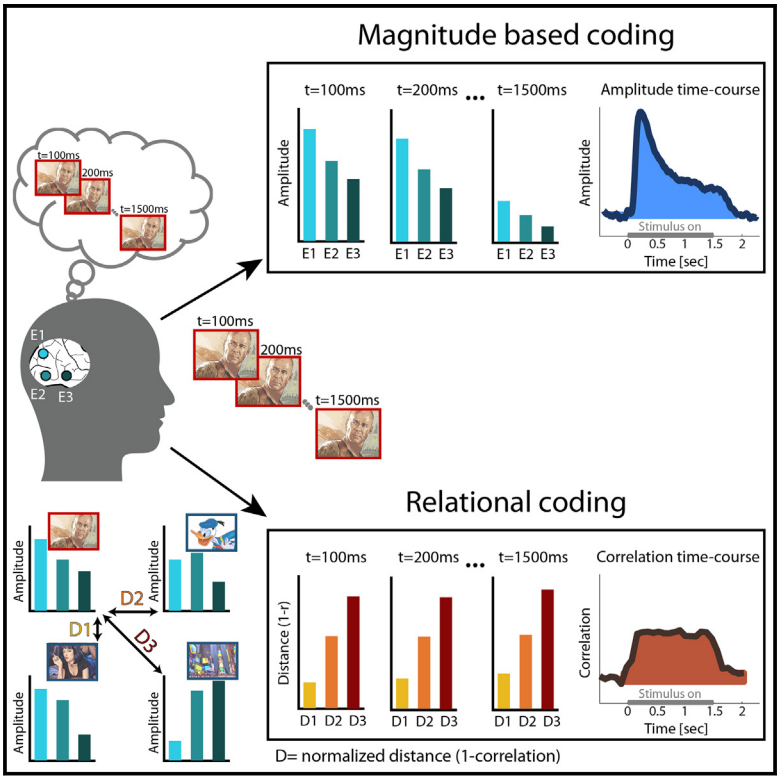
Cell Reports

Article

Perceptual stability reflected in neuronal pattern similarities in human visual cortex

Graphical abstract



Authors

Rotem Broday-Dvir, Yitzhak Norman, Michal Harel, Ashesh D. Mehta, Rafael Malach

Correspondence

rafi.malach@gmail.com

In brief

Neuronal activation magnitude is considered the key factor for visual perception yet shows a dramatic drop during prolonged visual stimulation while the visual experience remains stable. Broday-Dvir et al. find that relative population activity patterns and their relational geometries are sustained across time in high-order human visual cortex.

Highlights

- iEEG recordings in 13 patients during sustained (1.5 s) viewing of visual stimuli
- Neuronal response amplitude drops rapidly, while visual perception is maintained
- Relational population activity patterns remain stable across time
- The effects differ across the visual hierarchy

Broday-Dvir et al., 2023, Cell Reports *42*, 112614
June 27, 2023
https://doi.org/10.1016/j.celrep.2023.112614

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Perceptual stability reflected in neuronal pattern similarities in human visual cortex

Rotem Broday-Dvir,[1] Yitzhak Norman,[2] Michal Harel,[1] Ashesh D. Mehta,[3] and Rafael Malach[1,4,*]
[1]Department of Brain Sciences, Weizmann Institute of Science, Rehovot 76100, Israel
[2]Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA 94143, USA
[3]Department of Neurosurgery, Zucker School of Medicine at Hofstra/Northwell, and the Feinstein Institute for Medical Research, Manhasset, NY 11030, USA
[4]Lead contact
*Correspondence: rafi.malach@gmail.com
https://doi.org/10.1016/j.celrep.2023.112614

SUMMARY

The magnitude of neuronal activation is commonly considered a critical factor for conscious perception of visual content. However, this dogma contrasts with the phenomenon of rapid adaptation, in which the magnitude of neuronal activation drops dramatically in a rapid manner while the visual stimulus and the conscious experience it elicits remain stable. Here, we report that the profiles of multi-site activation patterns and their relational geometry—i.e., the similarity distances between activation patterns, as revealed using intracranial electroencephalographic (iEEG) recordings—are sustained during extended visual stimulation despite the major magnitude decrease. These results are compatible with the hypothesis that conscious perceptual content is associated with the neuronal pattern profiles and their similarity distances, rather than the overall activation magnitude, in human visual cortex.

INTRODUCTION

What is the neuronal code that underlies perceptual experiences in the human brain? A dominant theme supported by numerous studies, involving bistable illusions, binocular rivalry, backward masking, or attentional shifts,[1–7] as well as the study of spontaneous blinks,[8] is the notion that perception is associated with a rapidly increased magnitude of neuronal activity.

However, a puzzling contradiction to these converging lines of evidence is presented by the finding of a robust and rapid neuronal adaptation effect. Rapid adaptation refers to a dramatic reduction in the magnitude of neuronal activity in the visual cortex despite a constant visual input and its accompanied stable perceptual experience. These rapid adaptation effects appear a few hundreds of milliseconds after stimulus onset,[9,10] have been documented across the visual cortex,[8,9,11–13] and are unaffected by the stimulus presentation duration under laboratory[12] or ecological, free-viewing, conditions.[14] They likely share common mechanisms with the phenomena of visual adaptation and repetition suppression.[15–17]

The rapid adaptation phenomenon poses an obvious conundrum to activation magnitude-based theories of perception. Simply put, if the magnitude of neuronal activity determines perceptual awareness, how does perception remain stable despite this massive reduction? Complementarily, one may view this dissociation as an opportunity to uncover the neuronal codes that do remain stable in tandem with the stable nature of perception.

Interestingly, recently, the question of the neuronal correlates of perceptual stability has been identified as a key experimental test in an adversarial collaboration, proposed to form a differential prediction of two theories of conscious experience—the global workspace theory (GWT) and the integrated information theory (IIT).[18]

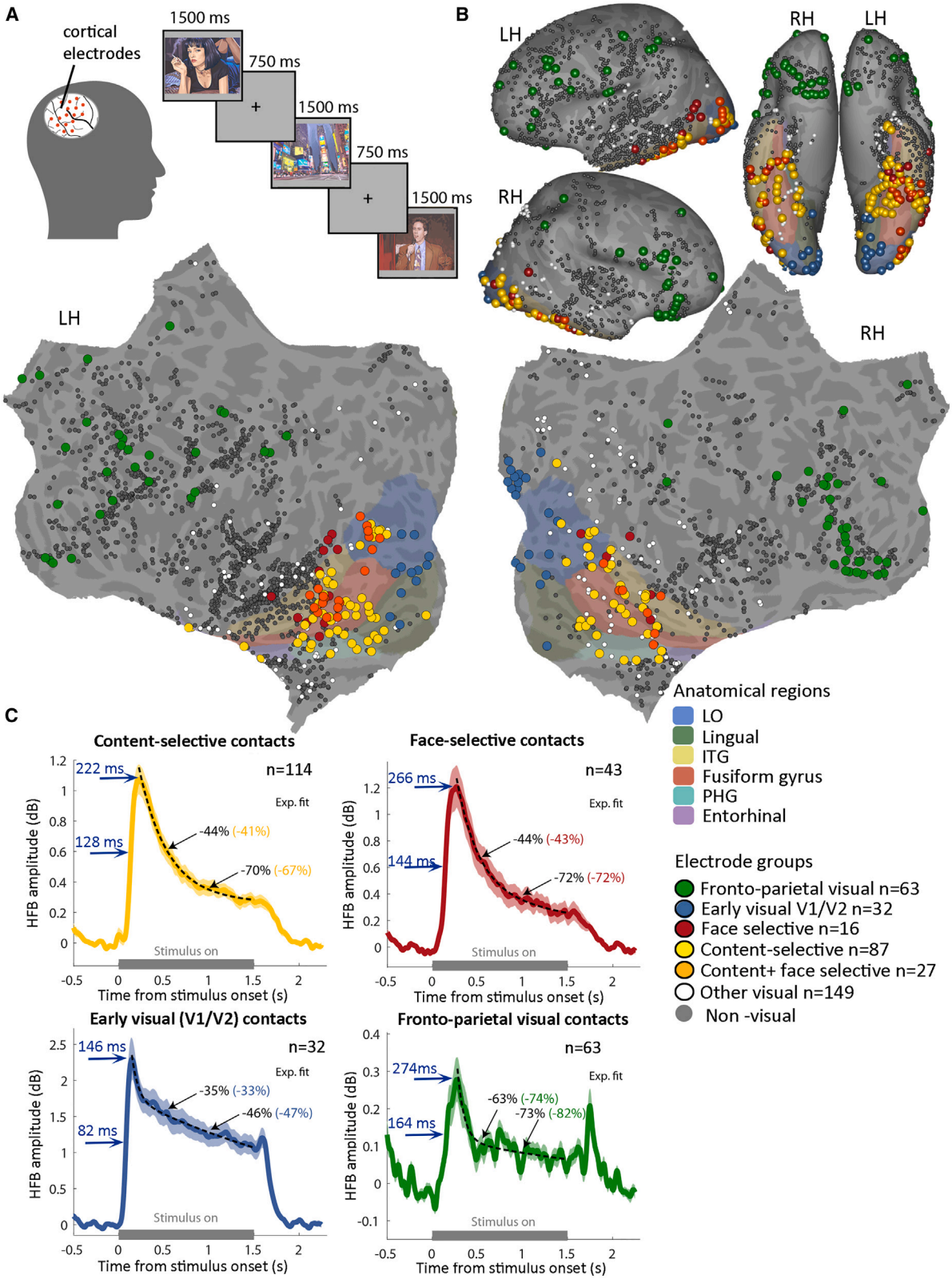
If magnitude is not the factor underlying perceptual stability, then what could be the alternatives? An attractive hypothesis, inspired by classical structuralist perspectives, proposes that activation patterns' profiles (orientation of the population vectors), and specifically their similarity relationships, may constitute such a code.[19–30] In the rest of the text, we will term these similarity distances the relational code.

Here, we set out to examine whether these pattern profiles and their relational codes remain stable across stimulus presentation time, thus enabling perceptual stability. Our results also allowed us to test the predictions made in the adversarial collaboration[18] and hence contribute experimental evidence relevant to different theories of consciousness in the brain.

RESULTS

The study is based on intracranial electroencephalographic (iEEG) recordings conducted for clinical purposes in 13 patients implanted with 2,571 electrode contacts (see Table S1 for details; data were collected as part of a previous study[13,31]).

iEEG recordings were taken while patients performed a simple visual task (see Figure 1A and STAR Methods), viewing colored



(legend on next page)

images of 28 familiar faces and places (14 from each category) for 1,500 ms, followed by fixation for 750 ms. Visibility was verified by successful post-session memory recall (see Norman et al.[13,31] and STAR Methods).

For analysis purposes, visually responsive electrode contacts were subdivided into four groups based on anatomical and functional criteria: early visual (n = 32); face-selective (n = 43); content-selective: contacts showing a significant preference for a specific subset of exemplars (n = 114); and fronto-parietal visually responding contacts (n = 63) (for further details, see STAR Methods and Table S1). Figure 1B depicts the location of the iEEG contacts, color-coded according to each group.

As a measure of neuronal activity, we have analyzed the broad-band power of high frequencies of the iEEG signal (HFBP [high frequency broadband power]: 60–160 Hz), which we and others have demonstrated to be tightly linked to average neuronal firing rates within the recording sites.[32–34] Figure 1C depicts the averaged HFB response magnitude in each contact group, revealing the response latency and the response dynamics elicited by the visual stimulation.

Response latency across cortical sites

The latency of visual responses and their delays across different cortical sites is of great theoretical interest.[18] The wide iEEG coverage in our visual task (see Figure 1B) provided a good opportunity to examine this question in the present experimental paradigm. Precise measurements of response latencies for both half-peak and peak response magnitudes (Figure 1C, blue arrows) revealed an activation delay between early visual cortex and high-order face-selective contacts of ~60 ms (half peak), while, interestingly, the half-peak latency of frontal contacts was 164 ms—showing only a ~20 ms delay relative to the face-selective contacts.

