Real-world structure facilitates the rapid emergence of scene category information in visual brain signals

Daniel Kaiser[1], Greta Häberle[2,3,4], Radoslaw M. Cichy[2,3,4,5]

[1]*Department of Psychology, University of York, York, UK*

[2]*Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany*

[3]*Einstein Center for Neurosciences Berlin, Berlin, Germany*

[4]*Berlin School of Mind and Brain, Humboldt-Universität Berlin, Berlin, Germany*

[5]*Bernstein Center for Computational Neuroscience Berlin, Berlin, Germany*

Correspondence:

Dr. Daniel Kaiser

Department of Psychology

University of York

Heslington, York

YO10 5DD, UK

danielkaiser.net@gmail.com

Abstract

In everyday life, our visual surroundings are not arranged randomly, but structured in predictable ways. Although previous studies have shown that the visual system is sensitive to such structural regularities, it remains unclear whether the presence of an intact structure in a scene also facilitates the cortical analysis of the scene's categorical content. To address this question, we conducted an EEG experiment during which participants viewed natural scene images that were either "intact" (with their quadrants arranged in typical positions) or "jumbled" (with their quadrants arranged into atypical positions). We then used multivariate pattern analysis to decode the scenes' category from the EEG signals (e.g., whether the participant had seen a church or a supermarket). The category of intact scenes could be decoded rapidly within the first 100ms of visual processing. Critically, after within 200ms of processing category decoding was more pronounced for the intact scenes compared to the jumbled scenes, suggesting that the presence of real-world structure facilitates the extraction of scene category information. No such effect was found when the scenes were presented upside-down, indicating that the facilitation of neural category information is indeed linked to a scene's adherence to typical real-world structure, rather than to differences in visual features between intact and jumbled scenes. Our results demonstrate that early stages of categorical analysis in the visual system exhibit tuning to the structure of the world that may facilitate the rapid extraction of behaviorally relevant information from rich natural environments.

Introduction

In everyday situations, the input to our visual system is not random. It rather arises from structured scenes, in which information follows meaningful distributions across space (Bar 2004; Kaiser et al., 2019a; Oliva & Torralba, 2007; Võ et al., 2019; Wolfe et al., 2011). Neuroimaging studies have shown that the visual system is sensitive to this structure, with cortical responses differing when scene elements do or do not adhere to typical real-world structure (Abassi & Papeo, 2019; Baldassano et al., 2017; Bilalic et al., 2019; Kaiser et al., 2014; 2019b; Kaiser & Peelen, 2018; Kim & Biederman, 2011; Roberts & Humphreys, 2010). Although such studies suggest that the presence of real-world structure aids efficient scene representation, it is unclear how real-world structure impacts the representation of scene content: Specifically, does the presence of real-world structure facilitate the extraction of categorical information from a scene?

Evidence for an increase of visual category information in the presence of real-world regularities has already been reported for individual object processing. Several studies showed that typical real-world positioning enhances the neural representation of object category (Chan et al., 2010; de Haas et al., 2016; Kaiser & Cichy, 2018; Kaiser et al., 2018): for example, neural responses to an airplane are better discriminable from responses to other objects when the airplane is shown in the upper visual field, where it is typically encountered in the real world. Does the presence of real-world structure similarly facilitate the representation of categorical scene content in scenes?

To address this question, we used a jumbling paradigm (Biederman, 1972; Biederman et al., 1974) that manipulates natural scenes' spatial structure: individual parts of the scene could either appear in their typical, "intact" positions or in atypical, "jumbled" positions (Figure 1). In a recent neuroimaging study (Kaiser et al., 2020a), we employed this paradigm to show that in scene-selective visual cortex (fMRI) and after 250ms of vision (EEG), spatially intact scenes were represented differently from jumbled scenes. Here, we analyzed the EEG data from this jumbling paradigm to investigate whether the typical real-world structure, in contrast to an atypical jumbled structure, facilitates the visual representation of scene category.

