2024.
First published February 28, 2024; doi:10.1152/jn.00450.2023

SHORT REPORT

Sensory Processing

Coherent categorical information triggers integration-related alpha dynamics

Lixiang Chen,[1,2] Radoslaw Martin Cichy,[1] and Daniel Kaiser[2,3]

[1]*Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany;* [2]*Mathematical Institute, Department of Mathematics and Computer Science, Physics, Geography, Justus-Liebig-Universität Gießen, Gießen, Germany; and* [3]*Center for Mind, Brain and Behavior (CMBB), Philipps-Universität Marburg and Justus-Liebig-Universität Gießen, Marburg, Germany*

Abstract

To create coherent visual experiences, the brain spatially integrates the complex and dynamic information it receives from the environment. We previously demonstrated that feedback-related alpha activity carries stimulus-specific information when two spatially and temporally coherent naturalistic inputs can be integrated into a unified percept. In this study, we sought to determine whether such integration-related alpha dynamics are triggered by categorical coherence in visual inputs. In an EEG experiment, we manipulated the degree of coherence by presenting pairs of videos from the same or different categories through two apertures in the left and right visual hemifields. Critically, video pairs could be video-level coherent (i.e., stem from the same video), coherent in their basic-level category, coherent in their superordinate category, or incoherent (i.e., stem from videos from two entirely different categories). We conducted multivariate classification analyses on rhythmic EEG responses to decode between the video stimuli in each condition. As the key result, we significantly decoded the video-level coherent and basic-level coherent stimuli, but not the superordinate coherent and incoherent stimuli, from cortical alpha rhythms. This suggests that alpha dynamics play a critical role in integrating information across space, and that cortical integration processes are flexible enough to accommodate information from different exemplars of the same basic-level category.

NEW & NOTEWORTHY Our brain integrates dynamic inputs across the visual field to create coherent visual experiences. Such integration processes have previously been linked to cortical alpha dynamics. In this study, the integration-related alpha activity was observed not only when snippets from the same video were presented, but also when different video snippets from the same basic-level category were presented, highlighting the flexibility of neural integration processes.

alpha rhythms; cortical feedback; multivariate pattern analysis; natural scenes; spatiotemporal coherence

INTRODUCTION

During everyday life, our visual system continuously receives intricate and dynamic information from our surroundings. To derive meaningful interpretations from these stimuli, the brain integrates dynamic sensory inputs across the visual field, culminating in a seamlessly unified, behaviorally adaptive percept of the world (1, 2).

Classic theories of vision conceptualize visual processing as a feedforward hierarchy, along which stimuli are reconstructed through hierarchical feature integration (3, 4). Under such theories, visual integration is solved along the feedforward cascade. Feedforward theories of vision, however, are challenged by the abundance of recurrent and feedback connections in the visual system (5), as well as the pivotal role of attentional feedback processes in constructing visual percepts (6). Our recent study (7) indeed revealed that feedback processes are critical for spatial integration when stimuli are spatiotemporally coherent and afford integration. Such feedback is evident from stimulus-specific representations in neural alpha dynamics, which can be spatially localized to early visual cortex. This result suggests that integration-related feedback traverses the hierarchy in alpha rhythms from high-level visual cortex all the way to retinotopic early visual cortex. Our findings align well with theories that posit a multiplexing of information, where feedback is specifically routed via low-frequency alpha or beta rhythms (8–11).

However, our previous study used stimuli that were either coherent at the level of the individual video (i.e., two parts of the same video played in the left and right hemifields) or

Correspondence: L. Chen (lixiang.chen@fu-berlin.de).
Submitted 6 December 2023 / Revised 23 January 2024 / Accepted 22 February 2024

highly incoherent (i.e., two entirely different videos in the two hemifields). We thus could not address what level of spatiotemporal coherence in the stimuli is needed to trigger integration-related alpha dynamics.

In this study, we address this question in an EEG experiment. We manipulated the degree of spatiotemporal coherence by presenting videos from the same or different categories through two apertures left and right of the central fixation. Our findings showed that stimuli coherent at the level of individual videos are coded in cortical alpha dynamics. Critically, similar representations in alpha rhythms were also observed when different videos from the same basic-level category were presented, but not when the videos were from the same superordinate category or from different superordinate categories. This suggests that neural integration exhibits some flexibility, so that broadly consistent videos from the same category can trigger alpha dynamics linked to integration.

MATERIALS AND METHODS

Participants

Twenty-five healthy participants (14 females, mean age: 24.1 ± 3.9 yr), with normal or corrected-to-normal vision, participated in the experiment. A minimum sample size of 24 was determined using G*Power (12), with an effect size of 0.25 (comparison of decoding performance between the coherent and incoherent conditions in the alpha frequency band in the EEG study) as derived from our previous study (7), a significance level of 0.05, and a power of 0.8. All participants provided written informed consent before taking part in the experiment and they received either course credit or monetary reimbursement for their participation. The experiment was approved by the ethical committee of the Department of Education and Psychology at Freie Universität Berlin and was conducted following the Declaration of Helsinki.

Stimuli and Design

We selected sixteen 3-s videos (30 Hz) depicting various everyday events for the experiment. The videos were from four categories (4 exemplars for each category): birds flying, camels walking, cars running, and trains moving (Fig. 1*A*). We presented videos through two apertures left and right of the central fixation (7). The apertures had a diameter of 6° visual angle, and the closest distance between the aperture and the central fixation point was 2.64° visual angle. The central fixation dot was displayed at a visual angle of 0.44°.

