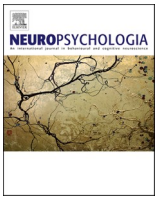
Neuropsychologia 165 (2022) 108111

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia



Single neuron evidence of inattentional blindness in humans

Brandon Freiberg [a,1], Moran Cerf [b,*,1]

[a] Columbia University, New York, NY, 10027, USA
[b] Northwestern University, Evanston, IL, 60208, USA

ARTICLE INFO

Keywords:
Inattentional blindness
Human electrophysiology
Singe neuron recording
Spike-field coherence
Consciousness

ABSTRACT

Recording directly from the brain of a patient undergoing neurosurgery with electrodes implanted deep in her skull, we identified neurons that change their properties when the patient became consciously aware of content. Specifically, we showed the patient an established clip of a gorilla passing through the screen, unnoticeable, in a classic inattentional blindness task, and identified a neuron in the right amygdala that fired only when the patient was aware of the gorilla. A different neuron coded the moment of insight, when the patient realized that she had missed the salient gorilla in previous trials. A third cluster of neurons fired when the patient was exposed to a post-clip question ("How many passes did you count?") and reflected on the content. Neurons in this cluster altered their response behavior between unaware and aware states.

To investigate the interplay between the neurons' activity and characterize the potential cascade of information flow in the brain that leads to conscious awareness, we looked at the neurons' properties change, their activities' alignment and the correlation across the cells. Examining the coherence between the spiking activity of the responsive neurons and the field potentials in neighboring sites we identified an alignment in the alpha and theta bands. This spike-field coherence hints at an involvement of attention and memory circuits in the perceptual awareness of the stimulus.

Taken together, our results suggest that conscious awareness of content emerges when there is alignment between individual neurons' activity and the local field potentials. Our work provides direct neural correlate for the psychological process by which one can look at things directly but fail to perceive them with the "mind's eye".

1. Introduction

Consciousness is one of the hallmarks of our existence. Studies of the neural correlates of consciousness (NCC) are nonetheless limited and focus primarily on behavioral outcomes of its experience. Despite the challenge in truly studying consciousness, works on attention and awareness have been able to characterize the limits of the conscious experience that can be witnessed outwardly (Koch and Tsuchiya, 2007). Further, studies with patients under minimal states of consciousness (i.e., comatose patients) have been used to infer the limits and boundaries of the conscious experience (Majerus et al., 2005). Similarly, usage of pharmaceuticals that alter the conscious states have allowed researchers access to the qualitative experience of consciousness (Koch, 2012). Finally, insights from animal studies, electrophysiological works, and theoretical neuroscience have led researchers to suggest hypotheses as to the neural sites likely to be implicated with NCC (Crick and Koch, 2005), the mechanisms by which consciousness emerges from the collective interaction of neural circuits (Tononi, 2008), and consciousness' theoretical attributes (Baars, 1997).

However, to date we do not have a single established empirical way of testing conscious experience. This is partially due to the challenge in detecting, replicating and objectively measuring the phenomenological attributes of consciousness (Chalmers, 2007), and partially due to the difficulty in generating tasks that elucidate varying states of consciousness. Specifically, it is challenging to alter consciousness while having a person reflect on the lack thereof, and it is impossible to conduct animal studies on the topic given the animals' inability to reflect on psychological experiences.

One experimental method that attempted to examine differences in individual behavior during states of awareness is "inattentional blindness" (Mack and Rock, 1998). Inattentional blindness is a phenomenon in which subjects fail to notice a fully visible visual stimulus that appears

* Corresponding author.
E-mail address: moran@morancerf.com (M. Cerf).
[1] Authors contributed equally.

https://doi.org/10.1016/j.neuropsychologia.2021.108111
Received 11 March 2021; Received in revised form 30 October 2021; Accepted 6 December 2021
Available online 10 December 2021

at the center of their fovea because of a manipulation that draws attention away from the stimulus. Specifically, while people directly look at an item on a screen — and the content is projected into their visual cortex accurately from the retina — they are not able to see the image in their mind's eye, as an information bottleneck prevents the imagery from arriving at perceptual neural circuits. For example, recent work (Simons and Chabris, 1999) has demonstrated the effect in a clip where a notable gorilla appears on the screen yet fails to draw attention from subjects who were tasked with counting the number of basketball passes between players.

The failure to consciously recognize the gorilla in the experimental paradigm is robust and allows for consistent replication. While numerous explanations for the failure to notice the gorilla are offered in the psychology literature, only a handful of works in neuroscience have looked at the neural correlates of inattentional blindness in the context of the dynamic stimulus. The existing works have investigated the phenomenon empirically using neuroimaging in humans (Shafto and Pitts, 2015) or electrophysiology in monkeys (Wegener et al., 2004) and have suggested that the inability to see presented stimuli may be driven by cognitive load (Cartwright-Finch and Lavie, 2007), by the allocation of resources to processes that bypass the attention foci (Koivisto et al., 2004), or by the deteriorates of attentional neural circuitry over time (Remington et al., 2014).

Two of the challenges in studying inattentional blindness neurally emerge from the fact that: 1) knowledge of the task makes it impossible to repeat it across trials (i.e., once subjects recognize the gorilla they are likely to never miss it again), and 2) generating stimuli that will elicit the blindness effect is extremely difficult. Additionally, the growing public awareness of the psychological phenomena makes it challenging to identify uninformed subjects for testing. Therefore, there is a limited concentrated effort by researchers investigating the neurobiological correlates of the phenomenon.

Within the context of this experiment, the challenge of finding subjects who are: 1) unaware of the task, 2) do not see the overt stimuli when presented with it, and 3) fail to see it *repeatedly*, is compounded by the unlikely chance of finding such person who also 4) has microelectrodes implanted in their brain, and 5) has one of the microelectrodes land near a neuron that is coding the concept of a gorilla (such that one can see whether the brain responds to a gorilla even if the person is not able to "see" it). Combining all those conditions leads to an infinitesimal likelihood of researchers being able to study consciousness in a direct way from the brains of humans.[2]

Nevertheless, we were fortunate to find such a subject, and tested her conscious awareness using single neuron recording from neurons responsive to the masked stimuli, repeatedly. The subject was a patient undergoing brain surgery for clinical resection of an epilepsy focus. During a screening test with the patient, a single neuron in the right amygdala was identified as selectively and invariantly responding to conceptual representations of gorillas. Given that the patient was unfamiliar with any inattentional blindness tasks, nor was she familiar with the specific popular clip of a gorilla emerging through the screen while basketball players pass balls among themselves in a distracting fashion, we used the validated stimulus to test her response mediation through multiple exposures to the clip.

We replicate the inattentional blindness experiment behaviorally, and show evidence for modulation of the neural responses in alignment with the subject's conscious awareness. Additionally, we show a change in response properties by neurons in other sites with the conscious modulation. Further, we note a variation in the coherence between the neurons and their input network that is reflected by modulation in the local field potentials (LFP). This study allows for a direct test of the gap between what a person can report seeing and what neurons in their brain reflect independently.

We find: 1) a neuron that responds invariantly to gorilla stimuli and modulates its activity properties — latency, duration and spike density — in alignment with the change in conscious awareness, 2) a second neuron that alters its properties when the subject recognizes that she failed to notice something prior and experiences a moment of "aha" indicating her learning of the masked gorilla, and 3) a cluster of neurons that respond when the clip is over and a question pertaining to its content appears.

Our empirical results support a hierarchical model of information processing in the brain where consciousness can be seen as a property of the network in relation to the individual neuron. In particular, our findings suggest such a model where 1) a finite locus of attention must be spread among various tasks – i.e., increased focus in one area may come at the expense of another, and 2) responses to disparate stimuli are aggregated in such a manner to form responses to new stimuli (e.g., a neuron that receives input from one neuron that encodes "gorilla" and receives input from a second neuron that encodes "video clip" may itself encode "gorilla specific to clip"). We speculate that such hierarchical cascade of processing may be optimal for consciousness in humans in terms of rapid processing that maximizes information integration, in line with suggested evolutionary work in animals (MacIver et al., 2017), and theoretical work on consciousness emergence in humans (Tononi, 2008).

2. Methods

2.1. Subject

A single female, in her forties participated in the study (specific details of the individual are not shared as these can be identifiable). The subject was a patient undergoing brain surgery who was implanted bilaterally with chronic intracerebral depth electrodes - primarily in the medial temporal lobe (MTL) - to localize the epileptic focus for possible clinical resection (see Fig. 1 for depiction of the electrodes' locations). Subject was right-handed and had normal vision. Prior to the experimental procedure, the subject took a variety of cognitive and affective tests and scored within normal range.

2.2. Data acquisition

Data were acquired from 96 micro-wires which were implanted and localized exclusively based on clinical criteria. Neuronal signals were recorded using 9–10 depth electrodes (Ad-Tech, Medical Instrument Corp., Racine, WI). Each depth ("macro") electrode contained a bundle of nine Platinum–Iridium micro-wires protruding from its tip. Eight of the micro-wires were high-impedance active recording channels, and the ninth was a low-impedance reference micro-wire. The neural signal was acquired using the Blackrock recording system (Blackrock Microsystems, Salt Lake City, UT) and Neuroport software. Behavioral data were acquired using a dedicated laptop running Matlab's Psychophysics toolbox. Subject's verbal reporting of her experience was recorded using a dedicated channel feeding into the Blackrock acquisition device. Video recording of the experiments was collected via the clinical acquisition system. The behavioral data were synchronized with the neural data using a dedicated USB cable feeding the timestamps from the laptop to the Blackrock device.

[2] To estimate the probability of this, we used the work of Waydo et al. (2006) suggesting that the chance of detecting a response to *any* concept is about 0.03% (and the chance of detecting a response to a *specific* concept such as a gorilla - out of about 100 shown – is, therefore, 0.0003%). Along with the chance of the gorilla being unseen in a single trial (34%, based on saliency estimation; Einhäuser et al., 2009), and in multiple (say, 3) trials being 0.0039, the probability of finding a patient with intracranial depth electrodes that shows a response to the stimuli is about 1:10, 000, 000,000.

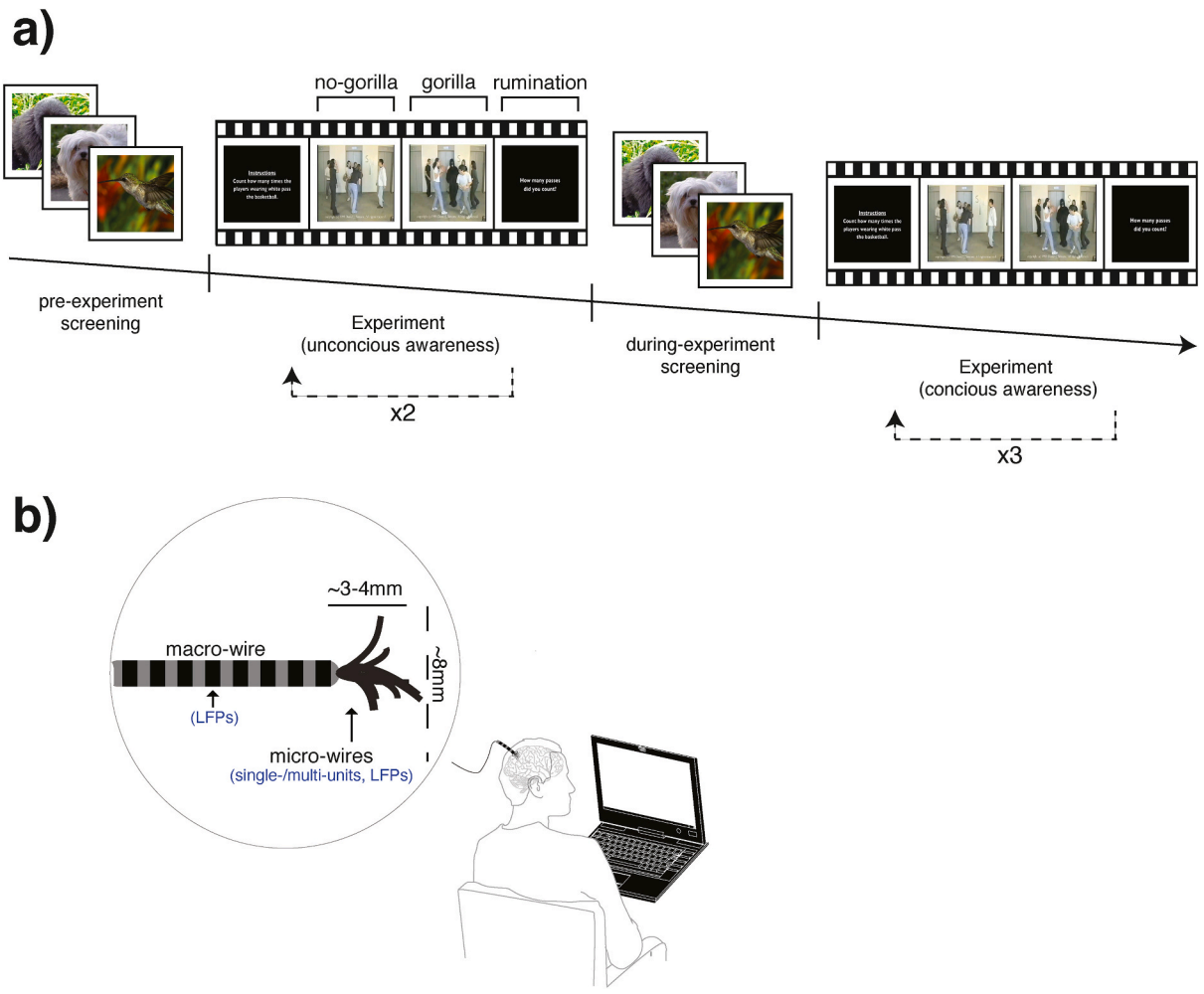


Fig. 1. Experimental procedure. **a)** subject performed a pre-screening prior to the study. Following, she participated in two unaware trials, a during-experiment screening, and three additional aware trials. **b)** illustration of the macro/micro-wires implanted in the subject's skull.