Response magnitude changes during sustained stimulation

As can be seen in Figure 1C, all four contact groups displayed a rapid adaptation effect characterized by an early onset response at 200–300 ms, followed by a striking amplitude decline despite the sustained, constant visual stimulation (decreasing in ~70% after 1 s; see Figure 1C for precise percentage decreases at times 0.5 and 1 s after stimulus onset). The adaptation effect was particularly prominent in the high-order visual contacts. Early visual contacts showed a less pronounced decline, while the fronto-parietal visual contacts exhibited a marked adaption but a far weaker signal overall that, interestingly, exhibited an offset response following stimulus termination.

Diversity in response patterns and their inter-stimuli distances

Inspecting the multi-contact population patterns of the high-order content-selective contacts revealed that each visual exemplar was associated with a distinct profile of a multi-contact pattern, i.e., a unique population vector (examples shown in Figure 2). For the convenience of comparison across these patterns, the contacts (n = 114) are arranged in descending order of their activation magnitude in response to an image of Bruce Willis, 300 ms after its onset. This serial order of the contacts was maintained in the three other pattern histograms as well. The electrode colors depict their response magnitude to each specific stimulus, from the strongest response (light blue) to the weakest (black). Two clear aspects can be gleaned from these examples: first, the activation encompassed the entire set of contacts. Second, each visual image elicited a unique profile or population activity vector across the set of contacts, metaphorically resembling a "barcode" of each image. In a similar manner, the relational coding, i.e., the similarity distances between the patterns, also varied, as can be visually appreciated by comparing the similarity of the Bruce Willis elicited pattern with the other three activation profiles, depicted in Figure 2. The similarities between activity-pattern pairs were quantified in the present study by calculating the Pearson correlation between the two population vectors (see STAR Methods). It should be noted that unlike coherence measures, Pearson correlation is sensitive only to the relative activation profiles within the patterns and not to their absolute magnitude levels. The inverse of the pairwise pattern similarity was defined as the neuronal distance (d) between the pair of stimuli-related patterns and was calculated as $1 - \text{Pearson's } r$.

Comparing the distances between the population vector activations to multiple repeats of the same image with the distances between different stimuli revealed a significant difference (unpaired two-tailed t test, $t = 13.2314$, $p < 0.0001$) with a smaller mean distance for within-stimuli repeats of 0.33 (±0.07 SEM) vs. the between-stimuli mean distance of 0.52 (±0.03 SEM).

Stability of population activity patterns across time

We first examined the stability of the stimulus-linked activation patterns (see STAR Methods for details of this analysis) in the

Figure 1. Experimental design, contact locations, and mean HFB responses to image presentations

(A) Patients undergoing intracranial recordings viewed 28 different images of famous faces and places. The stimuli were presented for 1.5 s each, with 750 ms inter-stimulus intervals (ISIs). Face and place stimuli shown in all figures are accurate illustrations of the original images used in the study, in compliance with copyright limitations.

(B) Multi-patient contact coverage, shown on inflated (lateral view: top left; ventral view: top right) and on flat common cortical surfaces. Visually responsive contacts were allocated to the following subgroups: early visual areas V1 and V2 (blue); face selective (red); content-selective electrodes in intermediate visual areas (yellow); contacts that are included both in the face-selective and in the content-selective groups (orange); frontal-parietal contacts showing visual selectivity (green); and other visual contacts (white) (see STAR Methods). Remaining contacts, which did not display visual responses, are marked in gray. RH, right hemisphere; LH, left hemisphere; LO, lateral occipital cortex; ITG, inferior temporal gyrus; PHG, parahippocampal gyrus.

(C) Mean HFB (60–160 Hz) responses time locked to stimuli onset, depicting the average visual response of each contact set. The gray line denotes the stimulus-on duration. Shaded areas represent ± SEM across contacts, and the contact number (n) of each set is denoted in the figure. The black dashed line indicates the two-term exponent fit to the data. Percentage signal change of the response amplitude at times 0.5 and 1 s after stimulus onset, compared with the peak response, are denoted in black for the exp. model fit, and in color for the measured data. Blue horizontal arrows point to the half-peak time and peak time, respectively.

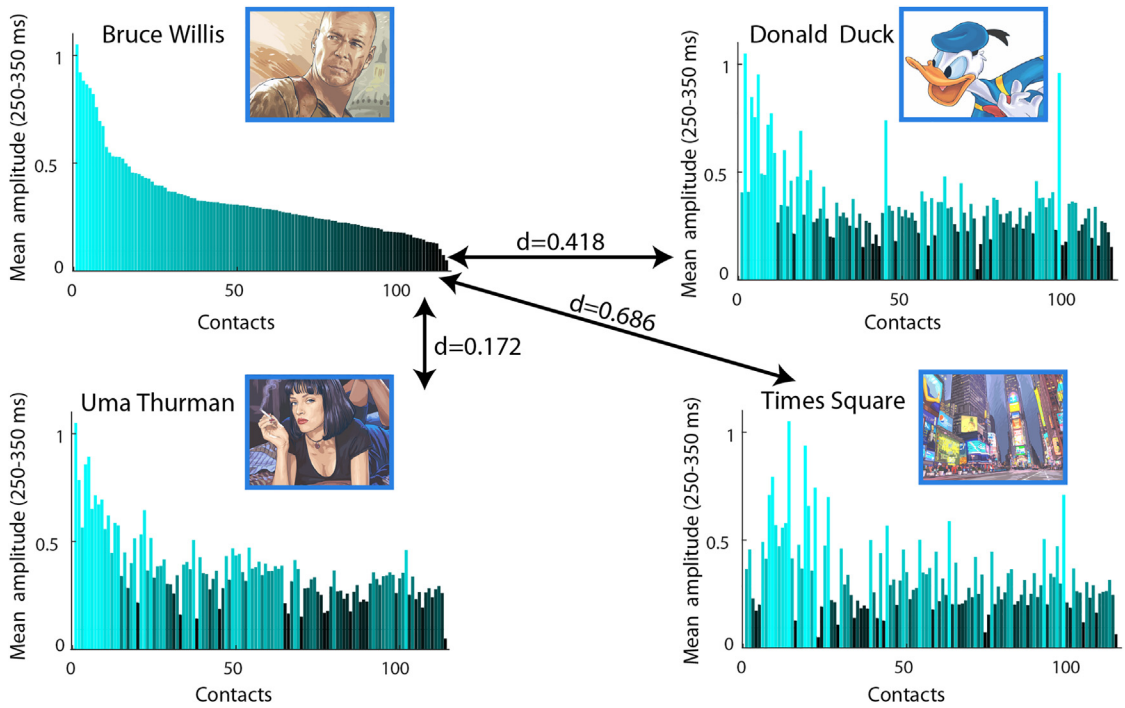


Figure 2. Activity patterns of all visual content-selective contacts (n = 114) to four different stimuli exemplars

Each histogram depicts the normalized HFB response of all contacts to the specific stimulus, averaged across time windows 250–350 ms post-stimulus onset. Contact order is preserved for all four stimuli and is arranged according to amplitude strength in response to the Bruce Willis image. Contact color depicts the amplitude level in response to each stimulus, from the strongest response (light blue) to the weakest (black). Distances (d) between the stimuli-evoked patterns are defined as 1 – Pearson's r and are depicted by the proportional length of the arrows. Face and place stimuli shown in all figures are accurate illustrations of the original images used in the study, in compliance with copyright limitations.

high-order content-selective visual contact group (all results from here on will refer to this set of contacts until noted differently). Figure 3A depicts the cross-time (inter-temporal) distances matrix, defined as 1 – Pearson's r, between the multi-contact activity patterns across all pairs of time bins during the trial (using 100 ms averaged time bins), calculated separately per stimulus, and then averaged across all stimuli (see STAR Methods). Figure 3B shows the across-stimuli average inter-temporal correlation profiles (essentially a horizontal cross-section line of the inverse of the distances matrix shown in Figure 3A), at a few example time points, revealing a significant increase in inter-temporal pattern correlations throughout the entire image presentation duration ($p < 0.005$, using 1,000 randomly shuffled permutations, false discovery rate [FDR] and cluster corrected). The inter-temporal correlations showed a transient increase in the early onset-related time bins, which later subsided into stable, significant correlation levels (see the correlation profile of 0.3 vs. 1.5 s in Figure 3B).

Stability of relational geometry

We next examined the stability of the relational geometry, calculated by comparing the representational dissimilarity matrix (RDM), the matrix containing all stimuli-pattern pairwise distances,[22,23,35] across time (see STAR Methods). Figure 3C shows the inter-temporal distance matrix for the relational geometries across all different time points, while Figure 3D depicts the profiles of the inter-temporal correlation time courses at selected time points. The overall picture largely resembled the stability of the activity patterns—displaying significant and sustained RDM correlations across the entire duration of the stimuli presentations, with a transient increase in the correlations during the early time windows ($p < 0.005$, across 1,000 shuffled permutations, FDR and cluster corrected).

Stability of relational coding

Next, we examined the stability of the relational coding of the stimuli, estimated by calculating the distances of a specific image-related pattern to all other stimuli-elicited patterns in our set, as is illustrated in Figure 2. Figure 4A shows an example of a complete relational code profile for the Bruce Willis image, derived from the content-selective contact group (n = 114). As can be seen, the distances were far from uniform varying over a wide (>3-fold) range. This broad range of pattern similarity distances was a common feature, as can be appreciated from Figure 4B, which depicts the average, sorted, similarity-distance profiles across all stimuli (see STAR Methods). An anticipated category-related effect, exhibiting smaller distances within as compared with across categories, is also clearly evident in Figures 4A and 4B (see also the inset in Figure 4B, depicting the mean distances across all stimuli pairs for within vs. between categories; $p < 0.0001$, two-sample t test).

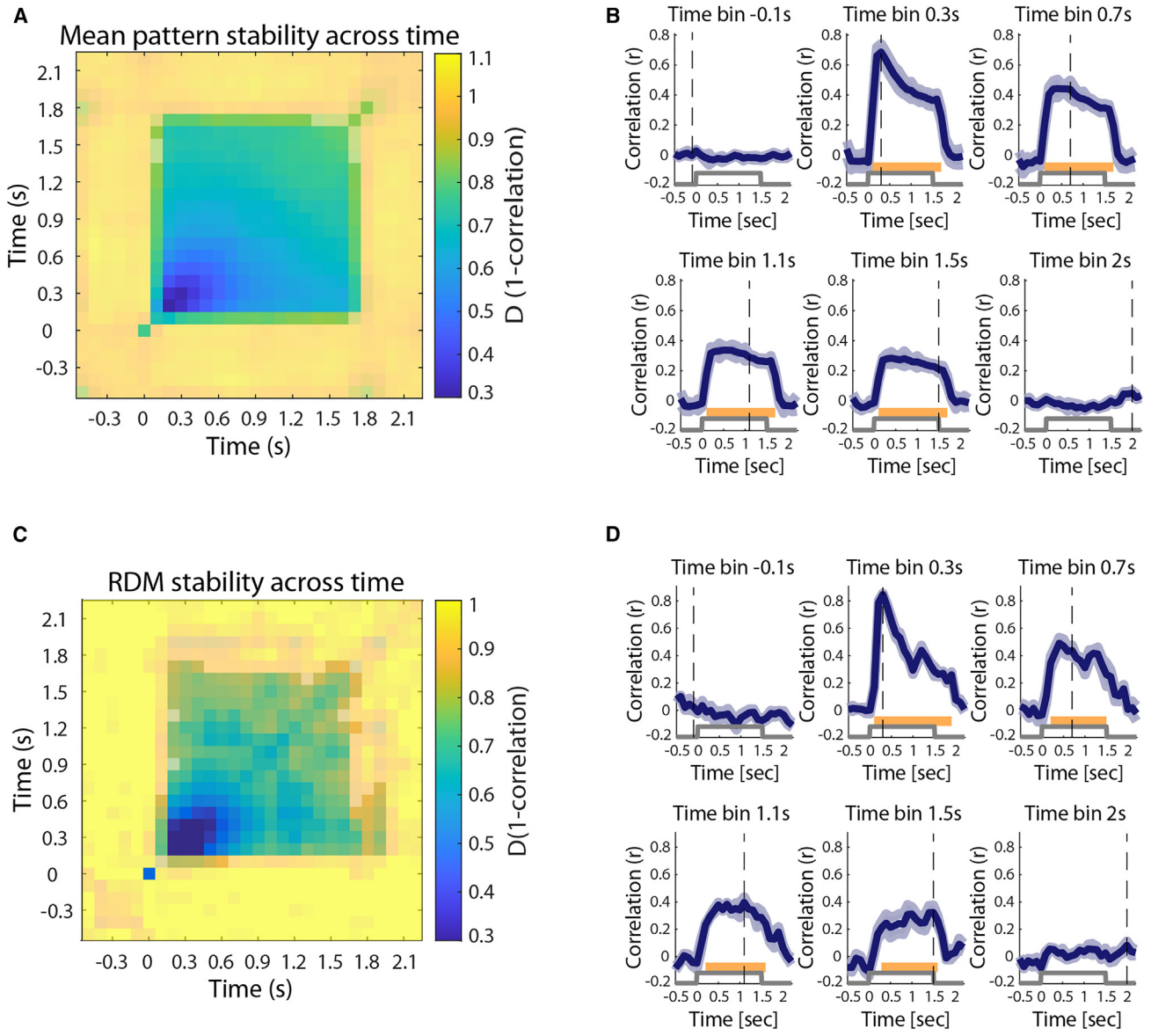


Figure 3. Multi-contact pattern and RDM stability across time in content selective contacts (n = 114)

(A) Mean inter-temporal pattern-distance matrix, exhibiting the distances between stimulus-evoked response patterns for all time bin pairs, using 100 ms time windows. Half-transparent colored regions mark distances that are non-significant, as tested relative to distances emerged from 1,000 random shuffling permutations, while distance values shown in full opaqueness are statistically significant ($p < 0.005$ permutation test, FDR and cluster corrected), displaying pattern stability across stimulus duration.