We designed four different conditions by showing parts from the same video or different videos (Fig. 1*B*). In the video-level coherent condition, we displayed two parts of the same video through the apertures. In the basic-level coherent condition, the two parts were from two different videos belonging to the same category (e.g., bird video 1 and bird video 2). In the superordinate coherent condition, the two parts were from two different videos belonging to the same superordinate category (e.g., bird video 1 and camel video 1). In the incoherent condition, the two parts were from videos belonging to different superordinate categories (e.g., bird video 1 and car video 1). In the basic-level coherent condition, the videos of each category were presented in fixed pairs (e.g., bird video 1 and bird video 2, bird video 3 and bird video 4). We similarly paired the videos for the superordinate coherent (e.g., bird video 1 and camel video 1, bird video 2 and camel video 2) and incoherent conditions (e.g., bird video 1 and car video 1, bird video 2 and car video 2). Therefore, there were a total of 64 unique video stimuli (16 stimuli for each of the 4 conditions).

Participants were comfortably seated at a distance of 60 cm from a monitor with a resolution of 1,680 × 1,050 pixels and a refresh rate of 60 Hz. The presentation of stimuli and recording of participants' behavioral responses were controlled using MATLAB and the Psychophysics Toolbox (13, 14). Each trial began with a 0.5-s fixation dot. Subsequently, a unique video stimulus was shown for 3 s, during which the color of the fixation changed periodically (every 200 ms) and turned either green or yellow at a single random point in the sequence (but not the first or last point). After the video, participants were presented with a response screen, prompting them to indicate whether a green or yellow fixation dot had appeared in the sequence. The next trial would not start until the participant's response was received. Participants were instructed to keep central fixation during the video presentation to ensure that the two videos presented stimulated different visual fields. An example trial for the basic-level coherent condition is shown in Fig. 1*C*. In the experiment, participants performed the color discrimination task on fixation with very high accuracy (video-level coherent: 95.8 ± 2.9%, basic-level coherent: 96.2 ± 3.0%, superordinate coherent: 96.3 ± 3.1%, incoherent: 96.0 ± 2.9%), indicating reliable fixation control. In the experiment, each of the 64 unique stimuli was shown 12 times. A total of 768 trials were presented in random order.

EEG Recording and Preprocessing

EEG data were acquired at a sampling rate of 1,000 Hz using an EASYCAP 64-electrode system with a Brainvision actiCHamp amplifier. Electrodes were arranged according to the 10-10 system. All electrodes were referenced online to the FCz.

We preprocessed the data using Fieldtrip (15). We first filtered the data at 1–100 Hz and epoched the data from −0.5 to 3.5 s relative to the onset of the stimulus. Then, we performed baseline correction by subtracting the mean signal in the prestimulus window (−0.5 to 0 s), after which we downsampled the data to 200 Hz. Next, we conducted visual inspection to exclude noisy trials and channels, and then interpolated the removed channels (2.6 ± 1.2 channels) using their neighboring channels. Finally, we used independent component analysis (ICA) to identify and remove artifacts associated with blinks and eye movements.

EEG Spectral Analysis

We performed spectral analysis on the preprocessed EEG data using FieldTrip, in the same way as in our previous study (7). For each trial, we estimated power spectra separately for each channel within the alpha (8–12 Hz), beta (13–30 Hz), and gamma (31–70 Hz) frequency bands. The analysis was done for the whole period of stimulus presentation (0–3 s). For the low frequency of 8–30 Hz, we applied a single taper with a Hanning window, with a step size of 1 Hz for the

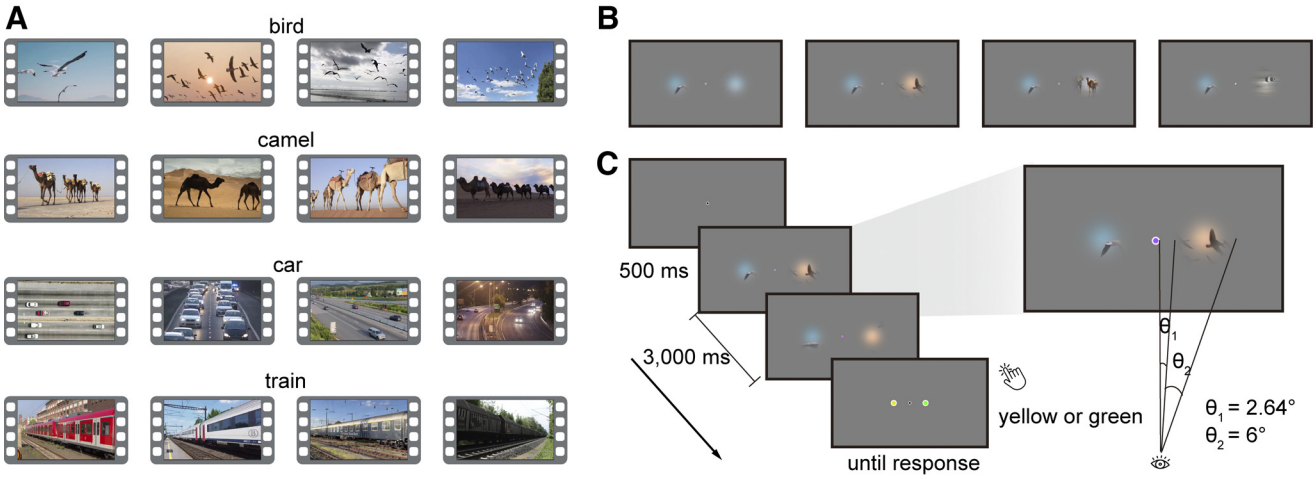

Figure 1. Stimuli and experimental design. *A*: snapshots from the video stimulus set. *B*: in the experiment, videos were presented through two apertures left and right of the central fixation, manipulated in four conditions: video-level coherent, coherent in the basic-level category, coherent in the superordinate category, and incoherent. *C*: during video presentation, the color of the central dot changed periodically (every 200 ms) and participants were asked to report whether there was a green or yellow fixation dot included in the sequence.