2.3. Spike detection

Differential neuronal signals (recording range ±3,200 mV) were sampled at 30,000 Hz. The extracellular signals were band-pass filtered (300 Hz to 3 kHz). Spikes were detected and pre-sorted automatically, offline, using Matlab's *wave_clus* toolbox. Manual verification and classification as an artifact, multi- or single-unit was based on spike shape, spike variance, inter-spike interval distribution per cluster, and the presence of a plausible refractory period. Twenty of the channels had no units in them. Twenty-eight of the remaining channels captured duplicate unit data (i.e., channels 73 and 74 showed identical spike shapes and times). For every duplicate channel, we removed one of the two channels, retaining the other one, if they were located within the same channel (i.e., two units captured by the same micro-electrode), or if they were within two separated channels within the same macro-electrodes. After removal of noisy and duplicate channels we isolated 85 units

Table 1
Neural responses breakdown.

Site[a]	Total neurons	Images		Clip		
		Pre-experiment screening	During-experiment screening	Beginning[b]	Rumination[c]	Gorilla visible
RA	16	1	2	5	13	2
RH	10			7	8	
REC	7			2	7	
ROF	8			8	8	
RAC	13			6	10	
RAF	7			1	7	
LA	6			5	3	
LH	9				5	
LEC	5			4	4	
LAC	4				4	

[a] RA = Right amygdala; RH = right hippocampus; REC = Right entorhinal cortex; ROF = Right orbitofrontal cortex; RAC = Right anterior cingulate; RAF = Right anterior frontal. L## corresponds to the *left* hemisphere.
[b] Between 0 and 2s from clip onset.
[c] Between 33 and 36s from clip onset (when the clip is over and a question about the number of throws seen appears).

across 48 channels (0.88 ± 0.76 units per channel; mean ± standard-deviation throughout the text unless stated otherwise). The micro-wires' locations spanned left/right hemispheres and included amygdala, hippocampus, entorhinal cortex, orbitofrontal cortex, cingulate and the frontal pole (see Table 1 and Supplementary Fig. 1 for a breakdown of sites, neurons, and responses). Spike detection was done using the default *wave_clus* parameters. Namely, minimum threshold for spike detection of 5 standard-deviations above the median, and artifacts rejected at 50 standard-deviations above the median. Sixty-four samples (20 before the spike peak and 44 after the peak; about 2 ms of data) were isolated from each spike for future spike sorting. Spike sorting was, initially, done automatically using the *wave_clus* sorting algorithm, default parameters. Following, a qualitative correction of the detection was conducted manually. Sorted units were classified as single units or multi units based on spike shape, spike variance, the ratio between spike peak value and noise level, the inter-spike interval, and the presence of a refractory period for the units (Quiroga et al., 2005). Of the 85 units detected, 18 were single units and the remaining multi units (see supplementary materials for details on the spike sorting and multi/single-unit classification). Following the spike sorting, all analyses pertaining to individual neurons were done at the unit level. That is, if multiple units were captured within a single micro-electro, we computed each unit's properties (i.e., mean firing rate) independently.

2.4. Neural responses to stimuli

A neuron was deemed responsive to a stimulus if the firing rate during a 1-s bin in which the stimulus was present exceeded 3 standard-deviations above the mean firing rate of that channel through the entire data acquisition. That is, we first binned the entire recording data to 1 s bins and calculated the mean and standard-deviation firing for that channel. Following, we used the window where a stimulus was present to estimate whether the neuron was considered responsive (see details of the method in Quiroga et al., 2005).

2.5. Stimuli

All stimuli were presented on a 15-inch laptop monitor (1680 *x* 1050 pixels). Subject's distance from the monitor was about 50 cm. The stimuli were presented in a rectangle at the center of the screen and occupied about 400 pixels, and about 11° visual angle. Stimuli were surrounded by a black background. Prior to each stimulus presentation a fixation cross was displayed at the center of the screen.

2.6. Experimental procedure

Prior to the experiment, the subject performed a screening session ("pre-experiment screening") where she was presented with 100 images that corresponded to concepts that she was familiar with and expressed interest in. The choice of stimuli was done based on a brief prior interview with the subject where her preferences (musical taste, political affiliation, food preferences, etc.) were discussed. Each image in the screening was presented six times in random order. Following the screening session, data were analyzed offline to identify neuronal responses that coded invariant concepts consistently (Quiroga et al., 2005). During the pre-screening, a single neuron which responded to an image of a gorilla was identified ("Gorilla" neuron in Table 2).

About 4 hours following the pre-screening our experimental procedure was initiated. The subject viewed a clip where a number of players wearing white/black shirts were tossing basketballs among themselves. About halfway through the clip a gorilla emerged from the right side of the screen, moved to the center of the screen, and made itself visible by tapping its chest before exiting from the left. The subject was asked to count the number of basketball passes conducted among the players wearing white shirts alone. This task was likely to distract her from noticing the salient gorilla. The clip effectively was expected to yield the inattentional blindness behavioral effect (Simons and Chabris, 1999).

The subject did not exhibit any signs of noticing the gorilla appearing during the clip. When the clip viewing was complete the subject reported the number of passes she had noticed. Following the first clip viewing the subject was asked to repeat the task again and count the passes a second time, to "ensure the accuracy of her report". The subject did not display any signs of noticing the gorilla in the second viewing as well. Once the second viewing was complete, an experimenter asked the subject explicitly if she had noticed "anything unusual about the clip" and the subject reported that she had not. We term the two initial viewings of the clip "unaware" trials.

Following the unaware trials, the experimenters verified the ability to detect the gorilla neuron as well as the subject's awareness of the

Table 2
Neural responses properties.

	n[a]	SU/MU[b]	Stimulus	Site[c]	Latency[d] (ms)	Duration (ms)	Baseline firing rate (Hz)	Response firing rate (Hz)
During-experiment screening	1	1/0	Gorilla 01	RA	370 ± 92	257 ± 141	0.12 ± 0.50	1.83 ± 1.11
	1	1/0	Gorilla 02		416 ± 111	426 ± 60		2.33 ± 1.44
	1	1/0	Gorilla 03		498 ± 251	504 ± 86		1.67 ± 1.56
	1[e]	1/0	Clip-gorilla	RA	449 ± 192	253 ± 202	0.83 ± 1.84	3.50 ± 1.88
Clip	13[f]	5/8	Rumination	RA	1162 ± 167	1464 ± 518	1.01 ± 1.74	5.95 ± 3.97
	8	1/7	Rumination	RH	978 ± 75	1820 ± 195	0.57 ± 1.28	4.67 ± 2.74
	7	3/4	Rumination	REC	1165 ± 189	1797 ± 384	0.65 ± 1.54	5.10 ± 2.37
	8	1/7	Rumination	ROF	931 ± 43	1927 ± 461	0.87 ± 2.16	7.41 ± 6.50
	10	1/9	Rumination	RAC	1050 ± 102	1486 ± 155	0.60 ± 1.12	3.16 ± 1.20
	7	0/7	Rumination	RAF	1069 ± 32	1900 ± 154	0.74 ± 2.27	5.25 ± 3.65
	3	0/3	Rumination	LA	1180 ± 149	611 ± 717	0.19 ± 0.10	0.91 ± 0.18
	5	0/5	Rumination	LH	1299 ± 201	1384 ± 526	0.27 ± 0.94	2.35 ± 0.53
	4	0/4	Rumination	LEC	1058 ± 62	1920 ± 374	0.93 ± 1.08	5.80 ± 2.43
	4	1/3	Rumination	LAC	1255 ± 50	700 ± 594	0.15 ± 0.41	2.22 ± 1.10

[a] Number of responsive units for stimulus, in site.
[b] SU = Single unit, MU = Multi unit.
[c] RA = Right amygdala; RH = right hippocampus; REC = Right entorhinal cortex; ROF = Right orbitofrontal cortex; RAC = Right anterior cingulate; RAF = Right anterior frontal. L## corresponds to the *left* hemisphere.
[d] All values (latency, duration, firing rate) reflect mean ± standard-deviation.
[e] The clip-gorilla neuron's response in the during-experiment screening was not significant. The response was significant during the clip viewing. Nevertheless, we report the neuron during-experiment properties as well, because of its relationship with the gorilla neuron (i.e., spike-distance from the gorilla neuron).
[f] When multiple units were responsive for the stimulus, we show in the firing rate columns (rightmost columns) the mean ± standard-deviation of the mean firing rate of all units.

gorilla in the clip by asking the subject to participate in a short screening session ("during-experiment screening"). In this screening the subject saw nine images, including the image of the gorilla used initially in the screening ("Gorilla 01"), two new images of gorillas that were not shown before ("Gorilla 02", "Gorilla 03"), and two still frames from the basketball clip with and without the gorilla ("Clip-gorilla", "Clip-No Gorilla"). In total, the subject saw four unique images containing a gorilla (one that appeared prior in the pre-screening, two new ones, and one from the clip). When the screening was complete the subject was asked to view the clip a third time. This time, rather than counting the passes the subject was instructed to observe the clip naturally and attend to unusual occurences. The subject identified the gorilla as it entered the screen in the third viewing. Finally, the subject was asked to view the basketball clip two additional times. In total, the subject viewed the clip three times while being aware of the gorilla ("aware" trials; Fig. 1).

3. Results

3.1. Neural correlate of conscious awareness of the triggering concept

We, first, validated that a single neuron shown to be responsive to invariant representations of a gorilla in the pre-experiment screening remained accessible in our experiment. In a dozen exposures to images pertaining to gorillas shown in random order the single unit (right amygdala, mean baseline firing rate: 0.12 ± 0.50 Hz; total number of spikes during the session: 369) significantly increased its firing rate (1.83 ± 1.11 Hz; $p < 10^{-6}$; *Wilcoxon rank-sum*) when the gorilla stimulus was present (Fig. 2 and Table 2). We further refer to this neuron as the "gorilla neuron".

To test whether we can identify a neural correlate of the conscious awareness of the gorilla, we investigated the gorilla neuron's activity when the subject was viewing the clip depicting a gorilla. The subject saw the clip five times. In the first two viewings the subject was not aware of the gorilla notably flailing its hands at the center of the screen. In line with the subjective experience of the subject, the gorilla neuron did not alter its firing rate when the gorilla was on the screen (Fig. 3). In the following three viewings of the gorilla, when the subject was aware of it, the neuron significantly increased its firing rate (0.37 ± 0.40 Hz; $T(18) = 2.905$; $p = 0.009$; *two-tail t-test* comparing mean number of spikes in 1-s bins, per trial, during the gorilla on-screen epoch of the aware trials and the unaware trials). As an additional way to estimate the ability to distinguish the two conscious states we used a *decision tree* classifier to decode whether the subject was viewing the clip in unaware/aware trials merely based on the spiking activity of the neuron. The decision tree (implemented using Python's *scikit-learn* library) used logistic regression regularization with uniform initial weights (set to 1.0), no dual formulation, L2 regularization, a tolerance for stopping of 0.0001, and maximum of 1500 iterations. The classifier was used with no cost-complexity pruning, Gini impurity function for the quality of split measure, minimum number of samples required to split of 2, and a minimum sample required for a node to act as a leaf of 1. We did not use parameter tuning beyond the default ones. The classifier's performance showed 77% accuracy in decoding the state given a single trial data (chance: 50%).