To extract differences in category information between intact and jumbled scenes with high sensitivity, we used a cumulative multivariate decoding approach (Ramkumar et al., 2013), which maximizes the amount of data available at every time point along the processing cascade. In line with previous reports (Dima et al., 2018; Kaiser et al., 2019c, 2020b; Lowe et al., 2018), this analysis showed that scene category information emerges rapidly, within the first 100ms of vision. Critically, within 200ms the emergence of scene category information was facilitated for intact compared to jumbled scenes. This benefit was only present for upright, but not inverted scenes, indicating that the early facilitation of scene analysis is related to the presence of real-world structure, rather than differences in basic visual features.

Materials and Methods

Participants

Twenty healthy adults (mean age 26.6 years, *SD*=5.8; 9 female) participated. All participants had normal or corrected-to-normal vision. Participants provided informed consent and received either monetary reimbursement or course credits. All procedures were approved by the ethical committee of the Department of Psychology at Freie Universität Berlin and were in accordance with the Declaration of Helsinki.

Stimuli

Stimuli were scenes from four different categories: churches, houses, roads, and supermarkets (Figure 1a). The stimuli were taken from an online resource (Konkle, Brady, Alvarez, & Oliva, 2010). For each category six different exemplars were used. To manipulate scenes' adherence to real-world structure, we first split each original image into quadrants. We then systematically recombined quadrants from different scenes such that the scenes' spatial structure was either intact or jumbled (Figure 1b). For the intact scenes, four fragments from four different scenes of the same scene category were combined in their correct spatial locations. For the jumbled scenes, four fragments from four different scenes of the same scene category were combined, but their spatial locations were arranged in a crisscrossed way. Additionally, the stimulus set entailed scenes that were jumbled in their categorical content (with the individual scene parts stemming from different categories); these scenes were created to answer a different research question (see Kaiser et al., 2020a) and not used in the analyses reported in this paper. In both conditions relevant for

this paper, we used fragments from four different scenes to equate the presence of visual discontinuities between fragments. Separately for each participant, 24 unique intact and 24 unique jumbled stimuli were generated by randomly drawing suitable fragments from different scenes. Each scene was presented upright and upside-down.

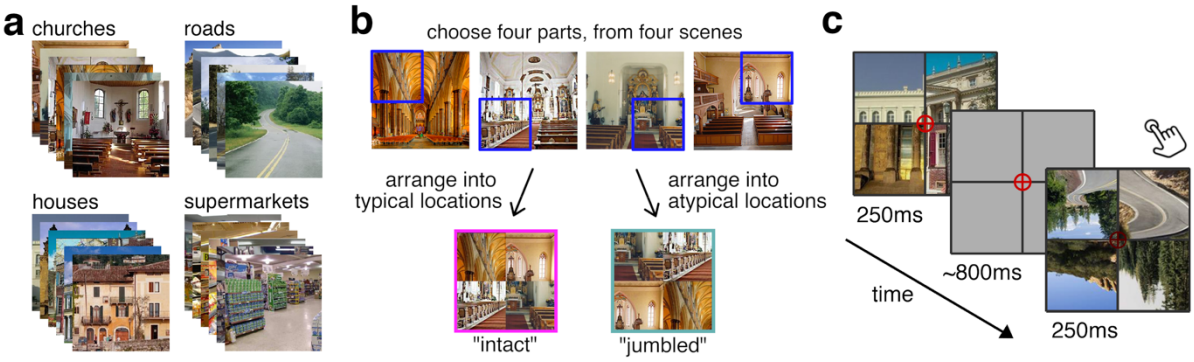

Figure 1. Experimental design. a) The stimulus set was constructed from natural scene photographs of four categories. b) Intact and jumbled scenes were created by combining parts of four different scenes of the same category in either typical locations or in locations (with positions swapped in a crisscrossed way). c) During the EEG experiment, participants viewed the scenes in upright and inverted orientation for 250ms each in random order. Participants performed an orthogonal task, where they responded whenever the fixation cross darkened.