(B) Example pattern correlation time courses, displaying the Pearson r coefficients between stimuli-related activity patterns at the indicated time bins in each panel vs. all 100 ms time bins between −0.5 and 2.2 s. Gray step plots mark stimulus duration time (0–1.5 s), and the dashed vertical lines indicate the selected time bin. Shaded areas represent ± SEM across the different stimuli (n stimuli = 28). Yellow lines indicate statistical significance at $p < 0.005$, FDR and cluster corrected (permutation test using 1,000 shuffled permutations).

(C) Inter-temporal RDM distance matrix, exhibiting the distances between RDMs calculated separately for each time bin, across all time bin pairs. Statistical significance assessed and denoted as in (A).

(D) Example RDM correlation time courses, depicting the Pearson r coefficients between the RDMs at the indicated time bins in each panel vs. the RDMs from all 100 ms time bins between −0.5 and 2.2 s. Shaded areas represent ± SEM across the leave-one-out iterations (n = 4; see STAR Methods for details). Statistical significance and notations as in (B).

To examine the stability of the relational code, we calculated the across-stimuli mean matrix of the inter-temporal relational code distances, displayed in Figure 4C (see STAR Methods). Figure 4D displays a few selected time point correlation examples. Together, they display a similar dynamic profile to the stimuli-activity patterns and RDM findings shown in Figure 3.

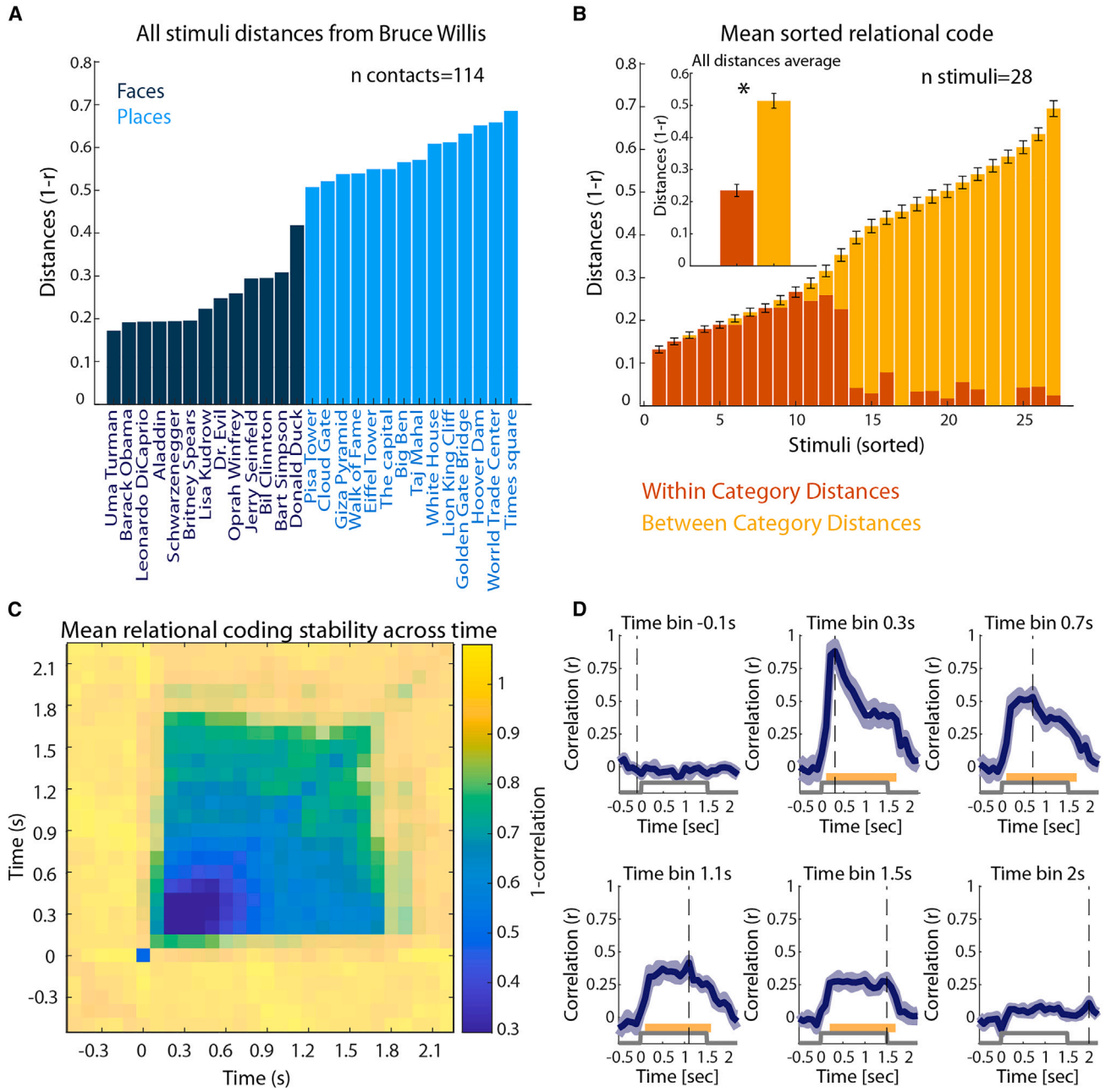


Figure 4. Stability of relational coding across time in content-selective contacts (n = 114)

(A) An example of the relational code of the Bruce Willis image. Histogram bars denote the distances of all other stimuli-linked patterns from the Bruce Willis activity pattern, at time 0.3 s post-stimulus onset.

(B) Sorted relational codes averaged across stimuli (n = 28), error bars represent SEM across the stimuli. For each stimulus, the relational code was calculated and sorted in ascending order, as shown in the example in (A); these sorted relational codes were then averaged across all stimuli, obtaining a mean distances vector, illustrating the reliability of this metric across stimuli. Proportions of within-category stimuli pairs vs. between-category pairs are marked by the relative orange (within category) vs. yellow (between categories) areas of each bar, reflecting that within-category pairs display shorter distances than between-category pairs. This is further illustrated in the inset bar plot, presenting the mean of all within- vs. between-category pair distances (unpaired two-tailed t test, $p < 0.001$, $t = -33.97$). Error bars denote SEM across all stimuli pairwise distances (within-category n = 364, between-category n = 392).

(C) Mean inter-temporal relational coding distance matrix, exhibiting the distances between the relational code vectors across all time bin pairs, averaged across stimuli. Statistical significance assessed and denoted as in Figures 3A and 3C.

(D) Example relational coding correlation time courses, displaying the Pearson r coefficients between the relational code at the indicated time bins in each panel vs. the relational codes from all 100 ms time bins between −0.5 and 2.2 s, averaged across all stimuli. Shaded areas represent ± SEM across the different stimuli (n stimuli = 28). Statistical significance and notations as in Figures 3B and 3D.

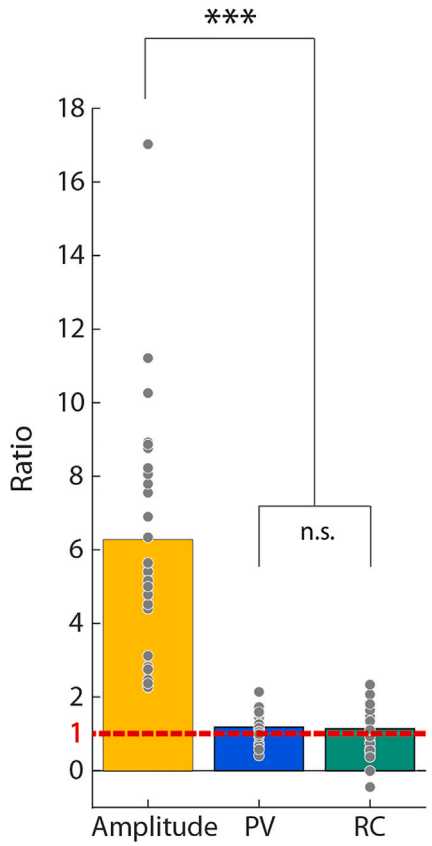


Figure 5. Across-time stability comparison of the response amplitude, population vectors (PVs) and relational code (RC) correlations, for the content-selective electrode set (n = 114)

Across-time stability is defined as the ratio between the values measured at time 0.2 s after stimulus onset and at time 1.2s (for additional details, see main text and STAR Methods). The bars indicate the across-stimuli average stability levels of the three parameters, and the gray dots indicate the individual stimuli values (n stimuli = 28). The dashed red line marks perfect across-time stability, i.e., a ratio of 1. ***p < 0.001, Bonferroni-corrected post-hoc comparisons test.

Dynamics and effects of signal-to-noise levels

What could be the source of the early transient increases observed in the inter-temporal correlations? A likely candidate may be a higher signal-to-noise ratio (SNR) in the initial, compared with the later, phase of the response. Examining the dynamics of the SNR in the content-selective contact group indeed revealed a major increase in SNR at the initial time windows (see Figure S1A), coinciding with the response amplitude increase, as well as a slight noise decrease. This phenomenon was common to all contact groups, as can be seen in Figure S2, albeit with the lowest overall SNR levels in frontal contacts and the highest in the early visual cortex.

In order to examine the potential effect of SNR on inter-temporal relational-code correlations, we compared correlations derived from time courses averaged across multiple trials with the noisier single trial-based time course correlations (see STAR Methods). This revealed that lower noise levels (i.e., in averaged trials) indeed resulted in higher inter-temporal correlation values (see Figures S1B and S1C).

Stability comparison across the different coding mechanisms

A quantitative comparison of the stability of the response amplitude, population patterns, and relational codes over time is presented in Figure 5. For this analysis, the stability of the response amplitude was defined by the ratio of the response at time 0.2 s with that of time 1.2 s. For the stability of the population vectors (PVs) and relational codes (RCs), we first calculated the correlations of the PVs and RC vectors at times 0.2 and 1.2 s with the vector at time 1.5 s as reference (see discussion for the rational of choosing this late time point for comparison, stemming from SNR differences) and then extracted the ratio between these correlation values. As can be seen, the response amplitude showed a greater level of transiency, with significantly higher early vs. late ratio levels, indicating less across-time stability, compared with the PV and RC measures, which were essentially unchanged across time, yielding an early/late ratio of ~1 (mean ratio values for response amplitude: 6.26 ± 3.44 SD, PV: 1.18 ± 0.37 SD, and RC: 1.14 ± 0.70 SD; Welch's ANOVA $F_{(2,\ 39.66)} = 27.82$, $p < 0.001$; Bonferroni post-hoc comparisons for amplitude > PV ratio: $p < 0.001$; amplitude > RC ratio: $p < 0.001$; PV > RC ratio: $p = 0.55$).