alpha band and 2 Hz for the beta band. For the gamma band, we used the discrete prolate spheroidal sequences (DPSS) multitaper method with ±8 Hz smoothing (in steps of 2 Hz).

Multivariate Decoding Analysis

We performed multivariate decoding analysis to investigate the frequency-specific representations of video stimuli using CoSMoMVPA (16) and LIBSVM (17). Given that integration-related alpha dynamics originate from retinotopic visual cortex (7), we selected 17 parietal and occipital (PO) channels (Oz, O1, O2, POz, PO3, PO4, PO7, PO8, Pz, P1, P2, P3, P4, P5, P6, P7, P8) over visual cortex (18) for our analysis. From these channels, we extracted the patterns of spectral power across these channels to classify the four video pairings within each condition (video-level coherent, basic-level coherent, superordinate coherent, incoherent), separately for the alpha, beta, and gamma frequency bands. We conducted the classification using the linear support vector machine (SVM) with leave-one-trial-out cross-validation. One trial was assigned to the test set, whereas the remaining $n - 1$ trials were used to train the classifier. We conducted the classification repeatedly until every trial was left out once, and averaged the resulting accuracies across trials. In each classification, we balanced the number of trials across categories, resulting in a maximum of 188 trials for the training set (47 for each category). To reduce the dimensionality of the data, we applied principal component analysis (PCA) to the data before classification (19). We performed PCA on the training data, and then projected the resulting PCA solution onto the testing data. We selected a subset of components that explained 99% of the variance of the training data. As a result, we obtained decoding accuracy for each frequency band and each condition, indicating the degree to which the video stimuli were accurately represented in different frequency bands. We performed a one-sample t test to compare the decoding performance against chance level (25%) for testing whether the stimuli could be represented in each frequency band [false discovery rate (FDR)-correction, $P < 0.05$]. Furthermore, to investigate whether the frequency-specific representations were modulated by the degree of stimulus coherence, we conducted a two-way ANOVA (4 conditions $\times$ 3 frequencies) and post hoc paired t tests to compare the decoding performance between conditions separately for each frequency band (FDR-correction, $P < 0.05$).

It is worth noting that, although in some conditions different categories were shown in both hemifields, decoding between the different videos should still be possible in principle: as the left and right side of the display are analyzed in the right and left hemisphere, respectively, each hemisphere offers information about the category presented in its contralateral hemifield.

To investigate where the effects are localized and whether the effects are maximum over visual cortex, we performed searchlight decoding analysis. For each channel, we defined a searchlight including itself and its 10 nearest neighboring channels and then used the spectral power patterns across these channels to decode between the four video pairings within each condition, separately for each frequency band (alpha, beta, and gamma). Identically to the decoding analysis using PO channels, we used leave-one-trial-out cross-validation and applied principal component analysis (PCA) for the classification. The whole classification process was iterated over all channels. As a result, we obtained decoding accuracy in each channel separately for each frequency band and each condition. To localize the significant decoding for each condition, we used a one-sample permutation test (10,000 iterations), comparing the decoding accuracy against the chance level (25%) in each channel and then performing cluster-based multiple comparison corrections ($P < 0.05$).

To investigate the representation of stimuli in time-locked broadband responses, we performed time-resolved decoding analysis. We classified between the four video pairings within each condition using broadband responses across PO channels at each time point from −0.1 to 1 s relative to stimulus onset (decoding already approached chances level well before 1 s). The decoding parameters were identical to the frequency-resolved decoding analysis using PO channels. The resulting decoding timeseries were smoothed with a moving average of five time points. Separately for each time

point, we used a one-sample t test to compare decoding against chance and paired t tests to compare the difference between conditions. Multiple comparison corrections were conducted using FDR ($P < 0.05$), and only clusters of at least five consecutive significant time points were considered (19).

Following our previous study (7), we primarily investigated integration-related effects in spectral EEG power. However, in principle, such effects may also be represented in the phase of neural rhythms (e.g., resulting from the different temporal dynamics of the videos). We performed the Fourier transform on the EEG preprocessed data and extracted the phase angles from the obtained complex Fourier spectrum. We then decoded the four video pairings within each condition using patterns of spectral phase across PO channels separately for alpha, beta, and gamma bands. Here, we performed the decoding analysis and statistical comparisons using the same approaches as in the frequency-resolved decoding analysis on spectral power.

Eye Tracking Recording and Processing

Eye movements were recorded monocularly (right eye) at 1,000 Hz with an Eyelink 1000 Tower Mount (SR Research Ltd., Mississauga, ON, Canada) using the Psychophysics and Eyelink Toolbox extensions (20). At the beginning of the experiment, we used a standard 9-point calibration to calibrate eye position.