Paradigm

During the EEG experiment, the different stimuli were randomly intermixed within a single session. Within each trial, a scene appeared for 250ms. Stimuli appeared in a black grid (4.5deg visual angle), which served to mask visual discontinuities between quadrants (Figure 1c). Each trial was followed by an inter-trial interval which randomly varied between 700ms and 900ms. For this paper, only parts of the collected data – spatially intact and spatially jumbled scenes in upright and upside-down orientation

– were analyzed. Each of these four conditions covered 384 trials (equating to 48 trials per scene category). Additionally, 1,152 target trials were measured. During the target trials the crosshair changed into a slightly darker red during the scene presentation. When detecting a target, participants had to blink and press a button (mean accuracy 78.1%, *SE*=3.6%). Target trials were not included in the subsequent analyses. Furthermore, 1,536 trials where the scene's categorical structure was altered have been recorded. This data has been analyzed elsewhere (see Kaiser et al., 2020a). Further, participants were instructed to maintain central fixation throughout the experiment. Stimulus presentation was controlled using the Psychtoolbox (Brainard, 1997).

EEG recording and preprocessing

The EEG data were the same as in Kaiser et al. (2020a). EEG signals were recorded using an EASYCAP 64-electrode system and a Brainvision actiCHamp amplifier. For two participants, due to technical problems, only data from 32 electrodes was recorded. Electrodes were arranged in accordance with the 10-10 system. EEG data was recorded at 1000Hz sampling rate and filtered online between 0.03Hz and 100Hz. All electrodes were referenced online to the Fz electrode. Offline preprocessing was performed using FieldTrip (Oostenveld et al., 2011). EEG data were epoched from -200ms to 800ms relative to stimulus onset, and baseline-corrected by subtracting the mean pre-stimulus signal. Channels and trials containing excessive noise were removed based on visual inspection. Blinks and eye movement artifacts were removed using independent components analysis and visual inspection of the

resulting components (Jung et al. 2000). The epoched data were down-sampled to 200Hz.

EEG decoding

Decoding analyses were performed using CoSMoMVPA (Oosterhof et al., 2016). To track cortical representations across time, we used a cumulative classification approach that for each time point across the epoch takes into account all time points prior to the current time point (Ramkumar et al., 2013). This classification technique uses larger amounts of data at each subsequent time point, while maintaining temporal precision in the forward direction (i.e., it only collapses across information backwards in time, but not forwards). Cumulative decoding thus provides increased sensitivity compared to standard timeseries decoding, where the data for each time point are subjected to separate decoding analyses (Grootswagers et al., 2017).