3.2. Neural correlate of internal rumination related to the awareness of the triggering concept

Given the clear distinction between the unaware/aware trials, we further pursued an exploratory investigation of other neurons' properties changes with a shift in awareness. We found a large number of units that activated during the occurrence of the question about the clip's content, which appeared after the clip was over. Specifically, 69 units showed an increased firing rate (above baseline, established as the mean response of a specific neuron 1000–300 ms prior to all image onsets) when the clip was over and the subject was asked to reflect on the content she had just viewed before answering the question "How many passes did you count?". These units were time-locked to the onset of the question and showed a consistent latency and firing rate (Table 2). We categorize this cluster of units as "rumination neurons" since they were responsive when the subject was reflecting on the recently viewed clip. The rumination neurons could be responding to reading the text shown at the end of the viewing, the effort expended towards the reading, the preparation to answer the question, or to other mental function. We use the rumination label based on the epoch during which they responded.

The rumination neurons significantly altered their properties between unaware and aware trials. To characterize the properties of the rumination neurons, we investigated the change in firing rate above baseline during the rumination epoch. Specifically, we estimated the percent change in firing rate of each rumination neuron between the "aware" trials and the "unaware" trials (Fig. 4). The rumination neurons significantly increased their mean firing density during the aware trials compared to the unaware ones ($T(136) = 3.782$; $p < 10^{-4}$; *two-tail t-test*). Interestingly, the rumination neurons were not isolated in a single site but were distributed both within the right amygdala as well as other MTL and frontal sites.

In order to rule out the possibility that the effect was driven by an increase in firing rate across the entirety of the clip (as opposed to just the response to the end of clip stimuli), we compared the percent change in firing rate during the rumination epoch to that of randomly selected epochs ("control"). The rumination neurons significantly changed their firing rate compared to the control ($T(68) = 3.784$; $p < 10^{-5}$; *two-tail t-test*). The majority of rumination neurons (72%) showed an increase in firing rate when the clip ended and the reflection epoch started, in the aware trials (mean change: 2.60 ± 1.87%, across 50 neurons that showed an increase). Notably, over half (55%) of the rumination neurons also exhibited an increase in firing rate at the clip onset – when the instructions to count basketball passes were displayed. This response, initially made us consider classifying the neurons' response differently (i.e., related to the appearance of a text on the screen, as this happens both in the beginning/end of the clip). However, the firing activity at the clip onset was not significant at the 3 standard-deviation benchmark for the majority of the neurons and, for those that it was, did not significantly change in response to the shift from unaware to aware trials. We report the units/sites that showed the increase in activity at the clip onset in Table 1 ("Beginning").

3.3. Evidence of revealed content learning

As the revelation by subjects that a gorilla was visible in the clip frequently leads to a surprise, accompanied by an immediate permanent learning (noted in the literature as a moment of insight, or "a-ha"; see Qiu et al., 2010) after which subjects consistently notice the gorilla in future viewings, we were interested in characterizing the initial recognition moment. To isolate this event, we chose to reveal the existence of the gorilla through exposing the subject to a still frame from the clip that clearly indicates that a gorilla was visible in the clip at the center of the frame (rather than debrief the subject verbally). The subject therefore was exposed to the content firsthand as she saw the image and recognized that it was in the clip throughout. This allowed us to time-lock the exact onset of the reveal and study the neural properties of this moment of learning.

A single neuron in the right amygdala that was located near the gorilla neuron, yet was not the same (see spike shapes difference in Figs. 2 and 5), showed a response profile that aligned with the exposure to the gorilla in the clip alone when the revelation occurred. We label the neuron responsive for the gorilla in the clip "clip-gorilla" neuron. The clip-gorilla neuron did not respond to other images of gorillas outside of the specific one in the clip (Supplementary Fig. 2). That is, this neuron was distinguished from the gorilla neuron in that it responded solely to the image of the gorilla in the clip when the subject learned about its existence. Notably, the clip-gorilla neuron showed distinct firing

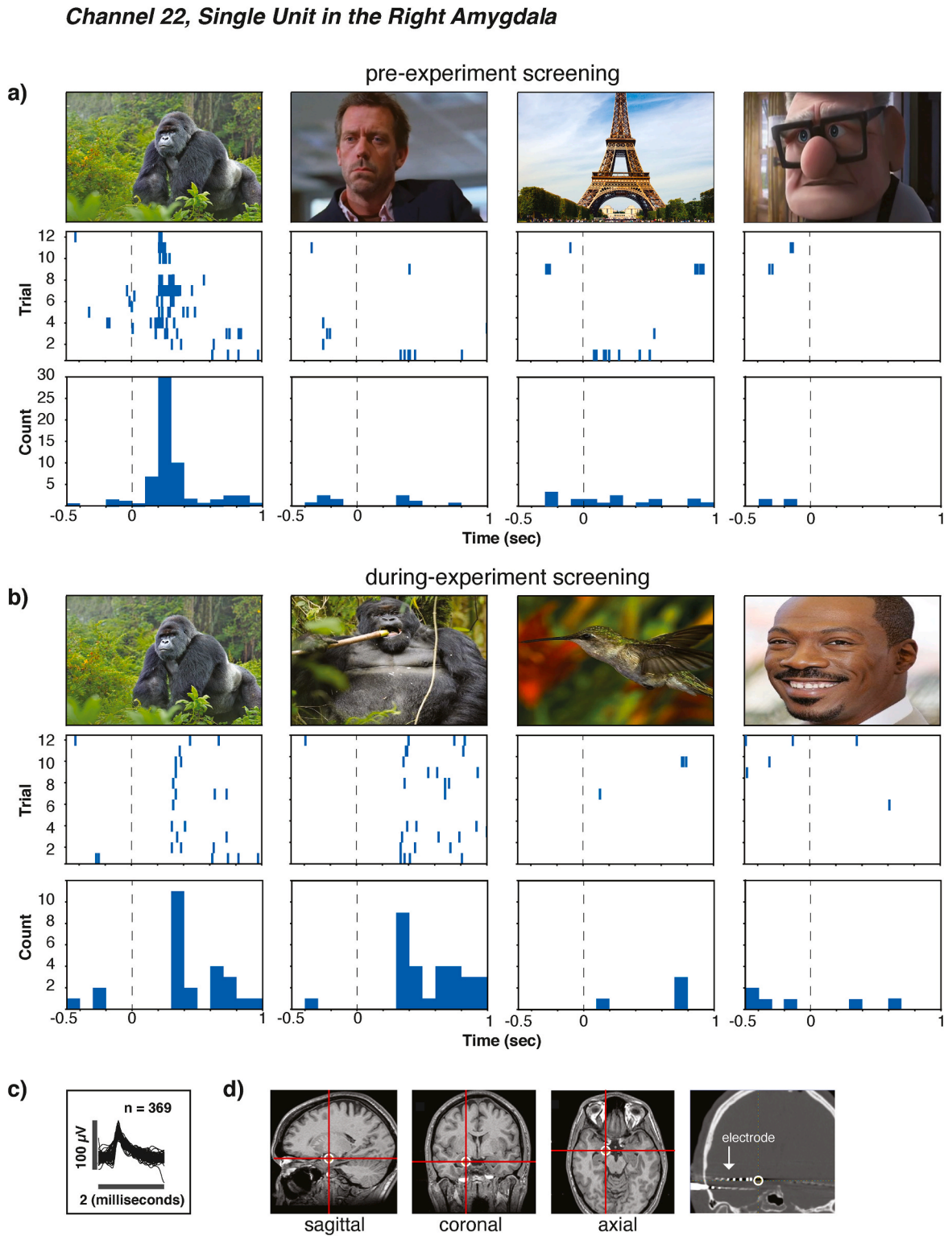


Fig. 2. Neural responses to the gorilla during screenings. Peristimulus time histogram (PSTH) of the a) pre-experiment screening and b) during-experiment screening. We show the response to a single gorilla image and non-gorilla images (selected arbitrarily) from the pre-experiment screening and to two invariant representations of a gorilla in the during-experiment screening. In the during-experiment screening we showed images of other animals (bird in the example here) and people (comedian Eddie Murphy in the figure) that did not yield a response. c) spike shapes from the activity in the during-experiment screening. d) illustration of electrode implant site for the right amygdala neuron based on sagittal, coronal, and axial CT and MRI scans fusion. The MRI scans were taken after the micro-wire implantation (see right panel).

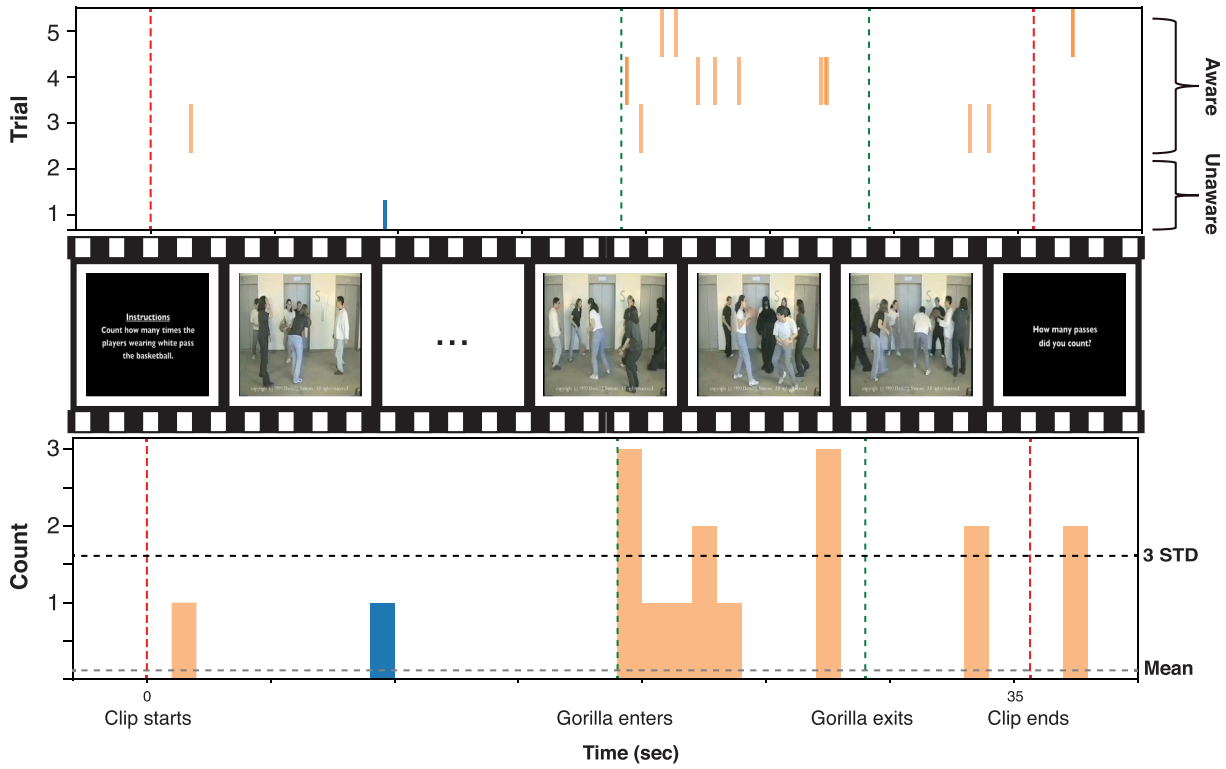


Fig. 3. Response difference between unaware/aware viewings. PSTH of the of the gorilla neuron across five clip viewings. The subject confirmed she was unaware of the gorilla during the first two exposures ("unaware" trials). Dashed lines in the histogram panel depict the mean firing rate for the unit and 3 standard-deviations above the mean. Blue indicates "unaware" trials and orange indicates "aware" trials.