Pattern stability in high- and low-order cortical areas

In addition to the content-selective contact group, shown above, we examined contacts both from earlier areas in the visual hierarchy (early visual cortex, n = 32) as well as the fronto-parietal cortex (fronto-parietal visual contacts, n = 63; see Figure 1B for the anatomical localization of these contact groups). Additionally, it should be noted that the content-selective set encompassed a relatively heterogeneous group of high-order visual contacts. To examine a more homogeneous set, we defined a smaller group of solely face-selective contacts (n = 43; see STAR Methods and Figure 1B for their locations).

Figure 6 exhibits the inter-temporal distance matrices depicting the stability of the activity patterns, RDMs, and RCs for the face-selective and early visual contact sets (Figures 6A and 6B, respectively, contact locations shown on inflated cortical surfaces on the left). As can be seen, the face-selective group showed a high resemblance to the content-selective group results ($p < 0.005$, permutation testing using 1,000 shuffled permutations, FDR and cluster corrections). By contrast, the early visual group manifested a robust pattern stability effect but showed only transient, variable, and unstable RDM and relational geometries representations across time.

The visual fronto-parietal contacts displayed the most prominent differences, with essentially no significant inter-temporal correlations of the relational geometries or relational coding, although some weak pattern correlations were observed. However, due to the sparse sampling of this cortical region in the available dataset, this may stem from insufficient sampling rather than an essential qualitative difference between the regions.

Could the difference between the pattern stability and RDM transiency in the early visual cortex be due to less distinct individual pattern profiles—i.e., smaller distances between different stimuli-evoked patterns in the early cortex—compared with high-order contacts? To examine this possibility, we calculated all across-time averaged distances between activity patterns in

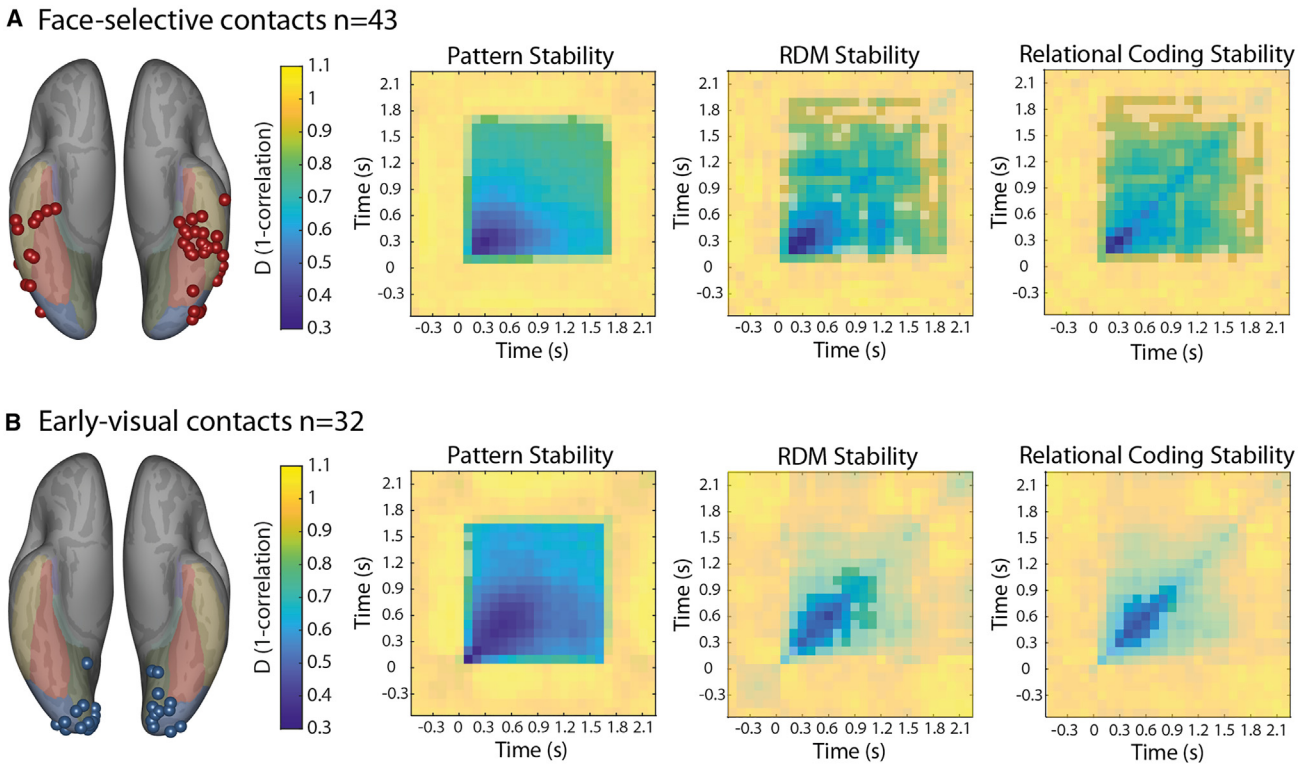


Figure 6. Mean activity patterns, RDMs, and relational coding stability across time in two additional visual functional groups

Contact locations are presented on inflated cortices in the left-side panels, for face-selective (n = 43) (A) and early visual (n = 32) (B) contacts. Inter-temporal distances (1 − r) of stimuli-evoked patterns, RDMs, and relational codes are calculated similarly to Figures 3A, 3C, and 4C, respectively, for each of the regions of interest (ROIs). Half-transparent colored regions mark distances that are non-significant, while distance values shown in opaque colors are statistically significant (1,000 random shuffling permutations, $p < 0.005$, FDR and cluster corrected).

early visual contacts and in the high-order content-selective contacts. The results revealed that the similarity distances in early visual cortex contacts were indeed significantly smaller than the distances in high-order content-selective contacts (two-tailed paired t test, $t_{405} = -10.37$, $p < 0.001$; see Figure S3 for the distributions of these distances). However, it is important to emphasize that due to a lack of a direct test for the effects of the smaller inter-stimuli distances, alternative sources for the transient results in the early visual cortex cannot be ruled out.

Stability of pattern decoding

An additional index of informational stability that is maintained in the pattern responses across time can be gained by examining the accuracy of inter-temporal decoding, of both the category and the single-item identity of the presented visual stimuli. In fact, inter-temporal category decoding was also chosen by the COGITATE adversarial collaboration as one of their two main measures when contrasting the predictions of the GWT vs. IIT theories.[18] Here, we trained a separate independent classifier on each single time point and then tested its performance across all other time bins, individually for exemplar and category models (see STAR Methods).

Decoding accuracy levels across time (see STAR Methods) are presented in the inter-temporal matrices in Figure 7. High-order visual contacts (content and face selective) showed a similar profile of significant ($p < 0.005$, shuffled permutation test, FDR and cluster corrections) inter-temporal decoding accuracy levels across all training vs. testing time bins during the stimulus presentation period, for both single-item and category decoding (see Figures 7A and 7B). The face-selective group showed slightly weaker decoding levels, likely due to the smaller number of contacts.

In order to rule out category-related effects in single-item classification, we additionally trained and tested separate exemplar decoders for each category class individually (see STAR Methods). Within-category exemplar decoding showed largely similar dynamics to the cross-category exemplar-decoding results, though they were more variable and less stable (see Figure S4).

Inter-temporal decoding results in early visual contacts (Figure 7C) interestingly indicated accurate decoding levels only for specific exemplars, with essentially no statistically significant category decoding. Note that this failure of category decoding was evident also when training and testing the decoder on the same time points, which are depicted in the main diagonal of the decoding matrix in Figure 7C (right panel), further indicating that this was not a result of a temporal instability effect but rather stemmed from the feature-based nature of early visual cortex representations.

Finally, although showing significant visual responses at the single contact level, the frontal visual contacts failed to show significant inter-temporal decoding for both single exemplars and categories.

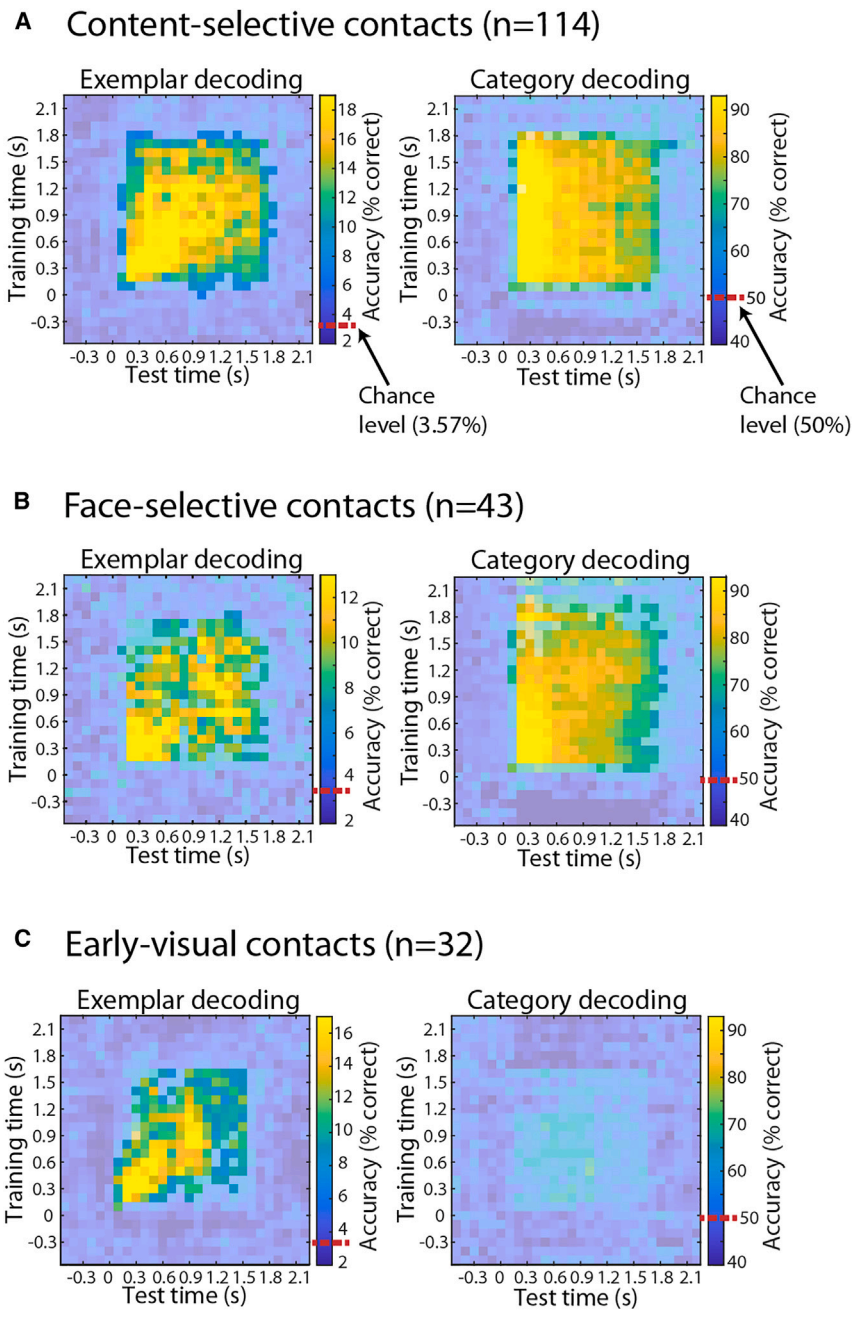


Figure 7. Exemplar and category decoding accuracy across time in the 3 contact groups

Content-selective contacts (n = 114) (A), face-selective contacts (n = 43) (B), and early visual (n = 32) (C) contacts. An individual classifier was trained on the activity patterns of each time bin, and then tested on all other time bins, using simple pattern-matching decoding for the exemplar decoding and a k-nearest neighbor (k-NN) classifier for category decoding. Dashed red lines on the color scale bar mark chance level (1/28 = 3.57% for exemplar decoding, and 1/2 = 50% for category decoding). Significant accuracy levels were calculated from a shuffled permutation test, comparing the real mean accuracy levels with the distribution of mean accuracy scores from 1,000 shuffled-label permutations. Statistically significant decoding- accuracy levels are shown in opaque colors ($p < 0.005$ permutation testing, FDR and cluster-based corrections), while insignificant values are shown as half-transparent. See STAR Methods for further details.