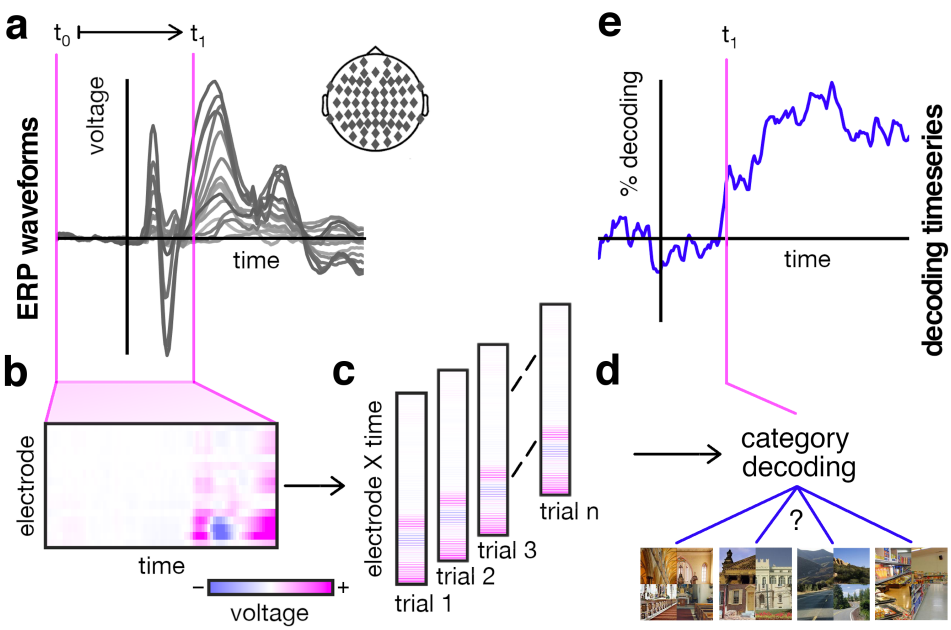

Figure 2. Schematic depiction of the cumulative decoding approach. a) For each time point t_1 across the epoch, a separate decoding analysis was performed. b) For each

of these analyses, we aggregated ERP waveforms across all EEG electrodes and all time points between t_1 and the beginning of the epoch (t_0). c) For each trial, we then unfolded these two-dimensional response pattern across electrodes and time into a one-dimensional response pattern. d) These one-dimensional response patterns were fed to linear classifiers, which were trained to discriminate the four scene categories. Decoding accuracy was computed by repeatedly assessing classifier performance on single trials left out during classifier training. e) Repeating this analysis across time yielded a decoding timeseries with 200Hz resolution. Importantly, the cumulative nature of this analysis allowed us to increase power by increasing the amount of data available to the classifier without losing temporal precision regarding the onset of category information.

We used these cumulative classifiers to discriminate between the four scene categories. This analysis was done separately for the intact and jumbled scenes. Classifiers were trained and tested on trial-wise response patterns across all electrodes in our montage. Classification analyses were performed repeatedly, with the amount of information available to the classifier accumulating across time (Figure 2). That is, for the first time point in the epoch, the classifier was trained and tested on response patterns across the electrodes at this time point. At the second time point in the epoch, the classifier was trained and tested on response patterns across the electrodes at the first and second time point in this epoch. Finally, at the last time point in the epoch, the classifier was trained on response patterns across all electrodes and at all time points in this epoch. To reduce the dimensionality of the data at each time point, we performed principal component analyses (PCAs; see Grootswagers et al., 2017). For each of these PCAs, we retained as many components as needed to explain 99% of the variance in the data. Cross-validation was performed in a leave-one-trial-out scheme. Therefore, classifiers were trained on

all but one trials and tested on the left-out trial. This was done repeatedly until every trial was left out once. Classification accuracies were averaged across these repetitions. The classification analyses resulted in separate decoding timeseries for intact and jumbled scenes, which reflect the temporal accrual of category information (i.e., how well the four categories are discriminable from the neural data).

Statistical testing

To compare decoding timeseries against chance level and the different conditions' decoding timeseries against each other, we used a threshold-free cluster enhancement (TFCE) procedure (Smith & Nichols, 2009). Multiple-comparison correction was based on a sign-permutation test (with null distributions created from 10,000 bootstrapping iterations) as implemented in CoSMoMVPA (Oosterhof et al., 2016). The resulting statistical maps were thresholded at $z>1.96$ (i.e., $p_{corr}<.05$). In an additional analysis, we analyzed the data in discrete time bins of 50ms. Here, we used conventional one-sample t-tests and false-discovery-rate (FDR) corrections to correct for multiple comparisons.

Data availability

Data are publicly available on OSF (doi.org/10.17605/OSF.IO/ECMA4).

Results

We first analyzed data from the upright scenes, where we expected a facilitation of category information for spatially intact, compared to jumbled, scenes. We found that EEG signals conveyed robust scene category information, as cluster-based TFCE tests indicated that categories were discriminable for both intact and jumbled scenes (Figure 3a): The clusters in the observed data extended from 65ms to 800ms for intact scenes and from 130ms to 800ms for jumbled scenes (Figure 3b). Crucially, we found significantly enhanced decoding for the spatially intact scenes, compared to the jumbled scenes, in a cluster between 155ms and 800ms (Figure 3c).