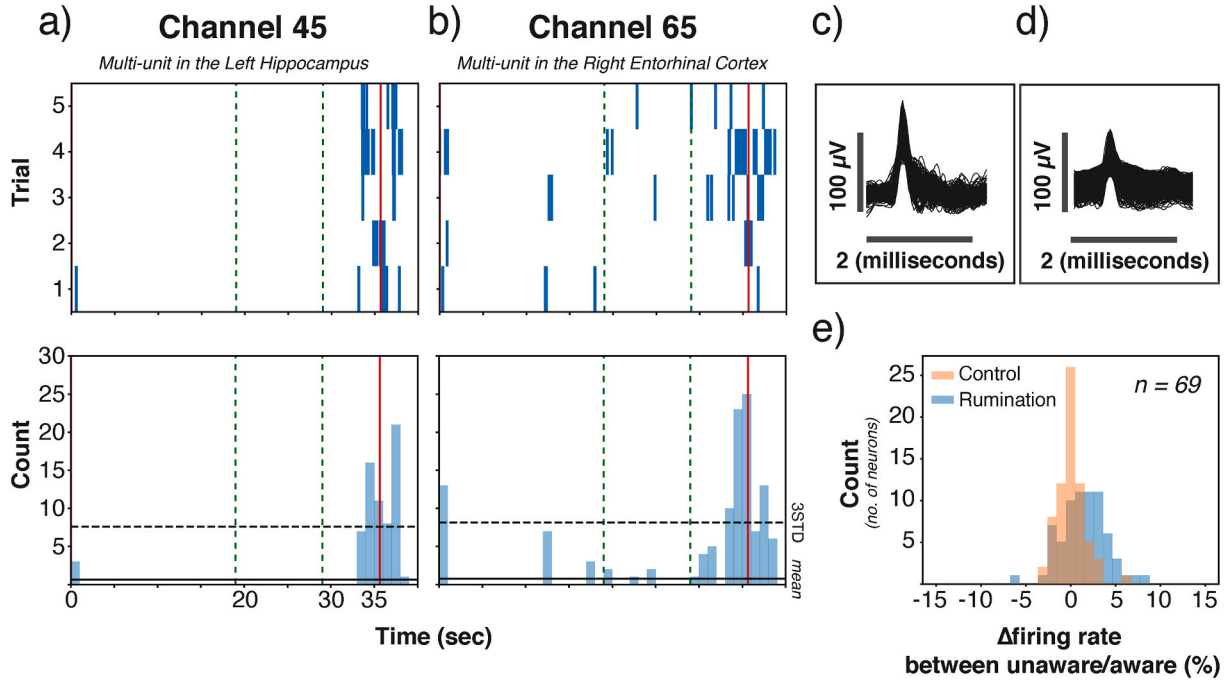


Fig. 4. Response of rumination neurons across unaware/aware viewings. a) two PSTH examples of neurons showing an increased mean response density during the rumination epoch in the unaware (1–2) and aware (3–5) trials. Spikes shapes for the single unit in channel 45 **(c)** and 65 **(d)** are shown to the right. **e)** overlapping histograms of change in mean density comparing firing difference during the rumination epoch (blue) and clip viewing epoch ("control"; orange). Results are statistically significant for both: 1) increase in firing rate across conditions, and 2) increase in firing rate within condition (between epochs).

properties that were different from the other responsive neurons. The clip-gorilla showed sustained response over a duration of 253 ± 202 ms, and a response onset (449 ± 192 ms) that was about 80 ms slower than that of the invariant gorilla neuron. The firing rate increased above the baseline (0.83 ± 1.84 Hz; total number of spikes during the entire session: 2655) when the clip-gorilla was visible (Fig. 5). Investigating the

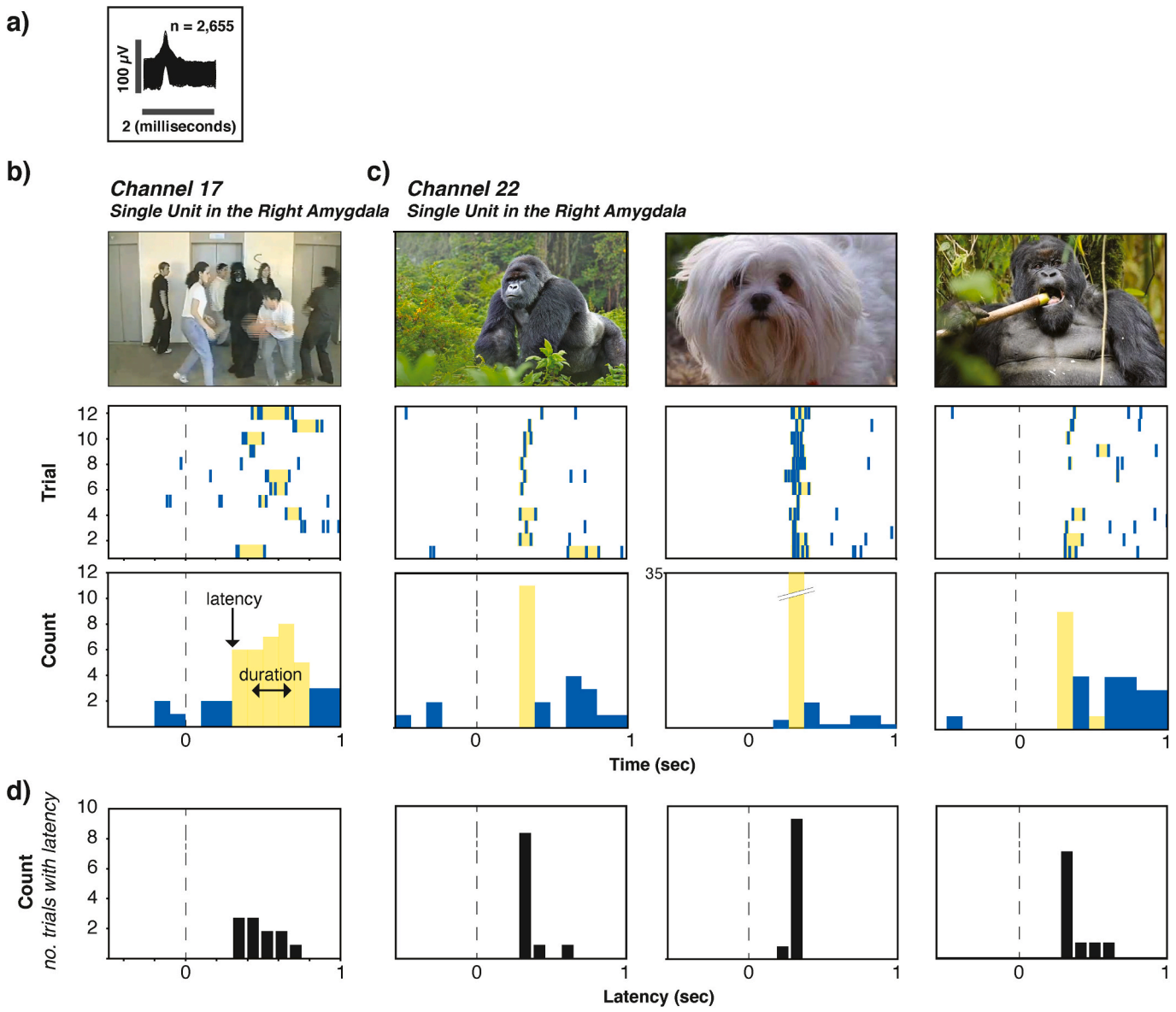


Fig. 5. A specific response distinguishing the clip-gorilla from all gorillas. a) spike shape of the clip-gorilla unit. **b)** response of the clip-gorilla unit to the still image from the clip in the during-experiment screening. Yellow highlights indicate the latency (onset of the response firing rate above the mean, within a 100-ms bin) and the duration until the firing rate declines below the threshold. We highlight in the bar graph below the onset latencies across all 12 trials. **c)** PSTH of the gorilla neuron across 3 responsive images (same as Fig. 2) illustrating the narrow onset window. **d)** count of the onset that start in each latency showing the distribution difference between the clip-gorilla and the gorilla neuron responses.

response properties of the clip-gorilla, we noted that the onset latency in each trial varied more than the typical latency expected by invariant representation neurons (see Mormann et al., 2008 for meta-analysis of such invariant neurons). The variance of the onsets was significantly different when comparing the clip-gorilla and the gorilla neuron's response onsets ($\chi^2(3) = 10.774$, $p = 0.013$; *Bartlett's test*). However, a comparison of each trial's mean latency between the clip-gorilla neuron and the gorilla neuron was not significant ($F(3,37) = 1$, $p = 0.402$; *ANOVA*). A wide range of onsets could suggest an interaction with a network or an additional processing (i.e., learning that occurs across multiple sites). Importantly, the clip-gorilla neuron did not respond to other images from the clip (i.e., still images of the clip without the gorilla) or to other images of gorillas or animals that were not the one shown in the clip. While we termed this neuron "clip-gorilla" we cannot rule out that it may reflect other subjective experiences (i.e., a "reveal of a surprise" neuron, or a neuron coding the "qualia of insight").

3.4. Suggested hierarchical coding of information

Given the adjacency of the two neurons in the right amygdala and the fact that their response properties reflect varying degrees of granularity of information (a response to *any* gorilla, versus a response to a *specific* gorilla in a recently viewed clip), we further investigated the relationship between the two neurons. Cross-correlation of the firing latency of the two neurons did not yield a significant lag. That is, we could not detect a clear time order between the two neurons that is suggestive of a cascade of activities. This may partially be due to the small number of spikes across all trials (of which two were "unaware" trials showing nearly no spiking activity in the gorilla neuron).

However, when comparing the spike distance (absolute time interval between spikes, across two neurons; see supplementary materials for details) between the gorilla neuron and the clip-gorilla neuron to the spike distance of any other 52 neurons (69 responsive ones in the study, of which 17 had no spikes during the investigated interval) we noted

that the gorilla/clip-gorilla distance was significantly lower than the distance between the gorilla neuron and any other neuron (Fig. 6). Simply put, the clip-gorilla is most likely to fire in time proximity to the gorilla neuron compared to other neurons. While the clip-gorilla neuron has the smallest average spike distance relative to the gorilla neuron, this does not necessitate that the distance is lower than any dyad of neurons. This may suggest that the two neurons receive inputs from the same network, or that they are activated by the same process, yet not necessarily part of a cascade of information flow (i.e., that the firing of one neuron leads to the firing of the other).

3.5. Change in coherence between the neuron and the neighboring network

To further investigate the potential relationship between the network and the firing activity of the responsive neurons we looked at the local field potential near the electrode location of the gorilla neuron and the clip-gorilla neuron. The LFP is the electric extracellular potential in relatively localized populations of neurons. It is often regarded as indicative of input to the neurons that is used to trigger activation. We first looked at the LFP in the channels from the same macro-electrode as the gorilla- and clip-gorilla neurons, during unaware and aware trials. Time-frequency analysis of the LFP within these right amygdala micro-wires during the clip viewing revealed no clear difference between the moments when the gorilla was visible and other epochs. However, comparing the activity of the unaware trials to that of the aware trials showed a difference in the activity within the alpha and theta bands during the rumination epoch (Fig. 7). The focus on alpha and theta band activities was directed by prior works showing that the LFP in these bands, primarily in the MTL, is related to attention (van Diepen et al., 2016) and memory (Rutishauser et al., 2010). As increase in the spiking activity during the rumination epoch was shown across neurons in various sites – not just the right amygdala – we conducted an exploratory investigation of all neurons and all pairs of unaware/aware trials to test whether an LFP modulation occurred. Looking at the alpha band activity across 4 seconds from the onset of the rumination epoch between the unaware/aware trials, we identified sixteen units that significantly changed their mean power during the rumination epoch (Bonferroni--corrected ANOVA; see Supplementary Fig. 3 for examples of the modulation). Nine of the units that have shown significant modulation also reflected a change in the spiking activity. Similarly, fourteen units have shown a significant modulation in the theta band during the rumination epoch. Among those fourteen units, four were located in the same macro-electrode as the gorilla neurons. Together, these modulations in LFP along with spiking activity supports a population coding of information processing across regions (see Reber et al., 2017).