We preprocessed eye-tracking data using Fieldtrip. Specifically, we segmented the data into epochs from −0.5 to 3.5 s relative to stimulus onset and downsampled the data to a sampling rate of 200 Hz. The preprocessed data were transformed from their original screen coordinate units (pixels) to visual angle units (degrees). We next excluded the trials that were removed in the EEG analysis. To check the fixation stability, we calculated the mean and standard deviation (SD) of the horizontal and vertical eye movements during video presentation (0–3 s) in each trial and then averaged the mean and SD values across trials separately for each condition. We found no significant differences in both horizontal [comparisons of mean: $F(3,72) = 0.74$, $P = 0.53$; comparisons of SD: $F(3,72) = 0.56$, $P = 0.64$] and vertical eye movements [comparisons of mean: $F(3,72) = 0.71$, $P = 0.55$; comparisons of SD: $F(3,72) = 0.97$, $P = 0.41$] between the four conditions.

RESULTS

To study the frequency-specific representations of video stimuli, we decoded between video stimuli within each condition (video-level coherent, basic-level coherent, superordinate coherent, and incoherent) using patterns of spectral power across channels separately for each frequency band (alpha, beta, gamma). In this analysis, we found significant above-chance decoding only in the alpha band and for the video-level coherent and basic-level coherent stimuli (Fig. 2*A*). Using a 4-condition × 3-frequency two-way ANOVA, we identified a significant interaction effect between condition and frequency [$F(6,144) = 3.75$, $P = 0.002$]. Subsequently, we conducted post hoc t tests to examine differences between conditions in each frequency band.

In the alpha band, we observed a decrease in decoding accuracy as the spatial coherence of stimuli reduced, indicating that integration-related alpha activity is modulated by the coherence of the stimuli. Specifically, the video-level coherent stimuli were decoded better than the superordinate coherent stimuli [$t(24) = 3.90$, $P < 0.001$; Fig. 2*A*] as well as better than the incoherent stimuli [$t(24) = 4.37$, $P < 0.001$; Fig. 2*A*]. Similarly, basic-level coherent stimuli were also more decodable than both the superordinate coherent stimuli [$t(24) = 3.317$, $P = 0.004$; Fig. 2*A*] and incoherent stimuli [$t(24) = 3.083$, $P = 0.005$; Fig. 2*A*]. We found no significant difference between the video-level coherent and the basic-level coherent conditions [$t(24) = 1.290$, $P = 0.314$]. Importantly, the difference in alpha decoding across conditions was not related to an absence of stimulus representation in the more incoherent conditions in the first place: When decoding from time-locked broadband responses, we found significant decoding for all conditions within the first 500 ms of processing that leveled off toward chance level during the first second (Fig. 2*B*). However, there was no significant between-condition difference in decoding from the time-locked responses (Fig. 2*B*), consistent with our previous results (7). In addition, we found no significant effects in the beta and gamma frequency bands (all $P > 0.05$).

Given that these analyses were only conducted on rhythmic patterns in the PO channels (see MATERIALS AND METHODS), we aimed to confirm that these effects indeed originate over visual cortex in a channel-space searchlight analysis (21). In this analysis, we observed significant decoding only in the alpha band, primarily in the PO channels, and only for the video-level coherent and basic-level coherent stimuli (Fig. 2*C*). We found no effects for the other two, more incoherent conditions. Together, these results suggest that alpha activity plays a key role in the integration of visual information across space. They further highlight that integration-related alpha dynamics are not only triggered when stimuli are video-level coherent (i.e., when the same video was shown through the apertures), but that integration processes are flexible enough to accommodate information that comes from videos belonging to the same basic-level category.

To investigate whether the integration-related effects were also represented in the phase of neural rhythms, we used the spectral phase to decode between stimuli. Although the basic-level coherent stimuli were decodable in the alpha band and incoherent stimuli were decodable in the beta band (Fig. 2*D*), we did not find reliable differences between conditions (all $P > 0.05$, FDR-corrected; Fig. 2*D*). This suggests that integration-related stimulus information is coded in the power of cortical alpha dynamics.

DISCUSSION

In this study, we investigated the involvement of alpha dynamics in the integration of visual information. We specifically asked whether integration-related alpha dynamics are also observed when videos are broadly consistent in category. Utilizing multivariate decoding analysis on spectrally resolved EEG data, we show that both video-level coherent and basic-level coherent stimuli were decodable from alpha-band EEG activity. In contrast, we found no alpha-band decoding for the superordinate coherent and incoherent stimuli. Our results suggest that categorical coherence of natural videos modulates the involvement of alpha-frequency activity: Alpha-related integration is triggered not