The inclusion of inverted scenes allowed us to investigate whether the effects of scene structure were genuinely related to the scenes adhering to real-world structure, rather than differences in their low-level visual attributes. If the enhanced category information for spatially intact scenes is indeed related to their adherence with real-world structure, then no effects should be seen when the same scenes are viewed upside-down, as all inverted scenes do not adhere to real-world structure.

Performing the category decoding analysis on the inverted scenes (Figure 3d) indeed revealed a qualitative difference to the upright scenes. Although the category of both intact scenes (in a cluster between 65ms and 800ms) and jumbled scenes (in a cluster between 70ms and 800ms) could be decoded from the EEG signals (Figure 3e), no significant differences between the intact and jumbled scenes were observed (Figure 3f). This indicates that the facilitation of scene category information for spatially structured scenes can be attributed to the scenes adhering to typical real-

world structure, rather than to low-level features differing between the intact and jumbled scenes.

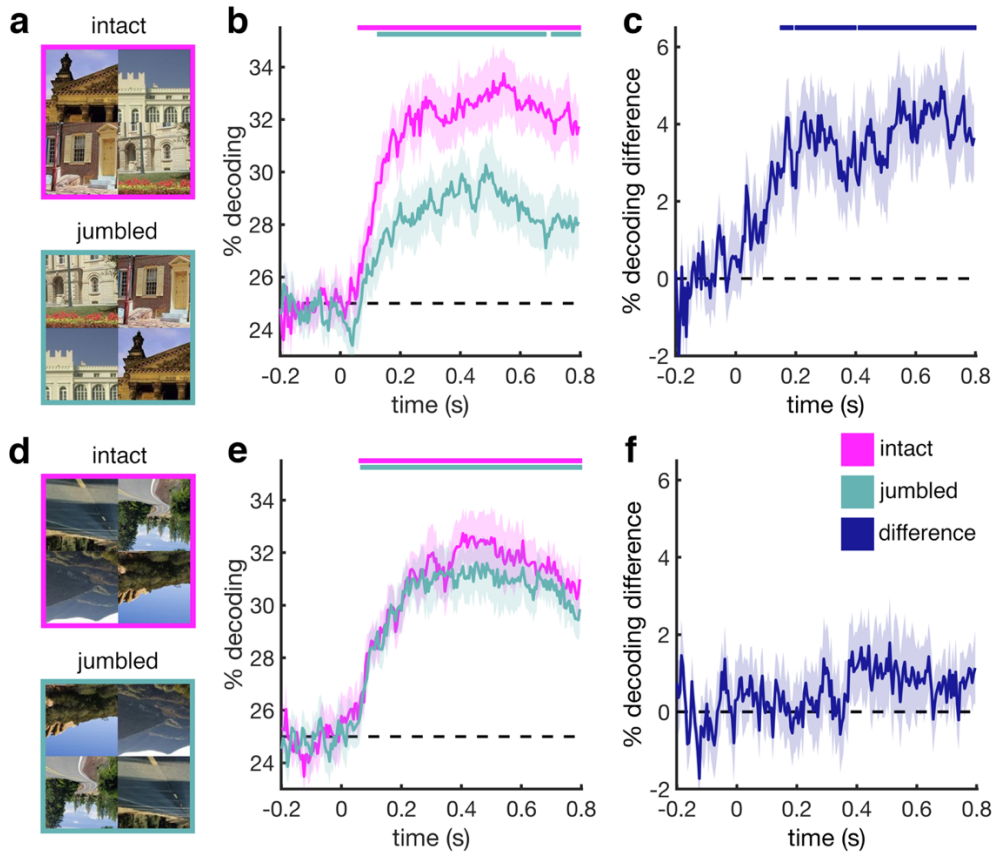

Figure 3. EEG decoding results. a) Firstly, we decoded the category of intact and jumbled scenes when they were presented upright. b) This analysis revealed widespread clusters of category decoding for both intact and jumbled scenes. c) Critically, in a cluster between 155ms and 800ms, we found more accurate decoding of scene category when the scene was intact, suggesting that adherence to real-world structure boosts visual category information. d) Secondly, we decoded the category of upside-down scenes. e) For upside-down scenes, category could be similarly decoded from the EEG signals. f) However, there was no benefit of intact scene structure when the scenes were inverted, suggesting that adherence to real-world structure, rather than low-level differences, explain the enhanced category decoding for structured scenes when they are upright. Error margins indicate standard errors of the mean. Significance markers indicate $p<0.05$, corrected for multiple comparisons.