To further investigate the interplay between the network activity and the single neuron activity we looked at the spike-field coherence (SFC). We filtered the raw channel signal in the alpha (8–15 Hz) and theta (4–7 Hz) bands and tested the spiking phase lock (location where the spike lands on the wave, between 0 and 2 π) during both unaware and aware trials. We divided the wave into 20 18° bins and tested whether a neuron's spikes landed more in a specific bin compared to chance. The gorilla neuron spikes, during both the clip viewing and the during-experiment screening, were not significantly different than a uniform distribution ($p = 0.103$; Rayleigh test). This was possibly because of the low spike count. Nevertheless, we noted visually that the spikes were clustered in certain bins more than others (see Fig. 8a–b). Accordingly, we quantified the mean phase locking "center of mass" (CoM) as an estimate of the SFC. While the SFC for the gorilla neuron did not significantly change during the clip viewing between the unaware and aware trials, the SFC for the clip-gorilla neuron did show a significant change in CoM (Fig. 8c–d; see also supplementary clip for an illustration of the neuron's CoM change during the clip). Specifically, the mean CoM of the clip-gorilla neuron shifted during the gorilla viewing epoch from 319° to 179° ($Z = 2.329$, $p = 0.019$, *Wilcoxon rank-sum test*) in the alpha band, and from 59° to 184° ($Z = -1.869$, $p = 0.061$, *Wilcoxon rank-sum test*) in the theta band (Fig. 8d). All other epochs and frequency bands did not exhibit a significant change. As a change in phase-locked SFC suggests a reset of the frequency bands during the clip, this alignment may be indicative of engagement of attention and memory in the stimulus processing.

4. Discussion

Because of a series of unique happenstances (having access to an individual patient implanted intracranially with depth electrodes, of which one was responsive to gorilla stimuli that was, prior, shown to be depicted in a popular clip demonstrating inattentional blindness; and the patient being unfamiliar with the clip despite its popularity, and failing to notice the gorilla repeatedly), we were able to test the coding of conscious awareness in humans using single neuron recording. This allowed us to gather insights into a phenomenon that is impossible to investigate behaviorally: a human brain's response to content that the

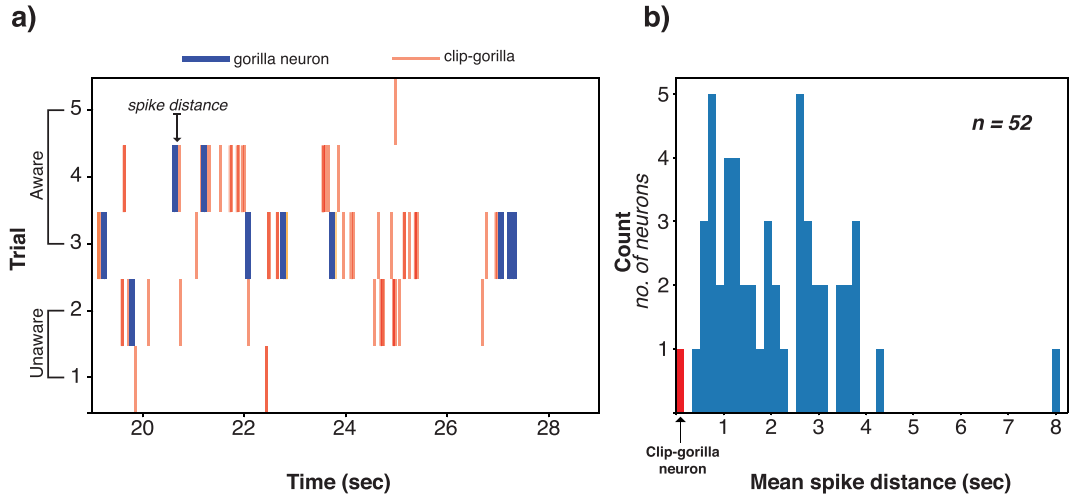


Fig. 6. Alignment between responsive neurons. a) spike depiction of the gorilla neuron (blue) and clip-gorilla (red) during the clip epoch in which the gorilla was on-screen. The time between each pair of nearby spikes in the two neurons ("spike distance") was compared to all neurons that showed spiking activity during the epoch. **b)** mean minimum spike distance between the gorilla-neuron and all other responsive neurons. Red bar marks the clip-gorilla neuron, which shows significantly higher alignment with the gorilla neuron compared to all other neurons.

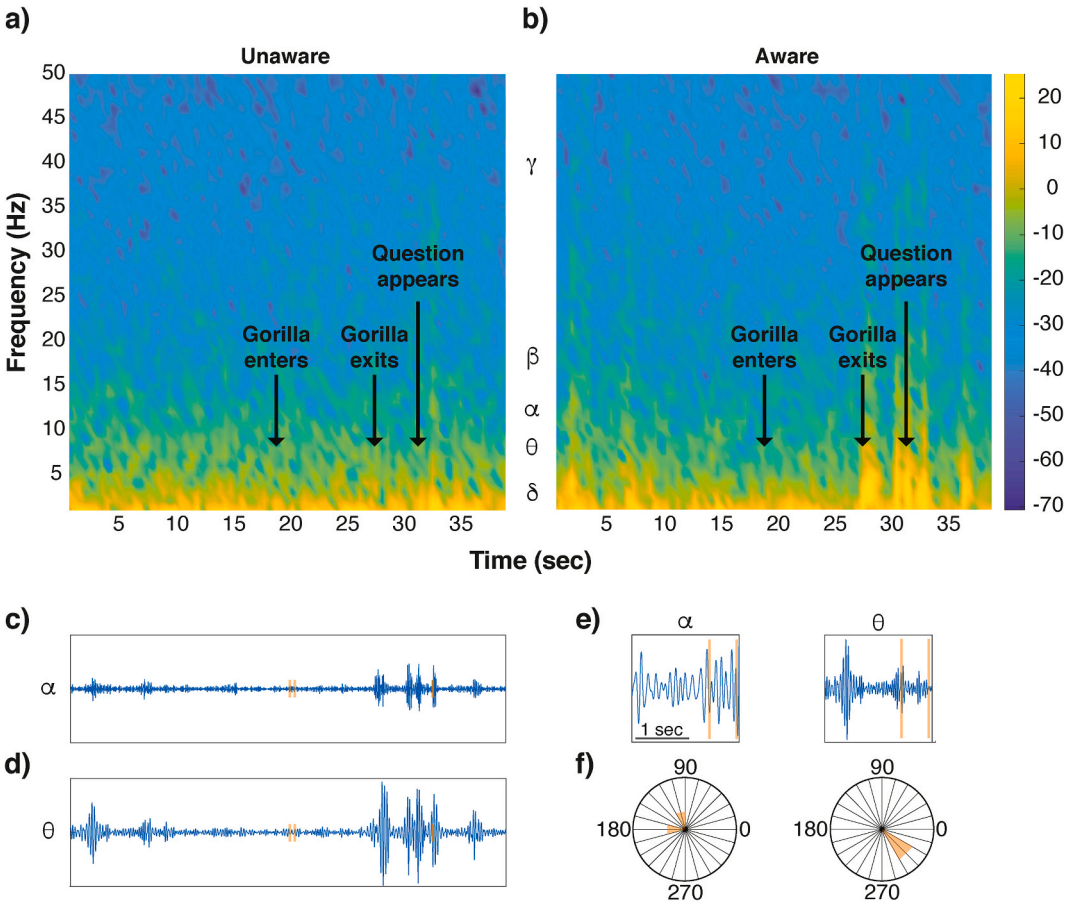


Fig. 7. Spike-field coherence calculation for unaware/aware trials. a) Time-frequency spectrogram of the channel 22 (gorilla neuron) local-field potential during the second unaware trial, and a following **b)** aware trial. The color reflects the *power* in each time-frequency (color code indicated on the right). Note the increased alpha and theta activities when the gorilla exits the screen during the aware trial. Extracting the alpha **(c)** and theta **(d)** bands from the aware trial and aligning the spikes in the trial with the local-field activity allows for detection of the spike-field coherence (see **e)** for depiction of 2 s of alpha and theta activity with two spikes aligned to the field phase. **f)** indicating the phase bin within the windows allows for a calculation of the spike-field coherence throughout the time window.

person herself does not perceptually notice.

We found a single neuron in the right amygdala that increased its firing rate when the subject saw an image of gorillas in a pre-experiment screening (Fig. 2). This neuron can be categorized as a "concept cell" (Cerf et al., 2010) that invariantly codes the notion of a gorilla. The neuron can also be characterized as an "animal" or "hairy creature" cell as it was responsive to at least one other animal (i.e., "Dog"; Fig. 5). We used the label "gorilla neuron" throughout the paper because of the focus of this work on the gorilla stimulus.

The single neuron's response was correlated with the *conscious awareness* of a gorilla in a clip (Fig. 3). That is, the neuron responded only when the subject exhibited conscious awareness of the content, and did not significantly alter its firing rate otherwise, even when the image of a gorilla was visible to the subject. The information pertaining to the existence of a gorilla presumably penetrated the subject's brain but did not rise to her conscious awareness. Put differently, the neuron responded only when the subject was able to perceptually experience the content. This result is in line with previous work showing that conscious awareness is linked with the firing of neurons in the MTL at the threshold of recognition (Quiroga et al., 2008). The previous work investigated this threshold's boundaries and suggested that exposures below dozens of milliseconds lead to no perceptual awareness and no change in firing rate. In our study the stimulus that should have evoked a response was on the center of the screen for seconds, occupied a sizeable part of the screen, and actively made itself noticeable - yet the neuron did not significantly shift its firing rate. Our results are also in line with another work showing an awareness-related activity modulation of single neurons in the MTL (Reber et al., 2017). The recent work suggests that the latency and strength of neuronal firing correlates with conscious perception, and that the responses align with an anatomical gradient of increased modulation. Our work offers an extension to this previous investigation in the form of a dynamic and continuous stimulus, and an attention manipulation that is driven by competing distraction (rather than consecutive ones).

The gorilla neuron's response properties - upon viewing still images of gorillas in a screening conducted between the clip viewings - were in line with those seen in previous work, showing that: 1) responses to animals are prominent in the right amygdala (Mormann et al., 2011), and 2) that the responses occur within about 397 ms from image onset (see Mormann et al., 2008; our results showed response latencies ranging from 370 ms to 498 ms; Table 2).

In addition to the change in the gorilla neuron between the unaware and aware trials, we identified a cluster of neurons, scattered across numerous sites, that altered their firing activity after the clip was over and the subject reflected on the recent clip viewing (Fig. 4). Those neurons significantly increased their firing rate between the unaware and aware trials, ostensibly as the subject shifted from counting basketball throws to thinking about the clip. We termed those neurons "rumination neurons" as they changed the firing properties primarily during the epoch where the subject was reflecting on the task (albeit nearly half also exhibited a visible, yet not statistically significant, change in firing rate in the beginning of the clip). Notably, the rumination neurons increased their firing rate above baseline *both* during the unaware trials and the aware trials (yet, significantly more so during the