DISCUSSION

A dissociation between perception and activation magnitude

Our results (Figure 1) replicate previous findings of a dramatic decline (to about 30%) in the magnitude of the visual neuronal responses in visual cortex as well as in fronto-parietal regions, within a rather brief time period, during constant visual stimulation.[8,12] These studies have demonstrated that, specifically in high-order visual areas[12] the transient nature of response magnitude drastically dissociates from the constant optical stimulus on the one hand and from the perceptual state on the other. Importantly, this phenomenon is not merely a byproduct of controlled laboratory conditions[12] but appears as robustly under ecological, natural behavioral conditions as well.[14]

Thus, it is clear that the magnitude of neuronal activation is unlikely to serve as the neuronal correlate of sustained conscious perception. By contrast, uncovering stable parameters of the response can point to the neuronal mechanism that underlies the stability of visual perceptual awareness. Here, we hypothesized, based on recent theoretical considerations,[19,20,35-] that such a stable coding mechanism is readily available in the profiles of activation patterns and their relations.

Stable profiles of activity patterns revealed in high-order visual areas

We examined the population activity profile and relational coding hypothesis in the context of an image viewing task in which familiar face and place images were presented to the patients for 1.5 s. Our results demonstrate that while activation magnitude rapidly declined within less than 500 ms, examining the similarity of pattern profiles across time (defined through correlation-based distances), as well as the stimuli representational geometry, or RDM, and the RC, revealed that they were all sustained across time for the entire stimulus duration. This general stability effect was evident in high-order content-selective as well as face-selective contacts but was not as marked in early visual areas (see Figures 3, 4, and 6). We have previously proposed that different cortical regions may manifest distinct relational geometries, endowing them with their unique functional

specializations.[19] Here, we demonstrate this in the temporal domain and show that, indeed, there were substantial differences across early and high visual cortical areas.

Further direct comparison of the stability of the response amplitude vs. the population patterns and relation code correlations across time, obtained by examining the ratio of these coding parameters at an early vs. late time point in the trial, revealed significant reductions in the magnitude, compared with nearly perfect stability (ratio of ~1) in the other two population codes (see Figure 5), in the content-selective contact set.

Thus, the inter-temporal correlations of the three pattern-representation measures in the high-order visual cortex nicely demonstrate that pattern-based stability is evident at all levels: in the individual stimuli-linked population patterns, in the entire stimuli-space geometry defined by the distances between these patterns (RDM stability), and in the specific RC of each stimulus within the image vector space.

Examining the magnitude of inter-temporal correlations revealed a clear and transient increase in the early time points (300–500 ms) (Figures 3 and 4). It could be argued that this transient enhancement may be due to summation with an additional, transient pattern response profile that subsides, exposing a more stable and persistent pattern. However, careful examination of the inter-temporal correlation dynamics at different time points argues against this interpretation. In particular, examining the profile of inter-temporal correlations to the late pattern at 1.5 s (Figures 3B, 3D, 4D, and 5) revealed that it remained stable across the entire duration of the stimulus, which is incompatible with the existence of a different transient pattern. A more plausible explanation may be the reduction in the SNR, which we demonstrated can reduce inter-temporal correlations (see Figure S1) and was a common phenomenon across all contact groups (see Figure S2).

These transient changes in the SNR are also directly related to the choice of the 1.5 s time point as the reference for calculating the PV and RC stability (Figure 5). In principle, and given that perception was stable throughout the entire stimulus duration, one would expect a measure that tightly follows the perceptual state to be invariant to the choice of the reference time point for which stability measures are calculated. However, as discussed above and shown in Figures 3B, 3D, and 4D, this was clearly not the case. Rather, the correlation time courses demonstrate highly non-symmetrical behavior, where choosing an earlier time point as reference resulted in a skewed time course with increased early time bin correlations, while choosing a late time bin reference showed stable, relatively constant cross-time correlation levels. As we discussed above and demonstrate using control analyses (shown in Figure S1), a likely explanation for this asymmetry is the drastic change in the SNR in early vs. late time points, which strongly modulates the correlation values. To avoid this confounding, we chose the latest time point (1.5 s after stimuli onset) as our reference point for determining the stability of the PV and RC codes, as it was least likely to be affected by the transient SNR effect.

Stability of RC in face-selective contacts

Due to methodological limitations inherent in iEEG, our sampling necessitated combining high-order visual contacts with diverse category selectivity. However, examining a smaller group of contacts that showed a single-category selectivity, i.e., face-selective contacts, largely replicated the main inter-temporal stability effects from the content-selective contact set (though noisier, likely due to the smaller contact number; see Figure 6A). This is in agreement with previous studies from our group[24] as well as others,[23] suggesting that these contacts comprise a separate, functionally defined cortical geometry.

Comparing the number of contact recordings across the three major visual groups, early visual (n = 32), content selective (n = 114), and face selective (n = 43), supports the notion that the stability profile was dominated by the hierarchical level rather than sampling density or contact number, e.g., the content-selective and face-selective sets had a very different contact number (~2.5-fold difference), while their results proved to be very similar.

Inter-temporal decoding and its relevance to theories of conscious experience

Recently, in an attempt to experimentally examine the predictions of two consciousness theories—the GWT and the IIT—an "adversarial" collaboration was proposed to examine inter-temporal visual information stability as an experimental test of these theories.[18]

Due to the sparse sampling of fronto-parietal contacts in our study, we cannot draw clear conclusions from our inter-temporal decoding analysis in these regions, especially in regard to the GWT predictions that focus on this region. However, it should be noted that our response latency measurements were incompatible with the GWT prediction of a ~100 ms activation lag between posterior visual areas and the fronto-parietal ignition.[18] Our results reveal a latency lag within the visual system proper—i.e., between early and high-order face-selective regions—but only a minimal latency lag (~20 ms) between high-order face-selective cortex and the fronto-parietal response (Figure 1C), also compatible with our previous results from large-scale intracranial recordings.[36]

As for IIT predictions, our results appear to be compatible at the qualitative level with their prediction of significant and persistent inter-temporal decoding in the visual cortex. However, it should be noted that IIT is still a theory that is widely debated in the field,[37] and the persistent effects we find are predicted also by non-IIT hypotheses.[19] Furthermore, the IIT formalism may not correspond in a straightforward manner to the similarity structure of sensory representations.[38]

Finally, it should be noted that our finding of a temporally stable neuronal code does not necessarily refute the theoretical possibility that cognitive and neuronal states may fail to correspond to each other over time.[39]

Possible roles for high neuronal activations

The present findings raise the question of the role of the early, transiently high neuronal activation levels, which are a common observation in visual responses and a consistent signature of crossing the conscious perceptual threshold.[4,40–42] At present, the function of these high activity bursts remains an open question—however, there are a number of attractive possibilities, such as the recent proposal[19] of a binding mechanism essential for establishing population and relational coding. Additional

potential functions of the initial high firing could be its rapid spread to down-stream cortical areas and its registration in working memory and long-term memory and in rapid motor actions. Clearly, future studies are needed to resolve all these fundamental questions.

Conclusions

To summarize, our findings provide a plausible experimental resolution of a fundamental conundrum in perceptual neuroscience: how do we maintain stable perceptual states given the transient nature of neuronal activation magnitude? Our results suggest that this problem is solved through relational coding. Rather than relying solely on the overall magnitude of the response, perception is coded via the relative activation of different neurons and their overall population patterns. Together with recent conceptual advances emphasizing such coding mechanisms,[19,43,44] these findings support a temporal stability aspect of relational coding in human cognitive neuroscience.

Limitations of the study

The central limitation of the study is the relatively sparse sampling of frontal cortical regions that were available. This limitation, due to the strictly clinical criteria applied in the targeting of recording sites, prevented us from achieving a conclusive analysis regarding the representation of sustained visual information in frontal areas. Another limitation, again inherent in the clinical setting, was the limited time available for experimental testing, resulting in the single-task nature of the experiment. This was particularly problematic for achieving thorough mapping of frontal representations, which are likely to be strongly task dependent.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Participants
- METHOD DETAILS
  - Experimental task and stimuli
  - Contacts implant and data acquisition
  - Anatomical localization of the contacts
  - iEEG data preprocessing and HFB estimation
  - Definition and grouping of visually responsive contacts
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Examining inter-temporal stability
  - Population activity patterns' inter-temporal stability
  - RDM computation and across-time stability analysis
  - Inter-temporal stability of relational coding
  - Signal-to-noise analyses
  - Stability (across-time ratio comparison) analysis
  - Single-exemplar and category decoding across-time

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112614.

ACKNOWLEDGMENTS

We are grateful to the patients for their kind cooperation. This work was supported by the US-Israel BinationalFoundation grant 2017015 (R.M. and A.D.M.) and the CIFAR Tanenbaum Fellowship (R.M.).

AUTHOR CONTRIBUTIONS

R.M. and R.B.-D. conceived the study. R.M. and Y.N. designed the experiment. R.B.-D. analyzed the data. R.M. supervised the analysis. A.D.M. performed the surgeries and supervised the experiments and all aspects of data collection. Y.N. preprocessed the data. M.H. contributed to electrode localization. R.M. and R.B.-D. wrote and edited the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: November 15, 2022
Revised: March 14, 2023
Accepted: May 22, 2023
Published: June 6, 2023

REFERENCES

1. Sheinberg, D.L., and Logothetis, N.K. (1997). The role of temporal cortical areas in perceptual organization. Proc. Natl. Acad. Sci. USA *94*, 3408–3413. https://doi.org/10.1073/pnas.94.7.3408.
2. Tong, F., Nakayama, K., Vaughan, J.T., and Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. Neuron *21*, 753–759. https://doi.org/10.1016/s0896-6273(00)80592-9.
3. Gelbard-Sagiv, H., Mudrik, L., Hill, M.R., Koch, C., and Fried, I. (2018). Human single neuron activity precedes emergence of conscious perception. Nat. Commun. *9*, 2057. https://doi.org/10.1038/s41467-018-03749-0.
4. Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., Andelman, F., Neufeld, M.Y., Kramer, U., Fried, I., and Malach, R. (2009). Neural "ignition": enhanced activation linked to perceptual awareness in human ventral stream visual cortex. Neuron *64*, 562–574. https://doi.org/10.1016/j.neuron.2009.11.001.
5. Hasson, U., Hendler, T., Ben Bashat, D., and Malach, R. (2001). Vase or face? A neural correlate of shape-selective grouping processes in the human brain. J. Cognit. Neurosci. *13*, 744–753. https://doi.org/10.1162/08989290152541412.
6. Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature *435*, 1102–1107. https://doi.org/10.1038/nature03687.
7. Davidesco, I., Harel, M., Ramot, M., Kramer, U., Kipervasser, S., Andelman, F., Neufeld, M.Y., Goelman, G., Fried, I., and Malach, R. (2013). Spatial and object-based attention modulates broadband high-frequency responses across the human visual cortical hierarchy. J. Neurosci. *33*, 1228–1240. https://doi.org/10.1523/JNEUROSCI.3181-12.2013.
8. Golan, T., Davidesco, I., Meshulam, M., Groppe, D.M., Mégevand, P., Yeagle, E.M., Goldfinger, M.S., Harel, M., Melloni, L., Schroeder, C.E., et al. (2016). Human intracranial recordings link suppressed transients rather than 'filling-in' to perceptual continuity across blinks. Elife *5*, e17243. https://doi.org/10.7554/eLife.17243.

9. Müller, J.R., Metha, A.B., Krauskopf, J., and Lennie, P. (1999). Rapid adaptation in visual cortex to the structure of images. Science *285*, 1405–1408. https://doi.org/10.1126/science.285.5432.1405.

10. Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. J. Neurophysiol. *97*, 3155–3164. https://doi.org/10.1152/jn.00086.2007.

11. Bandettini, P.A., Kwong, K.K., Davis, T.L., Tootell, R.B., Wong, E.C., Fox, P.T., Belliveau, J.W., Weisskoff, R.M., and Rosen, B.R. (1997). Characterization of cerebral blood oxygenation and flow changes during prolonged brain activation. Hum. Brain Mapp. *5*, 93–109.

12. Gerber, E.M., Golan, T., Knight, R.T., and Deouell, L.Y. (2017). Cortical representation of persistent visual stimuli. Neuroimage *161*, 67–79. https://doi.org/10.1016/j.neuroimage.2017.08.028.