Together, these results suggest a relatively early facilitation of category information when scenes adhere to real world structure. However, the onset of statistical significance for the TFCE method used may be biased by the presence of strong clusters following the onset and can therefore not be directly interpreted (Sassenhagen & Draschkow, 2019). To more precisely establish the emergence of effects in time, we performed an analysis where we averaged data within discrete time bins of 50ms. For each time bin separately, we then used FDR-corrected t-tests to establish differences between the intact and jumbled scenes.

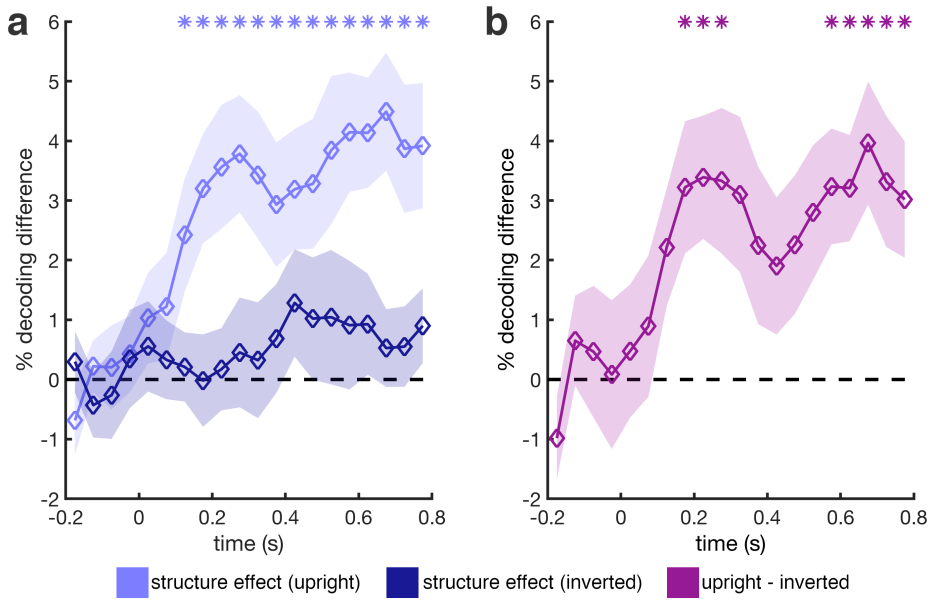

Figure 4. Effects of scene structure on category decoding for discrete time bins. a) When aggregating data within discrete 50ms bins, we found a strong effect of scene structure on category decoding for upright scenes, which started in the 100ms-150ms time bin; no significant effects were observed for the inverted scenes. b) Directly comparing the effects of scene structure for upright and inverted scenes revealed significantly stronger effects in the upright condition, which first emerged in the 150ms-200ms time bin. Error margins indicate standard errors of the mean. Significance markers indicate $p<0.05$, corrected for multiple comparisons.

This analysis provided evidence for an early enhancement of scene category information for typically structured scenes: already in the time bin from 100ms to 150ms ($t[19]=2.6$, $p_{FDR}=0.022$), category information was more pronounced for intact scenes than for jumbled scenes. This was not the case when the scenes were presented upside-down (Figure 4a). Further, when directly testing for differences between the upright and inverted scenes, we found that category information was more strongly enhanced for the upright, compared to the inverted, scenes (Figure 4b); this difference first reached significance between 150ms and 200ms ($t[19]=2.9$, $p_{FDR}=0.021$). This result demonstrates that adherence to real-world structure, rather than low-level differences between the upright and inverted scenes, enhances scene category information within the first 200ms of visual analysis.