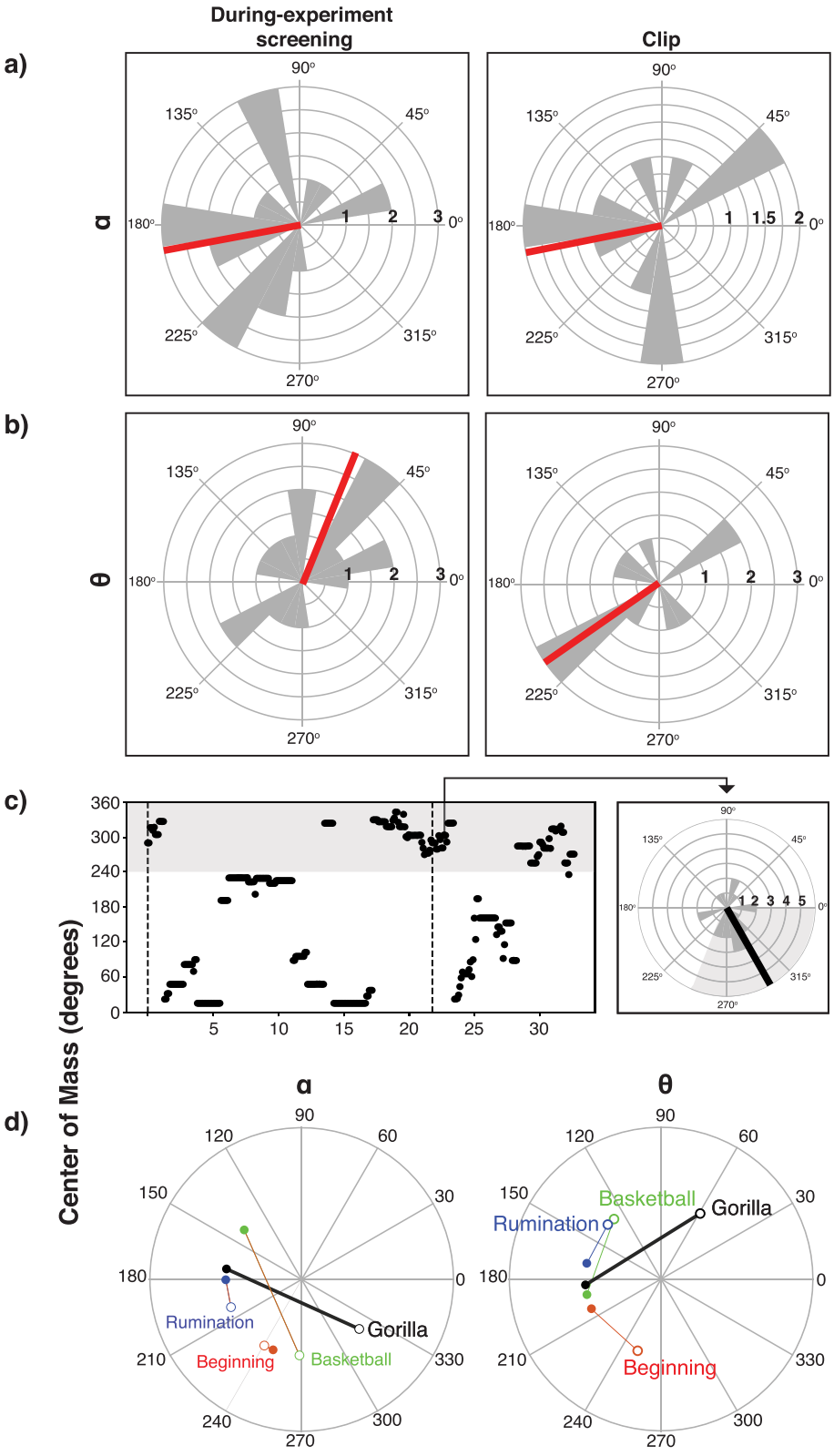


Fig. 8. Change in spike-field coherence during the viewing when the gorilla is visible. a) spike phase on the LFP alpha band during the 12 viewings of the gorilla image in the during-experiment screening (**left**) and the gorilla epoch in the clip's "aware" trials (**right**), for the gorilla neuron. The 20 bins correspond to the phase angles 0°-18°, 18°-36°, …, 342°-360°. Each bin reflects the spike count that landed in the specific phase. The red lines mark the Center of Mass (mean angle for the period). **b)** the spike-field coherence and center of mass for the same epochs as **(a)** at the theta band. **c)** the center of mass of the clip-gorilla neuron during the entire clip viewing. Each dot corresponds to a single 3-s window (100 ms step) of alpha band center of mass. The dots are pooled across all trials (both unaware and aware ones). Grey shaded area marks the prominent set of degrees (240°-360°) where the majority of rumination neurons showed a shift in center of mass. On the **right** is an expansion of a single dot. **d)** mean center of mass for each epoch in the clip (beginning, basketball passes, gorilla, rumination) during unaware (empty circles) and aware (filled circles) trials for the alpha and theta bands. Thick black lines denote significant change in phase ($p < 0.05$) during the gorilla epochs.

aware trials). The change in response patterns between unaware and aware trials aligns with prior studies of human single neuron, which show that a shift in locus of attention is often accompanied by an in increased firing rate of task-responsive neurons (Cerf et al., 2010).

In addition to the neurons that altered their response properties with conscious awareness, we identified a neuron (likely located in proximity to the gorilla neuron, estimated by the furthest possible spread of the micro-wires within the macro-electrode bundle; see Fig. 1) that fired significantly solely to the specific gorilla depicted in the clip, and not to other gorilla images (Fig. 5). This neuron did not respond to images of gorillas in previous screenings, nor did it respond the gorilla in the clip during unaware trials (yet, it did significantly increase its firing rate

during the aware trials). In the during-experiment screening, the neuron increased its firing rate in response to the specific gorilla depicted in the clip when the subject became aware of its existence. The response was not significant at the 3 standard-deviations cutoff, yet we report the neuron's properties as it, visually, seemed to alter its firing properties (Supplementary Fig. 2). The reason the responses do not reach significance may be due to the low firing rate in the first two exposures of the image (out of 12), when the firing rate was still low as the subject gradually became aware of the missed gorilla. This "aha" moment, when a person recognizes that they were blind to content that was shown in front of their eyes prior, is nearly impossible to replicate (see, however, Aziz-Zadeh et al., 2009), especially with single neuron studies. Once a subject is made aware of the stimulus, they presumably will be unable to both unsee it and be surprised by its occurrence in the future. We cannot rule out the possibility that the clip-gorilla response may reflect a cognitive processing that is not related to the appearance of the gorilla itself but rather to parallel learnings that occurred along with the reveal of the gorilla in the clip (e.g., focused attention on the novel stimulus, or an amusement from the evident prior detection failure).

As both the gorilla concept neuron and the clip-gorilla learning neuron related to the conscious identification of the gorilla and to the transition between awareness states, we investigated whether the neurons are connected by a direct hierarchical representation (i.e., a cascade of actions that depicts a gorilla in various levels of abstraction; see discussion in Quiroga et al., 2009). We did not find any significant cross-correlation lag between the neurons which would have offered strong support to an information cascade hypothesis (i.e., that one neuron responds to "all gorillas" and another to a "specific gorilla"). However, we did find a strong spike-distance relationship between the two neurons (at times one fires before the other, and at times the order is reversed, but the firing time difference is always shorter and more aligned than that of any other neuron; Fig. 6). Recent works investigating how network activity in one region can be modulated by interconnected sites in order locations has suggested that such response characteristics can be the output of circuits that are distributed across functionally and anatomically separated regions, but not hierarchically organized (i.e., via recurrent associated networks; Perich & Rajan, 2020).

Given the seeming interplay between various circuits in conscious processing we wondered whether a speculative model of consciousness as a stream of information that is processed first in early circuits (i.e., sensory processing), before being elevated to our awareness (Tononi, 2008), may be aligned with our results. In order to shed light on this theory we investigated the changes in activity at the network level (LFP modulation). The LFP is said to be the aggregate of dendritic inputs which may reflect an early processing of information leading to the activation of cells like the gorilla neuron upon crossing a threshold indicative of awareness. We investigated the LFP modulation between the aware and unaware trials. Time-frequency depiction of the LFP activity between unaware and aware trials (collapsing across the two unaware trials and comparing to the aggregated aware trials) showed a notable difference in alpha and theta bands. Our exploratory investigation focused on changes in alpha and theta bands because those bands are often implicated with modulation of attention and memory in the MTL. Prior works looking at alignment of spikes and LFP in the MTL suggest that increased phase locking in the theta band may be linked to enhanced memory (Rutishauser et al., 2010) and that modulation of alpha activity is triggered by shifts in attention (Klimesch, 2012; Sauseng et al., 2005; van Diepen et al., 2016).

We identified a significant modulation in the LFP signal in various sites between unaware/aware trials in the rumination epoch (Fig. 7 and Supplementary Fig. 3). An exploratory test of the LFP modulation within the alpha and theta bands, in the micro-wires pertaining to the rumination neurons, showed that 9 neurons (19%) in the alpha band, and 17 neurons (36%) in the theta band (with 8 neurons, 17%, in both bands) exhibited modulation. Five of the neurons in the macro-electrode within which the gorilla and clip-gorilla neurons were identified also showed significant theta activity modulation. As the LFP modulation investigation was exploratory we did not have a hypothesis as to the nature of the interaction between the rumination neurons' modulation and the task. However, given that a shift in properties between unaware/aware trials occurs both at the single neuron level as well as the LFP, we suggest that the change could be driven by: 1) increased attention during the aware trials (in line with the significant change in alpha frequency during the rumination epoch), 2) memory activation (as the subject reflects on the clip she has just seen; in line with the theta frequency modulation), 3) a potential decrease in engagement with the task at the conclusion of a demanding focus (Barnett and Cerf, 2016), or 4) a mere process of calculation of the output number (Kutter et al., 2018). Each of those explanations would align with our results and support an alternative hypothesis.

Linking the spiking activity and network activity using the spike-field coherence initially did not show significant periodicity. However, focusing the analyses on the center of mass of the spikes (the mean alignment between the network and the neuron in absence of changes to the network in the form of "reset" or other external modulations) showed a change between unaware and aware trials, during the epoch where gorilla is visible on-screen (Fig. 8).

4.1. Speculative theoretical interpretation

4.1.1. Hierarchical representation of information

Our results align with a number of interpretations that relate to prior works. First, we explore the possibility of explaining the interplay between the gorilla neuron, clip-gorilla neuron, and the various network modulations as indicative of an aggregated coding of information.

Given the unique response characteristics of the clip-gorilla neuron (only firing for the clip-gorilla and not to the invariant representation of a gorilla) we suggest that the neuron may code a specialized representation of the gorilla (i.e., isolating the unique gorilla from a gorilla concept). Indeed, the response properties (latency, duration, firing rate; Fig. 5) of the clip-gorilla neuron were different than that of a typical response of a concept cell (i.e., dense firing activity centered on a small duration of less than 100 ms; see Mormann et al., 2008). An interpretation of the difference in properties could be that the clip-gorilla neuron is part of a circuit that distinguishes a specific entity from its larger category and manifests the recognition of the importance of a single prototype from the archetype (Rey et al., 2018). The need for specialized neurons that depict concepts in various levels of granularity and resolutions suggests an analogy between concepts cells identified in the MTL and place cells in rodents that show coding of locations across various resolutions (Solstad et al., 2006). Rodent electrophysiology, in the context of place and grid cells, has suggested that clusters of neurons are organized such that they code location information in varying scales, ultimately forming a grid that maps to a single place cell. An analog in human concept cells suggests that clusters of neurons may code the existence of, say, animals, gorillas, up to the high resolution of a specific clip-gorilla. The consistent short spike distance between the gorilla neuron and clip-gorilla neuron (Fig. 6) provides support for such an interplay between clusters. Recent works, in single neuron recordings in humans conducting navigation tasks, have observed similar results in the domain of spatial coding (Herweg et al., 2020) and aggregated coding (Kunz et al., 2019). These works also identified an interplay between the LFP and spiking activity similar to the one shown here (Chen et al., 2018).

An alternative interpretation of the clip-gorilla neuron's distinguished response pattern could be that it reflects processing of novel learning (i.e., the failure to detect a stimulus in prior trials) and committing to register this information. Such an explanation aligns with our results yet suggests an interpretation that is not driven by hierarchical processing, but rather by independent parsing of the stimuli (i.e., one neuron responds to, say, animals, whereas the other to error in

content detection).

4.1.2. Labeling of rumination neurons' responses

Second, we speculate on the nature of the rumination neurons' responses. We termed these neurons rumination neurons as their activity significantly increased when the clip ended and the subject was asked to reflect on the clip. However, the epoch could also be labeled differently. As the response of the rumination neurons occurred both during the unaware and aware states, the post-clip epoch likely did not reflect counting basketball throws (as the subject already knew this is not truly the purpose of the task in the aware trials), but rather some other contemplative process. We do not have a clear hypothesis as to the role of these neurons in the population coding, but note that their large number suggests that they may code a generic contribution to processing (i.e., considering a recent past experience).

The rumination neurons' increased response to the stimuli during aware trials compared to unaware trials (which was already above the baseline activity of the neurons) could be explained by an inhibition of the subject's concentrated focus on counting basketball throws during aware trials. That is, under conditions of focus on a singular goal, alternative neural responses may suffer, but not entirely disappear (perhaps requiring higher thresholds to trigger; Barnett and Cerf, 2017).

As the correct detection of the number of basketball throws is linked to performance in the task during the unaware trials, an alternative interpretation of the rumination could be that the subject is merely highly engaged with the accuracy and ensuring no error. Indeed, prior work on mental accounting in the context of reward and financial decision-making has shown that single neurons activity is linked to value coding (albeit, in the prior works the focus was on nucleus accumbens responses; Patel et al., 2012).

Further investigation of the conditions that trigger the rumination neurons could help characterize their properties. Our experiment was not tailored for an investigation of those neurons and their emergence was the outcome of exploratory investigation.

4.1.3. Conscious awareness is driven by an interplay between the network and neuron thresholds

Finally, one could describe the entire set of results as the output of a process in which information that penetrates the brain is aggregated in early sensory circuits before being projected to the MTL where it emerges as a concept. The awareness of the incoming information is driven by an interaction between the network and the single neurons' coding the concepts. Cognitive demands from the network, distractions, or bypassing of the processing limit (such that information does not reach the threshold of perception) maintain the unaware experience. This suggested framework for consciousness aligns with our results and with existing models of consciousness (Tononi, 2008) and visual exploration (Einhäuser et al., 2009; Mackay et al., 2012). The framework also makes concrete predictions. For example, the prediction that stimulation of the network in the site corresponding to the concept neuron would make a person notice the gorilla earlier, or that reset of the network such that the CoM of a spike occurs within the 240°-360° of alpha bands will increase the likelihood of gorilla detection. More speculative, the results suggest that changes to the network may alter the conscious perception of content altogether, or alternatively, make it hard to attend to a task that demands resources that are not in line with the concept drawing the attention. These testable hypotheses are also supported by previous models of attention (Reynolds and Heeger, 2009).