13. Norman, Y., Yeagle, E.M., Khuvis, S., Harel, M., Mehta, A.D., and Malach, R. (2019). Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. Science *365*, eaax1030. https://doi.org/10.1126/science.aax1030.

14. Podvalny, E., Yeagle, E., Mégevand, P., Sarid, N., Harel, M., Chechik, G., Mehta, A.D., and Malach, R. (2017). Invariant temporal dynamics underlie perceptual stability in human visual cortex. Curr. Biol. *27*, 155–165. https://doi.org/10.1016/j.cub.2016.11.024.

15. Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron *24*, 187–203. https://doi.org/10.1016/s0896-6273(00)80832-6.

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

17. Malach, R. (2012). Targeting the functional properties of cortical neurons using fMR-adaptation. Neuroimage *62*, 1163–1169. https://doi.org/10.1016/j.neuroimage.2012.01.002.

18. Melloni, L., Mudrik, L., Pitts, M., and Koch, C. (2021). Making the hard problem of consciousness easier. Science *372*, 911–912. https://doi.org/10.1126/science.abj3259.

19. Malach, R. (2021). Local neuronal relational structures underlying the contents of human conscious experience. Neurosci. Conscious. *2021*, niab028. https://doi.org/10.1093/nc/niab028.

20. Lau, H., Michel, M., LeDoux, J.E., and Fleming, S.M. (2022). The mnemonic basis of subjective experience. Nat. Rev. Psychol. *1*, 479–488. https://doi.org/10.1038/s44159-022-00068-6.

21. Edelman, S., Grill-Spector, K., Kushnir, T., and Malach, R. (1998). Toward direct visualization of the internal shape representation space by fMRI. Psychobiology *26*, 309–321.

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

23. Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. Front. Syst. Neurosci. *2*, 4.

24. Grossman, S., Gaziv, G., Yeagle, E.M., Harel, M., Mégevand, P., Groppe, D.M., Khuvis, S., Herrero, J.L., Irani, M., Mehta, A.D., and Malach, R. (2019). Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. Nat. Commun. *10*, 4934. https://doi.org/10.1038/s41467-019-12623-6.

25. Davidesco, I., Zion-Golumbic, E., Bickel, S., Harel, M., Groppe, D.M., Keller, C.J., Schevon, C.A., McKhann, G.M., Goodman, R.R., Goelman, G., et al. (2014). Exemplar selectivity reflects perceptual similarities in the human fusiform cortex. Cereb. Cortex *24*, 1879–1893. https://doi.org/10.1093/cercor/bht038.

26. Kriegeskorte, N., and Diedrichsen, J. (2019). Peeling the onion of brain representations. Annu. Rev. Neurosci. *42*, 407–432. https://doi.org/10.1146/annurev-neuro-080317-061906.

27. Cichy, R.M., Pantazis, D., and Oliva, A. (2016). Similarity-based fusion of MEG and fMRI reveals spatio-temporal dynamics in human cortex during visual object recognition. Cereb. Cortex *26*, 3563–3579. https://doi.org/10.1093/cercor/bhw135.

28. Cichy, R.M., Pantazis, D., and Oliva, A. (2014). Resolving human object recognition in space and time. Nat. Neurosci. *17*, 455–462. https://doi.org/10.1038/nn.3635.

29. Carlson, T., Tovar, D.A., Alink, A., and Kriegeskorte, N. (2013). Representational dynamics of object vision: the first 1000 ms. J. Vis. *13*, 1. https://doi.org/10.1167/13.10.1.

30. Deitch, D., Rubin, A., and Ziv, Y. (2021). Representational drift in the mouse visual cortex. Curr. Biol. *31*, 4327–4339.e6. https://doi.org/10.1016/j.cub.2021.07.062.

31. Norman, Y., Yeagle, E.M., Harel, M., Mehta, A.D., and Malach, R. (2017). Neuronal baseline shifts underlying boundary setting during free recall. Nat. Commun. *8*, 1301. https://doi.org/10.1038/s41467-017-01184-1.

32. Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., and Malach, R. (2005). Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. Science *309*, 951–954. https://doi.org/10.1126/science.1110913.

33. Nir, Y., Fisch, L., Mukamel, R., Gelbard-Sagiv, H., Arieli, A., Fried, I., and Malach, R. (2007). Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. Curr. Biol. *17*, 1275–1285. https://doi.org/10.1016/j.cub.2007.06.066.

34. Manning, J.R., Jacobs, J., Fried, I., and Kahana, M.J. (2009). Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. J. Neurosci. *29*, 13613–13620.

35. Edelman, S. (1998). Representation is representation of similarities. Behav. Brain Sci. *21*, 449–467. https://doi.org/10.1017/s0140525x98001253.

36. Noy, N., Bickel, S., Zion-Golumbic, E., Harel, M., Golan, T., Davidesco, I., Schevon, C.A., McKhann, G.M., Goodman, R.R., Schroeder, C.E., et al. (2015). Ignition's glow: ultra-fast spread of global cortical activity accompanying local "ignitions" in visual cortex during conscious visual perception. Conscious. Cogn. *35*, 206–224. https://doi.org/10.1016/j.concog.2015.03.006.

37. Michel, M., Beck, D., Block, N., Blumenfeld, H., Brown, R., Carmel, D., Carrasco, M., Chirimuuta, M., Chun, M., Cleeremans, A., et al. (2019). Opportunities and challenges for a maturing science of consciousness. Nat. Hum. Behav. *3*, 104–107.

38. Pautz, A. (2019). What is the integrated information theory of consciousness? J Conscious Stud *26*, 188–215.

39. Herzog, M.H., Drissi-Daoudi, L., and Doerig, A. (2020). All in good time: long-lasting postdictive effects reveal discrete perception. Trends Cogn. Sci. *24*, 826–837.

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

41. Quiroga, R.Q., Mukamel, R., Isham, E.A., Malach, R., and Fried, I. (2008). Human single-neuron responses at the threshold of conscious recognition. Proc. Natl. Acad. Sci. USA *105*, 3599–3604. https://doi.org/10.1073/pnas.0707043105.

42. Moutard, C., Dehaene, S., and Malach, R. (2015). Spontaneous fluctuations and non-linear ignitions: two dynamic faces of cortical recurrent loops. Neuron *88*, 194–206.

43. Malach, R. (2022). The role of the prefrontal cortex in conscious perception: the localist perspective. J. Conscious. Stud. *29*, 93–114.

44. Lau, H., Michel, M., LeDoux, J.E., and Fleming, S.M. (2022). The mnemonic basis of subjective experience. Nat. Rev. Psychol. *1*, 479–488.

45. Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

46. Papademetris, X., Jackowski, M.P., Rajeevan, N., DiStasio, M., Okuda, H., Constable, R.T., and Staib, L.H. (2006). BioImage Suite: an integrated medical image analysis suite: an update. Insight J. *2006*, 209.

47. Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. Med. Image Anal. *5*, 143–156. https://doi.org/10.1016/s1361-8415(01)00036-6.

48. Fischl, B. (2012). Neuroimage *62*, 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021.

49. Argall, B.D., Saad, Z.S., and Beauchamp, M.S. (2006). Simplified intersubject averaging on the cortical surface using SUMA. Hum. Brain Mapp. *27*, 14–27. https://doi.org/10.1002/hbm.20158.

50. Bokil, H., Andrews, P., Kulkarni, J.E., Mehta, S., and Mitra, P.P. (2010). Chronux: a platform for analyzing neural signals. J. Neurosci. Methods *192*, 146–151. https://doi.org/10.1016/j.jneumeth.2010.06.020.

51. Van Der Maaten, L., Postma, E., and Van den Herik, J. (2009). Dimensionality reduction: a comparative. J. Mach. Learn. Res. *10*, 13.

52. Hentschke, H., and Stüttgen, M.C. (2011). Computation of measures of effect size for neuroscience data sets. Eur. J. Neurosci. *34*, 1887–1894. https://doi.org/10.1111/j.1460-9568.2011.07902.x.

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

54. Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage *17*, 825–841.

55. Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage *31*, 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021.

56. Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T.T., Mohlberg, H., Amunts, K., and Zilles, K. (2008). Cortical folding patterns and predicting cytoarchitecture. Cereb. Cortex *18*, 1973–1980. https://doi.org/10.1093/cercor/bhm225.

57. Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage *53*, 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010.

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

59. Parvizi, J., and Kastner, S. (2018). Promises and limitations of human intracranial electroencephalography. Nat. Neurosci. *21*, 474–483. https://doi.org/10.1038/s41593-018-0108-2.

60. Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Statist. *29*, 1165–1188.

61. Foxe, J.J., and Simpson, G.V. (2002). Flow of activation from V1 to frontal cortex in humans. A framework for defining "early" visual processing. Exp. Brain Res. *142*, 139–150. https://doi.org/10.1007/s00221-001-0906-7.

62. Raiguel, S.E., Lagae, L., Gulyàs, B., and Orban, G.A. (1989). Response latencies of visual cells in macaque areas V1, V2 and V5. Brain Res. *493*, 155–159. https://doi.org/10.1016/0006-8993(89)91010-x.

63. Groppe, D.M., Urbach, T.P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology *48*, 1711–1725.

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

65. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. J Neurosci Meth *164*, 177–190.

66. Levina, E., and Bickel, P.J. (2005). Maximum Likelihood Estimation of Intrinsic Dimension, pp. 777–784.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Original Data	This paper	https://doi.org/10.5281/zenodo.7813404
Software and algorithms		
MATLAB 2019b	MathWorks Inc.	https://www.mathworks.com/products/matlab.html
EEGLAB v14.1.2.	Delorme and Makeig[45]	https://sccn.ucsd.edu/eeglab/index.php
BioImage Suite	Papademetris et al.[46]	https://bioimagesuiteweb.github.io/webapp/
Presentation software	Neurobehavioral Systems, Inc.	https://www.neurobs.com/
FSL FLIRT	Jenkinson and Smith[47]	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslInstallation
FreeSurfer 6.0	Fischl[48]	https://surfer.nmr.mgh.harvard.edu/
SUMA	Argall et al.[49]	https://afni.nimh.nih.gov/pub/dist/doc/SUMA/suma/SUMA_do1.htm
Chronux	Bokil et al.[50]	http://chronux.org/
DRtoolbox	Van Der Maaten et al.[51]	https://lvdmaaten.github.io/drtoolbox/
MES toolbox	Hentschke and Stuttgen[52]	https://github.com/hhentschke/measures-of-effect-size-toolbox
Custom-developed analysis code	This paper	https://doi.org/10.5281/zenodo.7813404

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Rafi Malach (rafi.malach@gmail.com).

Materials availability

This study did not generate any new reagents.

Data and code availability

- Anonymized original data has been deposited on Zenodo and is available for public download: https://doi.org/10.5281/zenodo.7813404.
- Original code has been deposited on Zenodo and is available for public download: https://doi.org/10.5281/zenodo.7813404.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants

Intracranial recordings were obtained from 13 patients (10 females, mean age 34.7 ± 9.6), monitored for pre-surgical evaluation of epileptic foci due to pharmacologically resistant epilepsy, at the North Shore University Hospital in NY. As part of the clinical assessment, all patients were implanted with subdural or depth contacts (see Table S1 for individual demographic and electrode coverage details). The study was conducted in accordance with the latest version of the Declaration of Helsinki, and all patients provided a fully informed consent to participate, including consent to publish the results, according to the US National Institute of Health guidelines, monitored by the institutional review board at the Feinstein Institute for Medical Research. No epileptic clinical seizures occurred during the experiment.

METHOD DETAILS

Experimental task and stimuli

The participants viewed images of famous faces and places, as part of a longer study that included a later phase of free recall of the images that were presented.[13,31] The experiment was divided into two runs. Each run began with a 200 s resting-state period with eyes closed (the first two patients performed the resting-state task on a different day). Immediately afterward, 14 different images of well-known faces and places were presented to the participants (7 pictures from each category). In total, 28 different naturalistic and colorful stimuli were included in the experiment, 14 in each run. Due to copyright restrictions, examples of the stimuli presented in Figures 1A and 2, as well as in the graphical abstract, are accurate color illustrations of original stimuli exemplars, drawn by a graphical artist (https://www.fiverr.com/valentinagulina?source=gig_page), and acquired via the Fiverr platform (https://www.fiverr.com). These illustrations were purchased by and are now the property of the Weizmann Institute of Science.