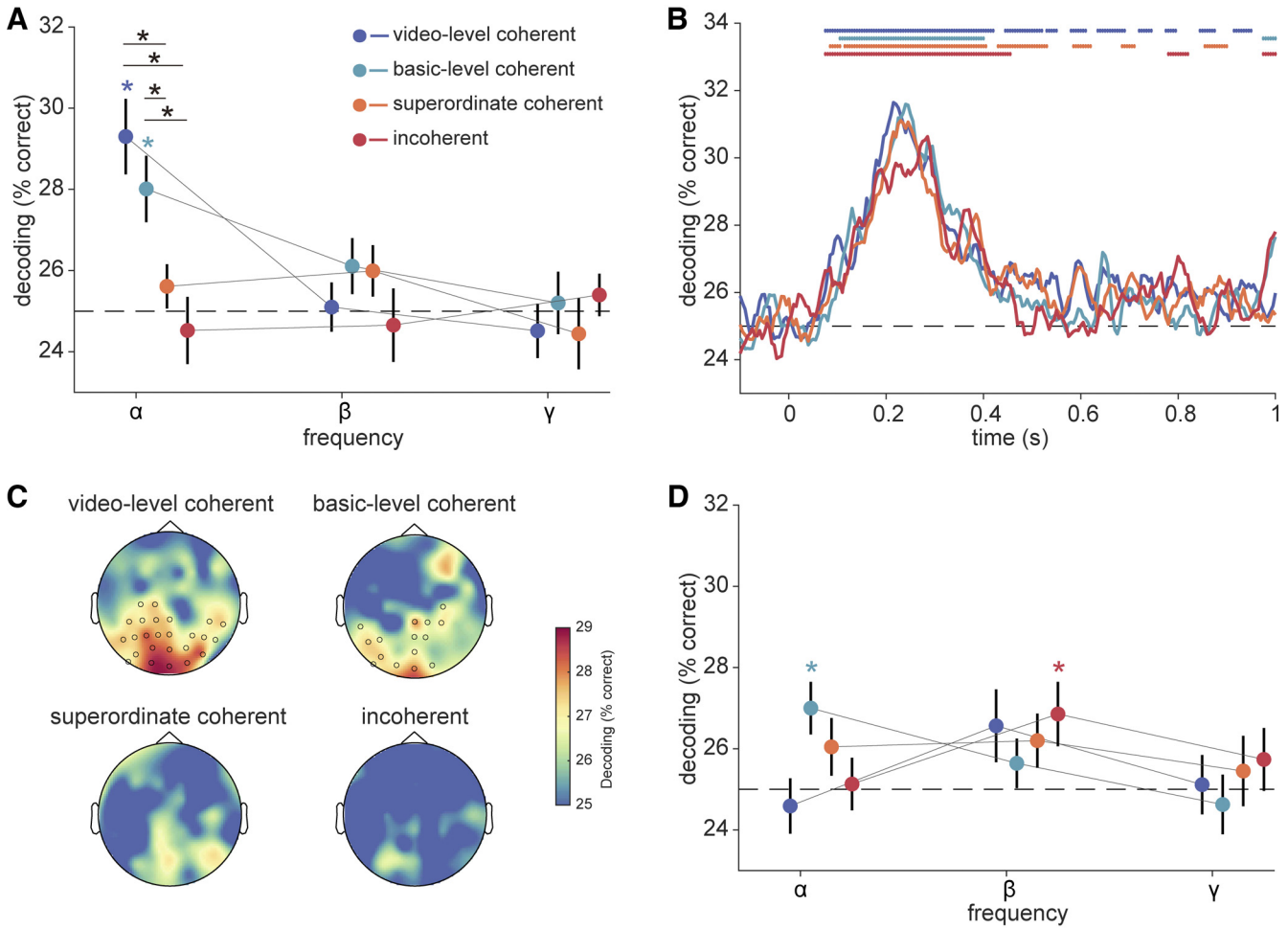

Figure 2. EEG decoding analysis. *A*: EEG frequency-resolved decoding analysis on spectral power. In each condition, we classified the four video pairings within each condition (video-level coherent, basic-level coherent, superordinate coherent, and incoherent) using patterns of spectral power across 17 parietal and occipital (PO) channels, separately for each frequency band (alpha, beta, gamma). We found significant decoding only in the alpha band and only for the video-level coherent and basic-level coherent stimuli (indicated by asterisks color-coded as result dots). In addition, the stimuli in the video-level coherent and basic-level coherent conditions were decoded better than the stimuli in the superordinate coherent and incoherent conditions (indicated by black asterisks over lines connecting compared data points). These results suggest that integration-related alpha dynamics are not only observed when videos are video-level coherent, but also when similar videos are from the same basic-level category. Error bars represent standard errors. $*P < 0.05$ (FDR-corrected). *B*: EEG time-resolved decoding analysis. We decoded between the four video pairings within each condition using time-resolved broadband responses across 17 PO channels at each time point from −0.1 to 1 s relative to the onset of the stimulus. We found significant decoding for all conditions within the first 500 ms of processing but no significant differences between conditions. Line markers denote significant above-chance decoding color-coded as result curves ($P < 0.05$, FDR-corrected). *C*: EEG searchlight decoding analysis. For each channel, we defined a searchlight including itself and its 10 nearest neighboring channels, and then used the patterns of alpha power across these 11 channels to decode between the four video pairings within each condition. We found significant decoding only for the video-level coherent and basic-level coherent stimuli primarily in PO channels (circles reflect significant channel locations). *D*: EEG frequency-resolved decoding analysis on spectral phase. We classified the four video pairings within each condition using patterns of spectral phase across 17 PO channels, separately for alpha, beta, and gamma frequency bands. We found no significant differences between conditions. Error bars represent standard errors. $*P < 0.05$ (FDR-corrected). FDR, false discovery rate.

only when visual stimuli are entirely coherent but also when these stimuli share common attributes resulting from their basic-level category membership.