4.2. Limitations

A notable limitation of the study emerges from the fact that it was conducted on a single individual. Despite the large number of neurons used and the robust results spread across multiple regions, epochs, and trials, we recognize that drawing any conclusions about population effects from a single example is challenging. While results in electrophysiology do typically require a similar number of neurons to derive a conclusion, established norms in neuroscience suggest that at least two primates are required for the results to be considered indicative of a pronounced effect.

Additionally, any study involving recording of single neurons in humans is also limited in that it observes only a small subset of regions in the brain and tries to draw conclusions about the entirety of a circuit. Our study is no different. Given that the work involves clinical patients, and that the choice of recording locations is determined solely by clinical requirement, we are limited in neural locations investigated. A future study with access to neurons in both the visual cortex and amygdala may be better able to trace consciousness - i.e., understand where visual stimuli enter and are ultimately lost in the unaware states.

Still, given the unique intersection of rare events required for such an experiment, we believe our findings to be valuable. We are not alone, as sufficient evidence of the value of single study neurobiological experiments exists for rare patients (see for example Adolphs et al., 1994).

Further, while our results speak to various psychological effects (i.e., rumination, learning, awareness) it is noteworthy that those labels are characterizing the responses post-hoc without true knowledge of whether those are indeed the cognitive processes effective in the brain. As all neural results that are drawn from activation and response in neuroscience are effectively a correlation between effect and neural responses, this limitation is true for the majority of neural works, but should be noted explicitly in our work since it is more prominent here. Specifically, it is not clear that the neuron we labeled, say, as "gorilla neuron" did not actually code a different concept (i.e., animals with fur) of which our gorilla was merely an exemplar. Similarly, the rumination neurons could in fact reflect a response to text appearing on the screen in white font. Since we did not test the reversed causation of the response (the triggering of the neuron leading to a subjective experience of text by the subject, for example) or a broader set of options that could be invariant triggers, we note that the labels we suggest should be seen as merely ways to create a systematic description that code the psychological experience. It is possible that other explanations could yield other equally valid predictions using the same data.

Finally, it is noteworthy that we reported some results despite the fact that they were not significant at the common criteria of numerous psychological works. At times, those were marginally significant, yet we posited that it was worth mentioning them despite the inconclusive clarity on their reliability, allowing the reader to reflect on their prominence. We do so since the study depicts an unusual acquisition circumstance which is unlikely to occur frequently and may benefit the reader despite uncertainty regarding the results' robustness.

4.3. Future work

In addition to improving upon the above limitations we suggest that future work should test the boundaries of our findings. Specifically, if conscious awareness is a function of the coherence between the individual neuron and its neighboring network circuits, then assessment of the interaction of the two could be predictive of deviations for normative states of consciousness (i.e., sleep, vegetative states, or even consciousness among animals with similar neural circuits).

Additionally, applications of the learning on network's ability to code hierarchical attention allocation could draw parallels in non-neural networks. Specifically, myriad examples exist in which networks (i.e., business organizations, schools of fish, or groups of humans exhibiting complex social dynamics) procure, aggregate and process information in a hierarchical fashion, yet fail to detect notable salient events (Couzin et al., 2005; Mentovich and Cerf, 2014). Might such networks suffer from inattentional blindness that parallels the one exhibited by our subject? And if that is the case – can the learnings from our work shed light on how to improve the internal dynamics of these networks? Finally, as invasive tools for neuroscience research are improving (Shachar et al., 2012), it may be possible to replicate studies like this one

with a growing population of individuals.

5. Conclusion

This work shows direct evidence of single neuron correlate of inattentional blindness in humans. It has not escaped our notice that the study offers an insight into a phenomenon that touches on the dynamics between our conscious experience of reality and the way it is reflected in our brain. Neural pathways that code the objective world are translated in our neural circuits through a cascade of processes to a perceptual representation that is altered by various mechanisms that we term "awareness" or "attention". The unlikely circumstances within which this experiment occurred (recording directly from neurons coding the experience of awareness in humans) allowed us to shed light on part of the perceptual processing. Contemporary endeavors that speak to the implantation of chronic electrodes in the brains of humans for non-clinical reasons may offer further insights into the qualitative experience of consciousness in humans, bridging neuroscience and psychology in a way that truly helps leverage our understanding of the subjective phenomenon of perception. Whereas psychological science often relies on reports and behavior in order to garner insights into the human psyche, the addition of neuroscience at the level of individual neurons can expose a fabric of understanding that cannot be accessed otherwise and reveal our unconscious properties to the scientific eye.

Author contributions

MC designed the study and collected the data. BF and MC analyzed the data and wrote the manuscript.

Funding

The authors received no funding or support for this work.

Data availability

The data and analyses codes (as *Jupyter* notebook) are available on www.morancerf.com/publications. Details on the data structure are available in a "Data in Brief" article titled "Human single neuron inattentional blindness data" (*Neuropsychologia* under preparation).

Declaration of competing interest

The authors declare no conflict of interest in the work.

Acknowledgment

The authors wish to thank the patient for her eager participation and insights during the study, which notably helped shape the research learning. Additionally, we wish to thank the anonymous reviewers for their valuable comments.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2021.108111.

References

Adolphs, R., Tranel, D., Damasio, H., Damasio, A., 1994. Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature 372 (6507), 669–672.
Aziz-Zadeh, L., Kaplan, J.T., Iacoboni, M., 2009. "Aha!": the neural correlates of verbal insight solutions. Hum. Brain Mapp. 30 (3), 908–916.
Baars, B.J., 1997. In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. J. Conscious. Stud. 4 (4), 292–309.
Barnett, S., Cerf, M., 2016. Method for Measuring Engagement. USPTO.
Barnett, S., Cerf, M., 2017. A ticket for your thoughts: method for predicting content recall and sales using neural similarity of moviegoers. J. Consum. Res. 44 (1), 160–181.
Cartwright-Finch, U., Lavie, N., 2007. The role of perceptual load in inattentional blindness. Cognition 102 (3), 321–340.
Cerf, M., Thiruvengadam, N., Mormann, F., Kraskov, A., Quiroga, R.Q., Koch, C., Fried, I., 2010. On-line, voluntary control of human temporal lobe neurons. Nature 467 (7319), 1104–1108. https://doi.org/10.1038/nature09510.
Chalmers, D., 2007. The Hard Problem of Consciousness. The Blackwell Companion to Consciousness, pp. 225–235.
Chen, D., Kunz, L., Wang, W., Zhang, H., Wang, W.-X., Schulze-Bonhage, A., Reinacher, P.C., Zhou, W., Liang, S., Axmacher, N., 2018. Hexadirectional modulation of theta power in human entorhinal cortex during spatial navigation. Curr. Biol. 28 (20), 3310–3315.
Couzin, I.D., Krause, J., Franks, N.R., Levin, S.A., 2005. Effective leadership and decision-making in animal groups on the move. Nature 433 (7025), 513–516.
Crick, F.C., Koch, C., 2005. What is the function of the claustrum? Phil. Trans. Biol. Sci. 360 (1458), 1271–1279.
Einhäuser, W., Schumann, F., Vockeroth, J., Bartl, K., Cerf, M., Harel, J., Schneider, E., König, P., 2009. Distinct roles for eye and head movements in selecting salient image parts during natural exploration. Ann. N. Y. Acad. Sci. 1164, 188–193. https://doi.org/10.1111/j.1749-6632.2008.03714.x.
Herweg, N.A., Kunz, L., Brandt, A., Wanda, P.A., Sharan, A.D., Sperling, M.R., Schulze-Bonhage, A., Kahana, M.J., 2020. A Learned Map for Places and Concepts in the Human MTL. BioRxiv.
Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. Trends Cognit. Sci. 16 (12), 606–617.
Koch, C., 2012. Consciousness: Confessions of a Romantic Reductionist. MIT press.
Koch, C., Tsuchiya, N., 2007. Attention and consciousness: two distinct brain processes. Trends Cognit. Sci. 11 (1), 16–22.
Koivisto, M., Hyönä, J., Revonsuo, A., 2004. The effects of eye movements, spatial attention, and stimulus features on inattentional blindness. Vis. Res. 44 (27), 3211–3221.
Kunz, L., Wang, L., Lachner-Piza, D., Zhang, H., Brandt, A., Dümpelmann, M., Reinacher, P.C., Coenen, V.A., Chen, D., Wang, W.-X., 2019. Hippocampal theta phases organize the reactivation of large-scale electrophysiological representations during goal-directed navigation. Science Advances 5 (7), eaav8192.
Kutter, E.F., Bostroem, J., Elger, C.E., Mormann, F., Nieder, A., 2018. Single neurons in the human brain encode numbers. Neuron 100 (3), 753–761.
MacIver, M.A., Schmitz, L., Mugan, U., Murphey, T.D., Mobley, C.D., 2017. Massive increase in visual range preceded the origin of terrestrial vertebrates. Proc. Natl. Acad. Sci. Unit. States Am. 114 (12), E2375–E2384.
Mack, A., Rock, I., 1998. Inattentional Blindness. MIT press.
Mackay, M., Cerf, M., Koch, C., 2012. Evidence for two distinct mechanisms directing gaze in natural scenes. J. Vis. 12 (4), 9.
Majerus, S., Gill-Thwaites, H., Andrews, K., Laureys, S., 2005. Behavioral evaluation of consciousness in severe brain damage. Prog. Brain Res. 150, 397–413.
Mentovich, A., Cerf, M., 2014. A psychological perspective on punishing corporate entities. In: Regulating Corporate Criminal Liability. Springer, pp. 33–45.
Mormann, F., Kornblith, S., Quiroga, R.Q., Kraskov, A., Cerf, M., Fried, I., Koch, C., 2008. Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. J. Neurosci. 28 (36), 8865–8872. https://doi.org/10.1523/JNEUROSCI.1640-08.2008.
Mormann, Florian, Dubois, J., Kornblith, S., Milosavljevic, M., Cerf, M., Ison, M., Tsuchiya, N., Kraskov, A., Quiroga, R.Q., Adolphs, R., Fried, I., Koch, C., 2011. A category-specific response to animals in the right human amygdala. Nat. Neurosci. 14 (10), 1247–1249. https://doi.org/10.1038/nn.2899.
Patel, S.R., Sheth, S.A., Mian, M.K., Gale, J.T., Greenberg, B.D., Dougherty, D.D., Eskandar, E.N., 2012. Single-neuron responses in the human nucleus accumbens during a financial decision-making task. J. Neurosci. 32 (21), 7311–7315.
Perich, M.G., Rajan, K., 2020. Rethinking brain-wide interactions through multi-region 'network of networks' models. Curr. Opin. Neurobiol. 65, 146–151.
Qiu, J., Li, H., Jou, J., Liu, J., Luo, Y., Feng, T., Wu, Z., Zhang, Q., 2010. Neural correlates of the "Aha" experiences: evidence from an fMRI study of insight problem solving. Cortex 46 (3), 397–403.
Quiroga, R Quian, Mukamel, R., Isham, E.A., Malach, R., Fried, I., 2008. Human single-neuron responses at the threshold of conscious recognition. Proc. Natl. Acad. Sci. Unit. States Am. 105 (9), 3599–3604.
Quiroga, R Quian, Reddy, L., Kreiman, G., Koch, C., Fried, I., 2005. Invariant visual representation by single neurons in the human brain. Nature 435 (7045), 1102–1107.
Quiroga, Rodrigo Quian, Kraskov, A., Koch, C., Fried, I., 2009. Explicit encoding of multimodal percepts by single neurons in the human brain. Curr. Biol. 19 (15), 1308–1313.
Reber, T.P., Faber, J., Niediek, J., Boström, J., Elger, C.E., Mormann, F., 2017. Single-neuron correlates of conscious perception in the human medial temporal lobe. Curr. Biol. 27 (19), 2991–2998.
Remington, A., Cartwright-Finch, U., Lavie, N., 2014. I can see clearly now: the effects of age and perceptual load on inattentional blindness. Front. Hum. Neurosci. 8, 229.
Rey, H.G., De Falco, E., Ison, M.J., Valentin, A., Alarcon, G., Selway, R., Richardson, M. P., Quiroga, R.Q., 2018. Encoding of long-term associations through neural unitization in the human medial temporal lobe. Nat. Commun. 9 (1), 1–13.
Reynolds, J.H., Heeger, D.J., 2009. The normalization model of attention. Neuron 61 (2), 168–185.