Each image was presented for 1500ms, with 750ms inter-stimulus intervals (ISIs) between them, during which a fixation cross was presented (see Figure 1A for task depiction). Single items were repeated four times each, in a pseudo-random order such that no image was repeated twice consecutively. The stimuli were presented on an LCD screen, using Presentation software (Version 0.70, www.neurobs.com), at a ~60 cm viewing distance (image size was 16.5° × 12.7). The participants were asked to carefully view these images and try to remember them as well as possible, including details regarding unique colors, facial expressions, perspective, lightning, etc. They were informed that after the viewing phase, they will be asked to recall the pictures they saw, and to specifically describe their prominent visual features, not just simply name them. This was done to ensure that they focused on visual information that was specific to the presented images, and not just their semantic details. All participants performed the task with a high success rate, with an average of 8.8 ± 2.7(SD) remembered items per run. No participants were excluded from the analysis based on their performance levels. For more details, see Norman et al..[13,31]

Contacts implant and data acquisition

All recordings were conducted at the Northshore University Hospital, Manhasset, NY, USA, at the patients' quiet bedside. Contacts were either subdural grids or strips, placed directly on the cortical surface, or depth electrodes (Ad-Tech, Racine, WI, Integra, Plainsboro, NJ, and PMT Corporation, Chanhassen, MN). Subdural contacts were either 1 or 3 mm in diameter, with inter-contact spacing of either 4 or 10 mm. In depth contacts were 2 mm platinum cylinders, with 4.4 mm inter-contact spacing, and a diameter of 0.8 mm. The intracranial signals were referenced to a vertex screw or to a subdermal electrode, and were electronically filtered between 0.1 and 200 Hz, and then sampled at a rate of 500 or 512 Hz. The data was stored for offline analysis using XLTEK EMU128FS or NeuroLink IP 256 systems (Natus Medical Inc., San Carlos, CA). Stimulus-triggered electrical pulses were sent upon stimuli onsets and recorded along with the iEEG data for precise alignment of task protocol to neural activity.

Anatomical localization of the contacts

Before the electrode implantation, the patients underwent a T1 weighted 1 mm isometric anatomical MRI scan using a 3T Signa HDx scanner (GE Healthcare, Chicago, Illinois). Following the implant, a computed tomography (CT) and a T1-weighted anatomical MRI scan on a 1.5-T Signa Excite scanner (GE Healthcare) were acquired, in order to enable precise localizations of each contact. All scans were skull-stripped using FSL's BET algorithm, and the post-implant CT was first co-registered to the post-implantation MRI scan, and then to the pre-implantation MRI anatomical scan, using rigid affine transformation, as implemented in FSL's Flirt.[47,53,54] Concatenation of these two co-registrations allowed visualization of the post-implant CT scan on top of the preoperative MRI scan, while minimizing potential errors due to possible surgery and implantation brain shifts. Individual contacts were next identified by manual inspection of the co-registered CT and post-implant MRI, and marked in each patient's preoperative MRI native space, using the BioImage Suite.[46] Next, electrode projection onto the cortical surface was performed as in the following manner[8,13]: First, individual patients' cortical surfaces were segmented and reconstructed from the pre-implant structural MRI, using FreeSurfer 6.0.[48] Then, each contact was allocated to the nearest vertex on the individual's cortical surface. Contacts that were farther than 8mm from the cortical surface were excluded from further analyses. In order to project all contacts from all patients onto a single template cortical surface, while maintaining their specific relations to individual gyri and sulci, the 3-D cortical mesh of each individual was resampled and standardized using SUMA,[49] allowing visualization of all contacts on a single common cortical template (FreeSurfer's FSaverage). Finally, colored labels on the cortical surface, as shown in Figure 1B, were derived from different surface-based anatomical atlases available in FreeSurfer,[55–57] including a probabilistic visual-retionotopic atlas.[58]

iEEG data preprocessing and HFB estimation

All data analysis was performed in MATLAB (MathWorks), using EEGLAB,[45] Chronux,[50] DRtoolbox,[51] MES toolbox,[52] and in-house developed code. Raw iEEG time-courses were inspected manually and statistically to detect noisy/corrupted channels, which were then excluded from further analysis. Signals that were recorded at a sampling rate of 512 Hz were downsampled to 500 Hz, for consistency. The 60 Hz power line interference, as well as its harmonics, were removed using zero-lag linear-phase Hamming-windowed FIR band-stop filters. Next, all contacts were re-referenced to a robust common average (excluding the corrupted channels).

The high-frequency broadband (HFB) signal was defined as the mean normalized power of frequencies between 60 and 160 Hz (High-gamma), the range of frequencies that is commonly used as the electrophysiological marker of neural population

activity.[13,31–34,59] HFB power was computed by filtering the signal into 20 Hz bands between 60 and 160 Hz, employing zero-lag linear-phase Hamming window FIR filters, using EEGLAB. Next, the momentary amplitude in each sub-range was calculated as the absolute value of the filtered signal's Hilbert transform.[4,13,24,36] Since the 1/f profile of the signal's power spectrum results in larger values of lower frequencies, we normalized each sub-range by dividing it with its mean value. Finally, we averaged the normalized values across all sub-ranges. The resulting HFB data was inspected for transient electrical artifacts, defined as signals above 5 SD in the common average signal (the averaged time-series across all electrodes). Time windows of 200ms around these peaks were removed from further analyses.

The HFB data was then epoched relative to stimulus onset (−550 to 2250 post stimulus onset). In order to normalize the data to account for differences in overall HFB amplitude levels in different contacts, event-related HFB responses to visual stimulation were normalized relative to a baseline period of −400 to −100 ms prior to stimuli onset, by dividing each time point in the epoched data by the mean baseline amplitude. Finally, since the HFB amplitude tends to follow a log-normal distribution, as do many other measures of population firing rate, the HFB values were log-transformed by 10*log10.[13,31]

Definition and grouping of visually responsive contacts

Visually-responsive contacts were identified by comparing the post-stimulus HFB response of each contact (averaged across the time window of 100 to 500ms post stimulus onset) to its pre-stimulus baseline (averaged from −400 to −100ms prior to stimulus onset), using a two-tailed Wilcoxon signed-rank test. All contact p values (from all patients) were pooled together to control for the false discovery rate (FDR).[60] Visually responsive contacts were defined as those that displayed a significant HFB response ($p_{fdr}<0.05$). Next, we grouped together visually responsive contacts based on their anatomical, functional, and response latency features, to the following subsets: early visual (V1/V2), face-selective, content-selective, and fronto-parietal visual contacts.

In order to calculate the response latency of each visually responsive contact, we compared the HFB amplitude in each post-stimulus time point to the pre-stimulus baseline, using a paired t test. Response latency was defined as the first time point in which the HFB amplitude was significantly higher than baseline ($p < 0.05$), and remained significant for the next 50 ms at least.[61] Previous single-unit studies in monkeys reported a maximal response latency of ∼180ms in early visual regions V1 and V2.[62] Thus, we defined early visual contacts (n = 32, marked in blue in Figure 1B) as the visually-responsive contacts that showed a response latency of 180ms or less, and were located in Brodmann areas 17 and 18 (V1 and V2), as based on the Brodmann atlas implemented in FreeSurfer.[56]

To define face-selective contacts, the mean HFB responses across the 100-500ms time window post-stimulus onset, were compared between faces and places using a Wilcoxon rank-sum test. Contacts that showed significantly higher activation in response to faces than places ($p_{fdr}<0.05$), and were anatomically located beyond early visual regions, excluding frontal cortex, were defined as face-selective contacts (n = 43, denoted in red and orange in Figure 1B).

An additional set of visual contacts was termed content-selective contacts, and included contacts that showed preferential responses to specific stimuli exemplars as compared to others; importantly, these were not necessarily only face or place selective (based on[13]). In order to define these contacts, we inspected the visually-responsive contacts located in 6 anatomical regions across the ventral visual hierarchy, based on the Desikan Killiany atlas,[55] including the lateral occipital cortex (LO), inferior temporal gyrus (ITG), lingual gyrus, parahippocampal gyrus (PHG), fusiform gyrus, and entorhinal gyrus, excluding the early-visual contacts that were already defined. Visually responsive contacts that were located in these ROIs, and showed significant content-selectivity in their responses, defined as a difference of at least 3.5 SD between the top 10 preferred images and bottom 10 images, were included in the content-selective contact set (n = 114, shown in yellow and orange in Figure 1B). Notice 27 contacts fitted the criteria for both the face-selective group and the content-selective groups, depicted in orange in Figure 1B.

Finally, in order to examine the activity patterns in prefrontal and parietal regions, we defined a fronto-parietal anatomical ROI based on the Desikan Killiany atlas labels, that included the superior frontal gyrus, rostral middle frontal gyrus, pars orbitalis, pars triangularis, pars opercularis, precentral gyrus, postcentral gyrus, supramarginal gyrus, orbital frontal gyrus, and the anterior cingulate. All visually-responsive contacts that were located within these regions were included in the fronto-parietal visual group (n = 63, shown in green in Figure 1B).

Additional visually-responsive contacts that did not fall under one of the above visually-selective subgroups were labeled ''other visually-responsive contacts'', and are shown in white in Figure 1B; non-visual contacts are marked in gray.

QUANTIFICATION AND STATISTICAL ANALYSIS

Examining inter-temporal stability

To quantify the stability of the three main coding parameters in this study-activity pattern stability, relational geometry (as reflected in the representational similarity distances) and relational coding stability, we correlated these measures across all pairs of time bins, in what we termed inter-temporal correlations between all time points. This generated temporal, time-by-time matrices, containing the similarities between these different features across the trial durations. The specific details of these analyses, for each type of coding measure, are detailed in separate sections below.

Population activity patterns' inter-temporal stability

In order to quantify the stability of the profiles of stimuli-evoked population activity patterns across time, we calculated the correlation based distances of these patterns between all time points. First, we split the HFB data into epochs relative to stimuli onset, including the 550ms pre-stimulus fixation, 1500ms stimulus presentation duration, and the 750ms post-stimulus fixation period (resulting in 2800ms long epochs). Each stimulus (n = 28) was presented 4 times, thus we had 112 such epochs in total per contact (28 stimuli*4 repetitions). Next, we down sampled the data by averaging it across time in 100ms non-overlapping time windows, hence resulting in 28 time bins per trial (or epoch). Then, for each contact subset (see "definition and grouping of visually responsive contacts" above), we arranged the data into a 3D matrix per each stimulus, resulting in 28 matrices depicting the activity patterns, of the following form: $A_{\textit{n contacts} \times \textit{28 time bins} \times \textit{4 repetitions}}$. Examples of such activity patterns (for the content-selective contact group), for time bin 0.3s post-stimulus onset, are shown for four different stimuli in Figure 2 (activity patterns were averaged across the 4 stimulus repetitions in this figure).

For every A_i matrix (i = 1,2 … 28 stimuli), we correlated the contact activity pattern vectors between all time-bin combinations, so that for each specific time bin we obtained the Pearson's r correlation values between the population patterns of all other time bins. In order to avoid auto-correlations and possible noise artifacts, the correlations were calculated in a leave-1-out procedure across the 4 stimulus repetitions, in the following manner: The correlation values for each time-bin pair were calculated 4 times, so that in every iteration we extracted one trial and averaged the pattern activity across the remaining 3 trials, and then correlated between the averaged pattern and the single-trial pattern. We then proceeded to calculate the average correlation value across these 4 iterations. Distances were defined as 1-Pearson's r. Thus, we obtained 28 inter-temporal distance matrices (for each of the 28 stimuli), of the following form: $D_{\textit{28 time bins} \times \textit{28 time bins}}$. Finally, we averaged across the individual stimuli D matrices, resulting in a single averaged inter-temporal distances matrix, shown in Figure 3A (and in Figure 6- denoted as the "Pattern Stability" matrices).