Discussion

Our results provide evidence that real-world regularities facilitate the extraction of scene category information during visual analysis. Using a cumulative decoding technique that is particularly sensitive for establishing differences in the initial emergence of information in EEG signals, we show that this facilitation of category information emerges within the first 200ms of vision. Our findings highlight the pervasive role of real-world structure in perceptual processing, suggesting that already at relatively early processing stages cortical scene representations are tightly linked to the typical composition of our daily surroundings.

The early facilitation of category information is consistent with results from single-object processing, where representations of individual objects are rapidly enhanced – within the first 150ms of vision – when the objects appear in their typical real-world locations, such as an eye in the upper visual field (Issa & DiCarlo, 2012) or a shoe in the lower visual field (Kaiser et al., 2018). Together, these findings therefore support the idea that real-world structure can boost basic visual analysis across diverse stimuli and processing levels (Kaiser et al., 2019a).

While our effects demonstrate an enhanced early representation of scenes that adhere to real-world structure, studies on object-scene consistency suggest that EEG waveforms only become sensitive to typical object positioning after around 250ms of vision (Coco et al., 2019; Draschkow et al., 2018; Ganis & Kutas, 2003; Mudrik et al., 2010, 2014; Võ & Wolfe, 2013). How do these early and late effect of scene structure relate to each other?

As one possibility, later effects may partly reflect increased responses to inconsistencies, rather than an enhanced processing of consistent scene-object combinations (Faivre et al., 2019). Together with our results, these findings may suggest that early responses are biased towards scenes that predictably follow real-world structure, whereas later responses may be more biased towards violations of this structure. This idea is consistent with a recent proposal in predictive processing, which suggests a temporal succession of more general processing biases, first towards the expected and then towards the surprising (Press et al., 2020).

Alternatively, the beneficial effects of real-world regularities may not immediately result in consistency signals: Whether visual inputs generally are consistent with our real-world experience may only be analyzed following more basic visual analysis. Supporting this idea, generic consistency signals in our data only emerge later than the enhanced category processing: As previously reported, intact and jumbled scenes (independent of their category) evoked reliably different responses only after 255ms of processing (Kaiser et al., 2020a).

More broadly, the finding that scene structure impacts early visual analysis adds to our understanding of efficient everyday vision. Even under challenging real-world conditions, human vision is remarkably efficient – in fact, much more efficient than findings from simplified laboratory experiments would predict (Wolfe et al., 2011; Peelen & Kastner, 2014). Our findings shed new light on this efficiency: Even early scene representations exhibit a tuning for real-world structure, and this early tuning may be a crucial asset in complex real-life situations, where we need to quickly make sense of our surroundings to perform adequate actions.

Acknowledgements

We thank Sina Schwarze for help in EEG data collection. D.K. and R.M.C. are supported by Deutsche Forschungsgemeinschaft (DFG) grants (KA4683/2-1, CI241/1-1, CI241/3-1). R.M.C. is supported by a European Research Council Starting Grant (ERC-2018-StG 803370). G.H. is supported by a PhD fellowship of the Einstein Center for Neurosciences Berlin.

References

Abassi E, Papeo L. (2020). The representation of two-body shapes in the human visual cortex. *J Neurosci, 40,* 852-563.

Baldassano C, Beck DM, Fei-Fei L. (2017). Human-object interactions are more than the sum of their parts. *Cereb Cortex, 27*, 2276-2288.

Bar M. (2004). Visual objects in context. *Nat Neurosci, 5*, 617-629.

Biederman I. (1972). Perceiving real-world scenes. *Science, 177*, 77-80.

Biederman I, Rabinowitz JC, Glass AL, Stacy EW. (1974). On the information extracted from a glance at a scene. *J Exp Psychol, 103*, 597-600.

Bilalic M, Lindig T, Turella L. (2019). Parsing rooms: the role of the PPA and RSC in perceiving object relations and spatial layout. *Brain Struct Funct, 224*, 2505-2524.

Brainard DH. (1997). The psychophysics toolbox. *Spat Vis, 10*, 433-436.