Our results support our previous finding (7) that alpha dynamics play a key role in the integration of visual information across space. Together with a significant correspondence between alpha activity and V1 response in our previous study, we interpret the coding of stimulus-specific information in alpha as integration-related feedback. This interpretation is in line with a series of studies demonstrating that alpha rhythms carry cortical feedback from higher-order brain regions (11, 22, 23), and also encode stimulus-specific information (24, 25). This perspective highlights the dynamic and active role that alpha rhythms may play in cognitive processes, in contrast to the passive or inhibitory roles often ascribed to them (26–30). As proposed in our previous study (7), the observed integration processes likely originate from the adaptive comparison of contralateral feedforward information with ipsilateral feedback information obtained through interhemispheric connections among regions in high-level visual cortex. This comparison then triggers rhythmic feedback processes that aid the analysis of inputs at lower levels of the hierarchy, for instance by predicting upcoming inputs from the spatiotemporal structure of the previous input.

Our results indicate that integration-related alpha dynamics can be triggered not only by the presentation of video snippets from the same video, but also by the presentation of different videos from the same basic-level category. This suggests a spectral signature of the category-level nature of feedback information used for visual integration. The basic level is defined as the level that has the highest degree of cue validity (31). Basic level categories maximize the number of attributes shared by members of the category while minimizing the number of attributes shared with other categories. This sweet spot might be the one also used by the brain when implementing integration. However, our study cannot entirely clarify whether the integration-related alpha activity is indeed triggered by a more abstract coherence in basic-level category, presumably coded in high-level visual cortex (32, 33) or by the spatiotemporal coherence of visual features associated with a category (34–36). Establishing a special role of the basic level in integration would require a systematic comparison of integration not only on the basic and superordinate levels but also on the subordinate level. Another interesting open question concerns whether different task demands (e.g., tasks requiring perceptual decisions on the stimuli themselves) modulate the neural correlated integration. For instance, the integration-related brain responses could vary as a function of the task requiring global versus local perception.

In our previous study, we found that gamma rhythms, previously associated with feedforward processing in visual cortex (8–11), carried more information about incoherent than about coherent inputs, suggesting that feedforward processing is to some degree dominated by integration-related feedback (7). By contrast, we were not able to decode between the videos from gamma rhythms across all four conditions in the current study. Several factors may explain this difference. First, a different group of participants were scanned with a different EEG system for the current study. As gamma activity can be weak and unreliable in EEG recordings, it may not be systematically observed in each experiment (37). Second, we used different stimuli than in our previous report. In the previous study, we tried to maximize incoherence by picking very different videos (featuring different colors, movements, etc.). Here we designed the experiment without focusing on maximizing such differences. However, such more drastic incoherence may be needed to induce reliable gamma activity: Given the extended presentation duration of the video stimuli (3 s) and the absence of rapid or unexpected visual events, reliable predictions may explain away feedforward inputs carried by gamma rhythms when the videos are similar enough on some dimensions. Further studies are needed to clarify the role of gamma dynamics in coding feedforward information propagation in similar paradigms.

Taken together, our findings emphasize the key role of alpha dynamics in the construction of coherent and unified visual percepts during naturalistic vision. They further suggest that integration-related alpha dynamics does not operate in an all-or-none fashion, but that the coarse coherence between inputs stemming from the same basic-level category can effectively trigger neural correlates of integration.

DATA AVAILABILITY

Data and code for the study are available at https://doi.org/10.17605/osf.io/phmda.

ACKNOWLEDGMENTS

The authors thank the HPC Service of ZEDAT, Freie Universität Berlin (38), for computing time. Preprint is available at https://doi.org/10.1101/2023.12.04.569908.

GRANTS

L.C. is supported by a PhD stipend from the China Scholarship Council (CSC). R.M.C. is supported by the Deutsche Forschungsgemeinschaft (DFG) under Grant Nos. CI241/1-1, CI241/3-1, CI241/7-1 and by a European Research Council (ERC) starting Grant ERC-2018-STG 803370. D.K. is supported by the Deutsche Forschungsgemeinschaft (DFG) Grants SFB/TRR135, Project Number 222641018; KA4683/5-1, Project Number 518483074, "The Adaptive Mind," funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art, and an ERC Starting Grant PEP, ERC-2022-STG 101076057.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

DISCLAIMERS

Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

AUTHOR CONTRIBUTIONS

L.C. and D.K. conceived and designed research; L.C. performed experiments; L.C. and D.K. analyzed data; L.C., R.M.C., and D.K. interpreted results of experiments; L.C. prepared figures; L.C. and D.K. drafted manuscript; L.C., R.M.C., and D.K. edited and revised manuscript; L.C., R.M.C., and D.K. approved final version of manuscript.

REFERENCES

1. **Block N.** Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav Brain Sci* 30: 481–499, 2007. doi:10.1017/S0140525X07002786.
2. **Cohen MA**, **Dennett DC**, **Kanwisher N.** What is the bandwidth of perceptual experience? *Trends Cogn Sci* 20: 324–335, 2016. doi:10.1016/j.tics.2016.03.006.
3. **Riesenhuber M**, **Poggio T.** Hierarchical models of object recognition in cortex. *Nat Neurosci* 2: 1019–1025, 1999. doi:10.1038/14819.
4. **DiCarlo JJ**, **Cox DD.** Untangling invariant object recognition. *Trends Cogn Sci* 11: 333–341, 2007. doi:10.1016/j.tics.2007.06.010.
5. **Kravitz DJ**, **Saleem KS**, **Baker CI**, **Ungerleider LG**, **Mishkin M.** The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn Sci* 17: 26–49, 2013. doi:10.1016/j.tics.2012.10.011.
6. **Roelfsema PR**, **Lamme VAF**, **Spekreijse H.** Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395: 376–381, 1998. doi:10.1038/26475.
7. **Chen L**, **Cichy RM**, **Kaiser D.** Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision. *Sci Adv* 9: eadi2321, 2023. doi:10.1126/sciadv.adi2321.