In order to assess the statistical significance of the inter-temporal pattern correlation/distance values, we shuffled the contact labels in the original data 1,000 times, and then proceeded to recompute the correlations in the same manner as described above, resulting in a shuffled 1,000 r-values distribution for each i^{th} time bin vs. j^{th} time bin entry. p-values were calculated as the proportion of correlation-based distance (1-r) values derived from the shuffled permutations that were smaller than the original distance value derived from the data (smaller d values indicate higher similarity). An FDR correction[60] was then applied on the resultant 784 p values (28 × 28 time-bin pairs), to control for a 0.5% false discovery rate across all time-bin comparisons ($p_{FDR} = 0.005$). In addition, we assessed the maximal cluster size in each shuffled iteration (using a cluster-defining threshold of p = 0.05). FWE-corrected p values were computed as the proportion of random clusters larger than or equal to the clusters observed in the actual data.[63–65] The distance values that survived the combined FDR and cluster-based corrections were marked as significant in the inter-temporal pattern stability matrices.

RDM computation and across-time stability analysis

In order to compute the stimuli pairwise distances matrix, i.e. the representational dissimilarity matrix (RDM), we first arranged the HFB epoched data, for each contact group separately, in neuronal stimulus representation matrices B, in the form of $B_{\textit{n contacts} \times \textit{28 stimuli} \times \textit{4 repetitions per stimulus}}$, with a separate $\mathbf{B_i}$ (1 = 1,2 … 28) matrix for each time bin. The pairwise distance between the neural response patterns of two exemplars was calculated as 1-Pearson's r, depicting the correlation between the activity pattern vectors of two stimuli. Distances were calculated between all possible stimuli pairs, separately for each $\mathbf{B_i}$matrix. The correlations were calculated in a leave-1-out manner to avoid autocorrelations and artifacts, as explained in the previous section (population activity patterns' inter-temporal stability), iterating across the 4 possible combinations of a single trial vs. 3-trials averaged activity vector, and deriving four correlation values for each stimuli pair, that were saved in separate RDMs. Thus, we obtained a total of 112 RDMs depicting the distances between all possible stimuli pairs, each in the form of $R_{\textit{28 stimuli} \times \textit{28 stimuli}}$, one for each of the four leave-1-out iterations, and for each time bin separately (n = 28).

In order to determine the stability of the RDMs across-time, as an indication to the stability of the stimuli representational geometry, we next examined the inter-temporal correlations of these RDMs. We first "unrolled" the top and bottom triangular halves of each RDM (including the main diagonal) into vectors, and averaged across them. This resulted in a vector of 406 distance values ($N * (N - 1)/2 + N$), depicting the distances between all stimuli pairs, including the distances between different presentations of the same stimulus (N = 28, denoting the number of stimuli). These pairwise distance vectors were obtained for each time bin, and for each of the four leave-1-out iterations. Next, we correlated these distance vectors between all possible time-bin pairs, again in a leave-1-out-manner, resulting in a 28 time bins × 28 time bins × 4 leave-1-out iterations matrix, which we then averaged across the third dimension. This resulted in a 28 × 28 distances matrix, depicting the inter-temporal stability of the RDM, or stimuli representational geometric space (as shown in Figure 3C, and in Figure 6- the matrices termed "RDM stability"). Statistical significance was assessed through 1,000 shuffled permutations, using FDR and cluster-based corrections, in the same manner as described above in "population activity patterns' inter-temporal stability".

Inter-temporal stability of relational coding

The relational code of a stimulus was defined as the vector of pattern distances from that specific stimulus to all other stimuli. In order to examine the across-time stability of these relational codes, we first extracted RDMs for each time bin separately, as described in the section above (see RDM computation and across-time stability analysis). Every row in these RDMs consists of a distances vector,

denoted d_i (i = 1,2 … 28 stimuli) that holds the relational code for a specific stimulus-i.e. the distances between one stimulus and all other stimuli, for a specific time bin. Next, we calculated the across-time correlations for every d_i vector separately, in a leave-1-out manner, as described in the two sections above. This resulted in 3D distance matrices of size $28_{\text{time bins}} \times 28_{\text{time bins}} \times 4_{\text{leave-1-out combinations}}$ which were averaged across the third dimension, resulting in a $28_{\text{time bins}} \times 28_{\text{time bins}}$ mean distances matrix for each individual stimulus. Lastly, we averaged across the 28 individual-stimulus distances matrices, resulting in the final grand average $28_{\text{time bins}} \times 28_{\text{time bins}}$ distances matrix, that reflected the average across-stimuli stability of the relational coding across trial duration (e.g. in Figure 4C). This was done separately for every contact group. Statistical significance was assessed through 1,000 shuffled permutations, using FDR and cluster-based corrections, in the same manner as described in the two sections above.

Signal-to-noise analyses

Signal to noise ratio (SNR) was calculated as the average HFB signal across all trials and stimuli, divided by the standard deviation computed across repetitions of the same stimulus, and averaged across all stimuli. It was then averaged across all contacts, separately for each ROI. Examining the SNR time course revealed that earlier time points in the trial, immediately following stimulus onset, displayed higher SNR levels. In order to explore how higher SNR levels can potentially contribute to the increased correlation values we obtained in these time points, we rerun the stimulus relational coding analyses pipeline twice, as described above, but using different SNR levels of the data. This was done by running the analysis once with averaging across individual repetitions of the same stimulus before calculating the correlations (in the leave-1-out manner as described above for the RDM computation, and then by averaging across pairs of RDMs before calculating their inter-temporal correlations); and once without averaging-i.e. calculating correlations between single trials, and only then averaging across the correlation values derived from the different single-trial combinations. Since averaging across trials increases the SNR, in this manner we could highlight the effects of higher SNR levels, which led to higher correlation levels in the relational coding stability analysis.

Stability (across-time ratio comparison) analysis

In order to quantify and compare the stability of the different coding mechanisms across time, we calculated the ratio between the parameter values measured early in the trial, at time 0.2s after stimulus onset, as compared to the same parameter measured at time point 1.2s after onset, for the content-selective contacts set. Perfect stability will yield a ratio of 1, while signal-information decay with time will lead to values larger than 1. Specifically, response amplitude stability was calculated as the individual stimuli responses at time 0.2, averaged across all electrodes and repetitions of each individual stimulus, divided by the mean responses at time 1.2s. Figure 5 presents the individual ratio values for each stimulus, as well as the average ratio across all stimuli. For the population pattern ratio calculation, we extracted the correlation time course of the population pattern at time 1.5s with the patterns from all other time bins (see population activity patterns' inter-temporal stability in STAR Methods and Figure 3B), separately for each stimulus, averaging the correlation time-courses across the four leave-1-out iterations based on the multiple stimulus repetitions. Stability was calculated as the ratio between the correlation with time bin 0.2s, divided by the correlation at time 1.2s. The individual-stimuli correlation ratios, as well as the across-stimuli average, are presented in Figure 5. Relational code stability was similarly calculated based on the correlation time-courses of the relational code vectors at time 1.5 with all other time bins (see inter-temporal stability of relational coding in STAR Methods and Figure 4D), separately for each stimulus, after averaging across the four leave-1-out iterations. Stability was calculated as detailed above for the population patterns.

The correlation time course with time point 1.5s was selected as the reference for the population pattern and relational code correlations, due to the transiently changing SNR levels, which were shown to significantly influence correlation levels, and were stabilized toward the end of the trials (see dynamics and effects of signal to noise levels and Figure S1). Statistical testing comparing the three stability measurements was performed using Welch's ANOVA, followed by post-hoc Bonferroni tests.

Single-exemplar and category decoding across-time

To test whether information regarding the identity of single exemplars, as well as the category of the stimulus, is sustained across time, we trained and tested inter-temporal exemplar and category decoders, separately for each contact group that we defined. First, we applied pattern dimensionality reduction (PCA) to the data, in order to reduce feature number before training the decoders.[13,51] This was done by averaging the epoched HFB data across the trial durations, and then across individual stimuli repetitions, constructing a mean feature matrix of the visual responses to all stimuli $(28_{exemplars} \times n_{contacts})$, and then applying the PCA algorithm on this matrix. In order to determine the optimal number of PCs to retain, we estimated the true dimensionality of the data (i.e. the intrinsic dimension) using a Maximum Likelihood Estimation (MLE) algorithm.[66] We proceeded to maintain the first 12 PCs, that accounted for 89.2% of the variance in the data for the content-selective contact group. Next, we applied the linear transformation obtained from the PCA to the original data (using an out-of-sample extension of the PCA), resulting in a 12-dimensional linear space, which was arranged in a grand four-dimensional matrix G of the form: $12_{PCs} \times 28_{time\ bins} \times 28_{exemplars} \times 4_{repetitions\ per\ exemplar}$. Thus, entry $G_{i,j,k,l}$ refers to the response level of the i^{th} PC, at the j^{th} time bin (100ms bins), for the k^{th} exemplar and the l^{th} repetition trial of that exemplar.

In order to train the inter-temporal single-exemplar decoders, we applied a simple template matching decoding technique, in the following manner: We looped over all individual time bins in two nested loops, in order to train and test every possible combination of training a decoder on time bin m, and testing it on time bin n, for a total of $28^2 = 784$ time-bin combinations. For each unique training

vs. testing time-bin combination, we ran 1,000 decoding iterations. In each iteration, we randomly chose a single trial for each exemplar, extracted the response it elicited at time bin *n* from all PCs, and assigned it to a test pattern 2D matrix $T\ (28_{exemplars} \times 12_{PCs})$. The responses at time bin *m* were averaged across the remaining 3 trials for each exemplar, and assigned to a ref. 2D matrix $R\ (28_{exemplars} \times 12_{PCs})$. Thus, every decoding iteration for training time point *m* vs. testing time point *n* began with a 2D test matrix T_n and reference, or training, 2D matrix R_m: row *i* in matrix T_n is the vector of responses of all PCs to the randomly chosen single trial of the i^{th} exemplar at time *n*; row *i* in matrix R_m is the vector of averaged responses across the remaining 3 trials from all PCs, at time *m*, to the i^{th} exemplar. Next, we assigned 28 decoded labels based on the maximal correlation value between each row in the test and reference matrices: on every decoding step, the correlations between all rows of matrix T_n and all rows of matrix R_m were obtained, and the maximal value was detected. The label of the corresponding row in the reference matrix (e.g. "Face 1") was then assigned to the row with the highest correlation to it from the test matrix. The assigned pair of test-reference rows were then excluded from subsequent steps in the current decoding iteration, such that every reference and test rows were assigned only once per iteration. Then, the next highest correlation value between the T_n and R_m rows was detected, and so on. After assigning all test-reference row pairs, the number of correctly assigned stimuli labels was recorded for that iteration. One thousand decoding iterations were performed for every training time bin *m* and testing time bin *n* combination, and the decoding accuracy for a specific *m* vs. *n* time bin combination was defined as the mean percentage of accurately decoded single exemplars across the 1,000 iterations. This resulted in a decoding accuracy matrix of size $28_{training\ time\ bins} \times 28_{testing\ time\ bins}$ in which each entry held the decoding accuracy for that specific training vs. testing time bins pair.

In order to obtain inter-temporal category decoding, we trained a k-nearest neighbors (k-NN) decoder (using k = 1) on the PCA transformed data, iterating through all possible training vs. testing time bin combinations, similarly to the procedure in the single-exemplar decoding, described above. Classification accuracy was obtained via a leave-1-out validation method. Thus, for every training time bin *m* and testing time bin *n* (*m*,*n* = 1,2 … 28) combination, we trained the k-NN decoder on the category labels of 111 trials at time *m*, and tested its performance on the remaining trial at time *n* (28 different exemplars × 4 repetitions each = 112 trials in total). Category classification accuracy was defined as the percent of correctly decoded iterations, for every *m* vs. *n* time bins combination. This resulted in a $28_{training\ time\ bins} \times 28_{testing\ time\ bins}$ matrix, similarly to the exemplar decoding matrix, in which each entry detained the category decoding accuracy level for that specific training vs. testing time bins pair.

Statistical significance for the decoding analyses were assessed through random-shuffling of the item labels (single-exemplars/category labels, in accordance), across 1,000 permutations. In each shuffled permutation, the classifier accuracy scores we recomputed, as described above. p-values for each entry in the inter-temporal decoding accuracy matrices were defined as the proportion of shuffled accuracy scores larger than the actual classifier score. Correction for multiple comparisons was achieved via combined FDR correction (p = 0.005) and cluster-size based correction (as detailed above in population activity patterns' inter-temporal stability).