Chan AW, Kravitz DJ, Truong S, Arizpe J, Baker CI. (2010). Cortical representations of bodies and faces are strongest in commonly experienced configurations. *Nat Neurosci, 13,* 417-418.

Coco MI, Nuthmann A, Dimigen O. (2019). Fixation-related brain potentials during semantic integration of object-scene information. *J Cogn Neurosci*, doi.org/10.1162/jocn_a_01504.

de Haas B, Schwarzkopf DS, Alvarez I, Lawson RP, Henriksson L, Kriegeskorte N, Rees G. (2016). Perception and processing of faces in the human brain is tuned to typical facial feature locations. *J Neurosci, 36,* 9289-9302.

Dima DC, Perry G, Singh KD. (2018). Spatial frequency supports the emergence of categorical representations in visual cortex during natural scene perception. *Neuroimage, 179*, 102-116.

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

Faivre N, Dubois J, Schwartz N, Mudrik L. (2019). Imaging object-scene relations processing in visible and invisible natural scenes. *Sci Rep, 9*, 4567.

Ganis G, Kutas M. (2003). An electrophysiological study of scene effects on object identification. *Brain Res 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.

Issa B, DiCarlo JJ. (2012). Precedence of the eye region in neural processing of faces. *J Neurosci, 32*, 16666-16682.

Jung TP, Makeig S, Humphries C, Lee TW, McKeown MC, Iragui V, Sejnowski TJ. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*, 163-178.

Kaiser D, Cichy RM. (2018). Typical visual-field locations enhance processing in object-selective channels of human occipital cortex. *J Neurophysiol*, *120*, 848-853.

Kaiser D, Häberle G, Cichy RM. (2020a) Cortical sensitivity to natural scene structure. *Hum Brain Mapp, 41*, 1286-1295.

Kaiser D, Moeskops MM, Cichy RM. (2018) Typical retinotopic locations impact the time course of object coding. *NeuroImage, 176*, 372-379.

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

Kaiser D, Stein T, Peelen MV. (2014). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proc Natl Acad Sci USA, 111*, 11217-11222.

Kaiser D, Peelen MV. (2018) Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *NeuroImage 169*, 334-341.

Kaiser D, Inciuraite G, Cichy RM. (2020b). Rapid contextualization of fragmented scene information in the human visual system. *bioRxiv,* doi.org/10.1101/2020.01.06.895870.

Kaiser D, Turini J, Cichy RM. (2019b). A neural mechanism for contextualizing fragmented inputs during naturalistic vision. *eLife, 8*, e48182.

Kim JG, Biederman I. (2011). Where do objects become scenes? *Cereb Cortex, 21*, 1738-1746.

Konkle T, Brady TF, Alvarez GA, Oliva A. (2010). Scene memory is more detailed than you think: the role of categories in visual long-term memory. *Psychol Sci, 21*, 1551-1556.

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 sensory processing: replication and extension. *Neuropsychologia, 56*, 447-458.

Oliva A, Torralba A. (2007). The role of context in object recognition. *Trends Cogn Sci, 11*, 520-527.

Oostenveld R, Fries P, Maris E, Schoffelen JM. (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,* 20.

Peelen MV, Kastner S. (2014). Attention in the real world: toward understanding its neural basis. *Trends Cogn Sci, 18*, 242-250.

Press C, Kok P, Yon D. (2020). The perceptual prediction paradox. *Trends Cogn Sci, 24*, 13-24.

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.

Roberts KL, Humphreys GW. (2010). Action relationships concatenate representations of separate objects in the ventral visual cortex. *NeuroImage 52*, 1541-1548.

Sassenhagen, J, Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or onset. *Psychophysiology, 56*, e13335.

Smith SM, Nichols TE. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage, 44,* 83-98.

Võ MLH, Boettcher SEP, Draschkow D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Curr Opin Psychol, 29*, 205-210.

Võ MLH, Wolfe JM. (2013). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychol Sci, 24*, 1816-1823.

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