8. **Bastos AM**, **Usrey WM**, **Adams RA**, **Mangun GR**, **Fries P**, **Friston KJ.** Canonical microcircuits for predictive coding. *Neuron* 76: 695–711, 2012. doi:10.1016/j.neuron.2012.10.038.
9. **van Kerkoerle T**, **Self MW**, **Dagnino B**, **Gariel-Mathis M-A**, **Poort J**, **van der Togt C**, **Roelfsema PR.** Alpha and γ oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc Natl Acad Sci USA* 111: 14332–14341, 2014. doi:10.1073/pnas.1402773111.
10. **Fries P.** Rhythms for cognition: communication through coherence. *Neuron* 88: 220–235, 2015. doi:10.1016/j.neuron.2015.09.034.
11. **Michalareas G**, **Vezoli J**, **van Pelt S**, **Schoffelen J-M**, **Kennedy H**, **Fries P.** Alpha-β and γ rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* 89: 384–397, 2016. doi:10.1016/j.neuron.2015.12.018.
12. **Faul F**, **Erdfelder E**, **Lang A-G**, **Buchner A.** G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39: 175–191, 2007. doi:10.3758/BF03193146.
13. **Brainard DH.** The Psychophysics Toolbox. *Spat Vis* 10: 433–436, 1997. doi:10.1163/156856897X00357.
14. **Pelli DG.** The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10: 437–442, 1997. doi:10.1163/156856897X00366.
15. **Oostenveld R**, **Fries P**, **Maris E**, **Schoffelen J-M.** FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011: 156869, 2011. doi:10.1155/2011/156869.
16. **Oosterhof NN**, **Connolly AC**, **Haxby JV.** CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Front Neuroinform* 10: 27, 2016. doi:10.3389/fninf.2016.00027.
17. **Chang C-C**, **Lin C-J.** LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2: 1–27, 2011. doi:10.1145/1961189.1961199.
18. **Kaiser D**, **Turini J**, **Cichy RM.** A neural mechanism for contextualizing fragmented inputs during naturalistic vision. *eLife* 8: e48182, 2019. doi:10.7554/eLife.48182.
19. **Chen L**, **Cichy RM**, **Kaiser D.** Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations. *Cereb Cortex* 32: 3553–3567, 2022. doi:10.1093/cercor/bhab433.
20. **Cornelissen FW**, **Peters EM**, **Palmer J.** The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. *Behav Res Methods Instrum Comput* 34: 613–617, 2002. doi:10.3758/BF03195489.
21. **Kaiser D**, **Oosterhof NN**, **Peelen MV.** The neural dynamics of attentional selection in natural scenes. *J Neurosci* 36: 10522–10528, 2016. doi:10.1523/JNEUROSCI.1385-16.2016.
22. **Bastos AM**, **Vezoli J**, **Bosman CA**, **Schoffelen J-M**, **Oostenveld R**, **Dowdall JR**, **De Weerd P**, **Kennedy H**, **Fries P.** Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85: 390–401, 2015. doi:10.1016/j.neuron.2014.12.018.
23. **Bastos AM**, **Lundqvist M**, **Waite AS**, **Kopell N**, **Miller EK.** Layer and rhythm specificity for predictive routing. *Proc Natl Acad Sci USA* 117: 31459–31469, 2020. doi:10.1073/pnas.2014868117.
24. **Xie S**, **Kaiser D**, **Cichy RM.** Visual imagery and perception share neural representations in the α frequency band. *Curr Biol* 30: 2621–2627.e5, 2020 [Erratum in Curr Biol 30: 3062, 2020]. doi:10.1016/j.cub.2020.04.074.
25. **Kaiser D.** Spectral brain signatures of aesthetic natural perception in the α and β frequency bands. *J Neurophysiol* 128: 1501–1505, 2022. doi:10.1152/jn.00385.2022.
26. **Pfurtscheller G**, **Stancák A Jr**, **Neuper C.** Event-related synchronization (ERS) in the α band—an electrophysiological correlate of cortical idling: a review. *Int J Psychophysiol* 24: 39–46, 1996. doi:10.1016/S0167-8760(96)00066-9.
27. **Romei V**, **Brodbeck V**, **Michel C**, **Amedi A**, **Pascual-Leone A**, **Thut G.** Spontaneous fluctuations in posterior α-band EEG activity reflect variability in excitability of human visual areas. *Cereb Cortex* 18: 2010–2018, 2008. doi:10.1093/cercor/bhm229.
28. **Jensen O**, **Mazaheri A.** Shaping functional architecture by oscillatory α activity: gating by inhibition. *Front Hum Neurosci* 4: 186, 2010. doi:10.3389/fnhum.2010.00186.
29. **Haegens S**, **Nácher V**, **Luna R**, **Romo R**, **Jensen O.** α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proc Natl Acad Sci USA* 108: 19377–19382, 2011. doi:10.1073/pnas.1117190108.
30. **Clayton MS**, **Yeung N**, **Cohen Kadosh R.** The roles of cortical oscillations in sustained attention. *Trends Cogn Sci* 19: 188–195, 2015. doi:10.1016/j.tics.2015.02.004.
31. **Rosch E.** Principles of categorization. In: *Cognition and Categorization*, edited by Rosch E, Lloyd B. Hillsdale, NJ: Lawrence Erlbaum, 1978, p. 27–48.
32. **Walther DB**, **Caddigan E**, **Fei-Fei L**, **Beck DM.** Natural scene categories revealed in distributed patterns of activity in the human brain. *J Neurosci* 29: 10573–10581, 2009. doi:10.1523/JNEUROSCI.0559-09.2009.
33. **Proklova D**, **Kaiser D**, **Peelen MV.** Disentangling representations of object shape and object category in human visual cortex: the animate–inanimate distinction. *J Cogn Neurosci* 28: 680–692, 2016. doi:10.1162/jocn_a_00924.
34. **Coggan DD**, **Baker DH**, **Andrews TJ.** Selectivity for mid-level properties of faces and places in the fusiform face area and parahippocampal place area. *Eur J Neurosci* 49: 1587–1596, 2019. doi:10.1111/ejn.14327.
35. **Coggan DD**, **Watson DM**, **Wang A**, **Brownbridge R**, **Ellis C**, **Jones K**, **Kilroy C**, **Andrews TJ.** The representation of shape and texture in category-selective regions of ventral-temporal cortex. *Eur J Neurosci* 56: 4107–4120, 2022. doi:10.1111/ejn.15737.
36. **Robert S**, **Ungerleider LG**, **Vaziri-Pashkam M.** Disentangling object category representations driven by dynamic and static visual input. *J Neurosci* 43: 621–634, 2023. doi:10.1523/JNEUROSCI.0371-22.2022.
37. **Pitts MA**, **Padwal J**, **Fennelly D**, **Martínez A**, **Hillyard SA.** Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness. *NeuroImage* 101: 337–350, 2014. doi:10.1016/j.neuroimage.2014.07.024.
38. **Bennett L**, **Melchers B**, **Proppe B.** Curta: A General-Purpose High-Performance Computer at ZEDAT, Freie Universität Berlin (Online). Freie Universität Berlin, 2020. https://refubium.fu-berlin.de/handle/fub188/26993 [2023 March 17].

5.4 Author contributions

Declaration pursuant to Sec. 7 (3), fourth sentence, of the Doctoral Study Regulations regarding my own share of the submitted scientific or scholarly work that has been published or is intended for publication within the scope of my publication-based work

I. Last name, first name: Chen, Lixiang
Institute: Department of Education and Psychology, Freie Universität Berlin
Doctoral study subject: Psychology
Title: How the brain uses real-world structure to generate coherent visual experiences

II. Numbered listing of works submitted (title, authors, where and when published and/or submitted):

1. Chen, L., Cichy, R. M.*, & Kaiser, D.* (2022). Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations. Published in Cerebral Cortex.

2. Chen, L., Cichy, R. M.*, & Kaiser, D.* (2023). Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision. Published in Science Advances.

3. Chen, L., Cichy, R. M., & Kaiser, D. (2024). Coherent categorical information triggers integration-related alpha dynamics. Published in Journal of Neurophysiology.

*The authors contributed equally.

III. Explanation of own share of these works:

The contribution is assessed on the scale: all, vast majority, most, part.

Regarding II. 1.: conceptualization and design (part), programming of paradigm (all), data collection (all), data analysis (all), discussion of results (most), writing/revising the manuscript (vast majority)

Regarding II. 2.: conceptualization and design (part), programming of paradigm (vast majority), data collection (most), data analysis (all), discussion of results (most), writing/revising the manuscript (most)

Regarding II. 3.: conceptualization and design (most), programming of paradigm (all), data collection (all), data analysis (all), discussion of results (vast majority), writing/revising the manuscript (vast majority)

IV. Names, addresses, and e-mail addresses or fax numbers for the relevant co-authors:

Regarding II. 1.: Chen, L. (1)
Cichy, R. M. (1)
Kaiser, D. (2, 3)

Regarding II. 2.: Chen, L. (1)
Cichy, R. M. (1)
Kaiser, D. (2, 3)

Regarding II. 3.: Chen, L. (1, 2)
Cichy, R. M. (1)
Kaiser, D. (2, 3)

(1) Department of Education and Psychology, Freie Universität Berlin, Berlin 14195, Germany
(2) Mathematical Institute, Justus-Liebig-University Gießen, Gießen 35392, Germany
(3) Center for Mind, Brain and Behavior (CMBB), Philipps-Universität Marburg and Justus-Liebig-Universität Gießen, Marburg 35032, Germany

Date, doctoral candidate signature . . . . . .

The information in III. must be confirmed in writing by the co-authors. I confirm the declaration made by Lixiang Chen under III.:

Name: Radoslaw Martin Cichy, Signature: . . . . . .

Name: Daniel Kaiser, Signature: . . . . . .

5.5 Selbstständigkeitserklärung

Hiermit erkläre ich,

- dass ich die vorliegende Arbeit eigenständig und ohne unerlaubte Hilfe verfasst habe,
- dass Ideen und Gedanken aus Arbeiten anderer entsprechend gekennzeichnet wurden,
- dass ich mich nicht bereits anderwärtig um einen Doktorgrad beworben habe und keinen Doktorgrad in dem Promotionsfach Psychologie besitze, sowie
- dass ich die zugrundeliegende Promotionsordnung vom 08.08.2016 anerkenne.

Berlin, 03.06.2024

Lixiang Chen