Rutishauser, U., Ross, I.B., Mamelak, A.N., Schuman, E.M., 2010. Human memory strength is predicted by theta-frequency phase-locking of single neurons. Nature 464 (7290), 903–907.
Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., Gruber, W.R., Birbaumer, N., 2005. A shift of visual spatial attention is selectively associated with human EEG alpha activity. Eur. J. Neurosci. 22 (11), 2917–2926.
Shachar, J., Chen, T., Farkas, L., Wu, W., Zimmerman, K., Cerf, M., Marx, B., Johnson, D., Farkas, L., 2012. Brain Retractor Apparatus for Measuring and Predicting Electrophysiological Parameters. USPTO.
Shafto, J.P., Pitts, M.A., 2015. Neural signatures of conscious face perception in an inattentional blindness paradigm. J. Neurosci. 35 (31), 10940–10948.
Simons, D.J., Chabris, C.F., 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception 28 (9), 1059–1074.
Solstad, T., Moser, E.I., Einevoll, G.T., 2006. From grid cells to place cells: a mathematical model. Hippocampus 16 (12), 1026–1031.
Tononi, G., 2008. Consciousness as integrated information: a provisional manifesto. Biol. Bull. 215 (3), 216–242.
van Diepen, R.M., Miller, L., Mazaheri, A., Geng, J.J., 2016. The role of alpha activity in spatial and featured-based attention. Eneuro.
Waydo, S., Kraskov, A., Quiroga, R.Q., Fried, I., Koch, C., 2006. Sparse representation in the human medial temporal lobe. J. Neurosci. 26 (40), 10232–10234.
Wegener, D., Freiwald, W.A., Kreiter, A.K., 2004. The influence of sustained selective attention on stimulus selectivity in macaque visual area MT. J. Neurosci. 24 (27), 6106–6114.

SUPPLEMENTARY MATERIALS

for

SINGLE NEURON EVIDENCE OF INATTENTIONAL BLINDNESS IN HUMANS

Brandon Freiberg and Moran Cerf

Includes:

- Detailed methods.
- Supplementary table.
- Supplementary figures.
- Supplementary clip.
- Supplementary analyses codes.

DETAILED METHODS

Spike sorting

We used the *wave_clus* Matlab toolbox for the spike detection and sorting. We initially used the default parameters from *wave_clus* version 3.0.3 for the spike detection (namely, the parameters initialized using the *set_parameters_default.m* script; i.e., high/low pass filtering of 300/3000Hz, minimal standard-deviation for spike detection of 5, and 50 standard-deviations for artifact detection, etc.). Following, we, initially, used the *wave_clus* default sorting code for spike clustering. After the initial unsupervised clustering, we manually clustered all units using the *wave_clus* user interface. While the criteria for the manual clustering is qualitative, we used the following guidelines in the process: 1) we opted to minimize the number of units detected (forcing a merging of visually similar clusters), 2) we opted to reject units that showed a notable inter-spike-interval below 3 milliseconds, 3) we elected to *force* spikes from the noise cluster onto units (adding noise to the data but decreasing the likelihood of missing a responsive spiking activity) to ensure that as little data as possible are classified as noise; this action also captures all the noise in as single cluster that can then be rejected, 4) we *rejected* units with a low spike count (i.e., below 10 spikes), 5) we favored low number of units within a channel to numerous units with potential over-clustering. While this manually supervised clustering allows for subjective interpretation, it is common in human single neuron recording where the unit counts are low. We included a larger set of units that may have been noisy in the initial calculation since the data was unique and we wanted to ensure that we maximize the statistical yield.

Following the manual spike sorting we looked at all the units that survived the clustering and manually observed units that showed similar spike shapes within adjacent micro-wires (i.e., units that may have been exposed to the same noise source). We rejected units that seemed

identical in their noise artifact characteristics. We tested both the manual rejection and one done using an automatic tool (Dehnen et al., 2021). Units that were deemed identical were removed from further analyses. As with the sorting, we erred on the side of including units that may have emerged from duplicate sources, despite the risk of using noisy data. We estimated the maximal number of potential duplicate neurons as 28 out of the total unit count (20%). Without any rejection, the total number of units would have been 134. Notably, we focused our clustering efforts on the two units that were at the center of our analyses (the units responding to the variation of the task) and ensured that those units show the optimal clustering we could yield.

Units were manually classified as multi- or single-unit based on spike shape, spike variance, inter-spike interval distribution per cluster, and the presence of plausible refractory period.

Spike distance calculation

To estimate the lag between two neurons we calculated the absolute distance separating two spikes within those neurons. We computed the absolute distance by subtracting the onset times of pairs of spikes across two neurons, for each spike. That is, for each spike in neuron *i*, we calculated the lag to the closest spike in neuron *j*. Following, we averaged the time lags across all spikes to yield a single number reflecting the mean time difference between neurons_{ij}. We repeated this calculation for all pairs of neurons (gorilla-neuron versus all other responsive neurons) in the duration investigated (clip viewing). As an intuition, a low spike distance estimate suggests an alignment between two neurons (yet, not necessarily causal interaction).

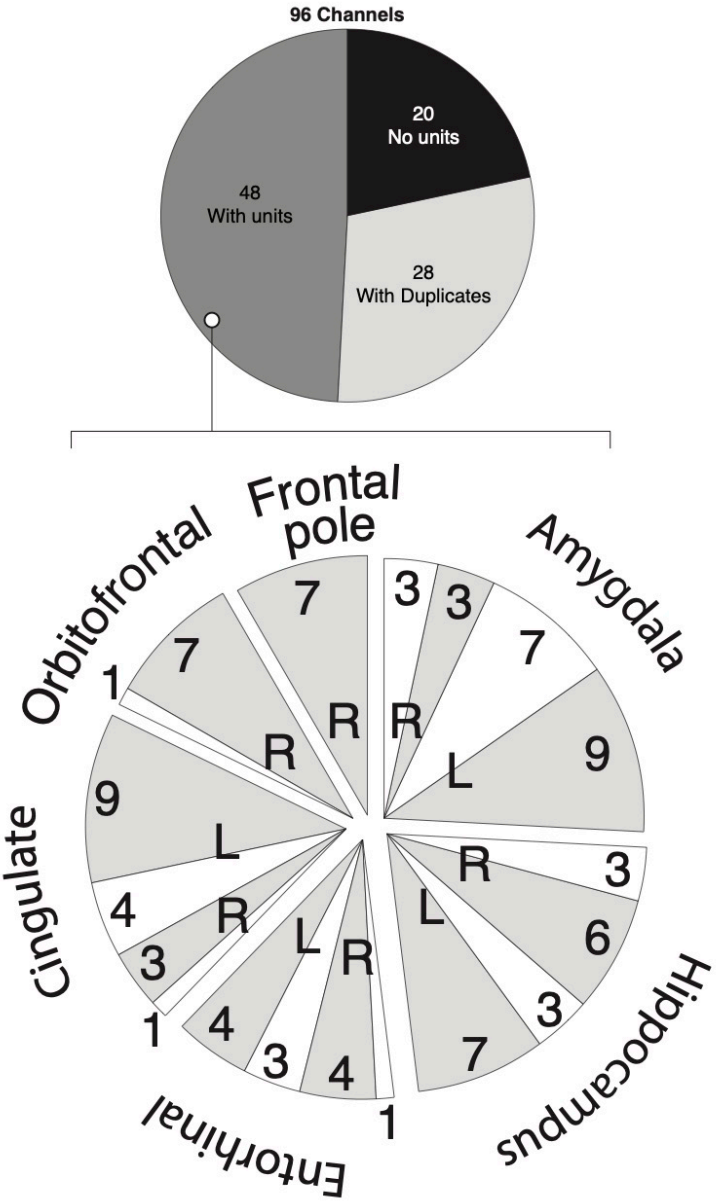
Spike field coherence

We identified the phase of spikes within the LFP using Python's *scipy signal.find_peaks* function with the wave's peak set to 0^{o}. Following, we allocated the spike to one of 20 bins within the wave (from 0 to $2\boldsymbol{\pi}$).

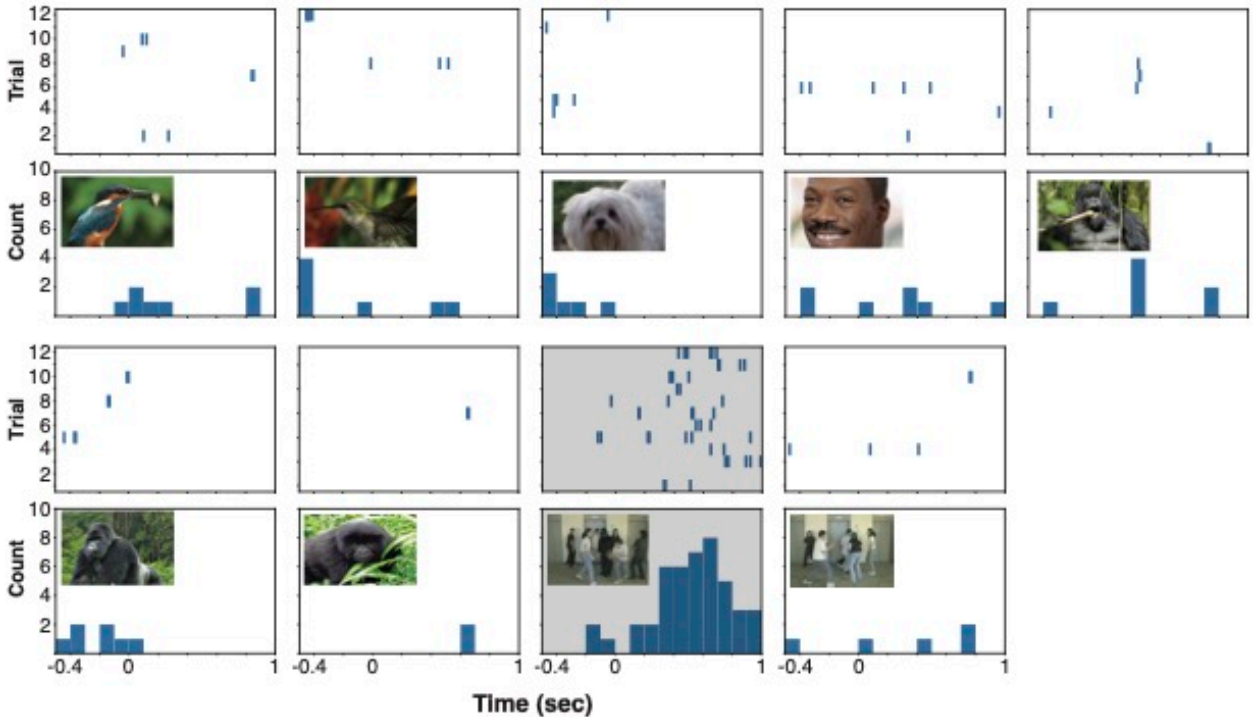
Supplementary table 1. List of units and spike count in each region.

Unit number	Unit label	SU/MU	Site	Spike count
1	Clip-gorilla	S		2,655
2	Rumination	S		464
3	Rumination	M		3,570
4	Rumination	S		147
5		S		223
6		M		6,844
7	Rumination	M		3,390
8	Rumination	M	**RA**	3,423
9	Rumination	M		4,389
10	Rumination	M		6,864
11	Rumination	S		6,797
12	Gorilla neuron	S		369
13	Rumination	M		3,823
14	Rumination	S		192
15	Rumination	M		6,395
16	Rumination	M		2,234
17	Rumination	M		353
18	Rumination	M		1,485
19	Rumination	M		1,413
20	Rumination	S		95
21		S	**RH**	43
22	Rumination	M		1,449
23	Rumination	M		1,735
24	Rumination	M		2,943
25		S		59
26	Rumination	M		5,054
27	Rumination	M		3,754
28	Rumination	S		343
29	Rumination	M		3,786
30	Rumination	M	**REC**	3,617
31	Rumination	S		20
32	Rumination	M		2,925
33	Rumination	S		97
34	Rumination	M		5,587
35	Rumination	M		496
36	Rumination	S		92
37	Rumination	M	**ROF**	5,025
38	Rumination	M		735
39	Rumination	M		7,181
40	Rumination	M		2,328
41	Rumination	M		767

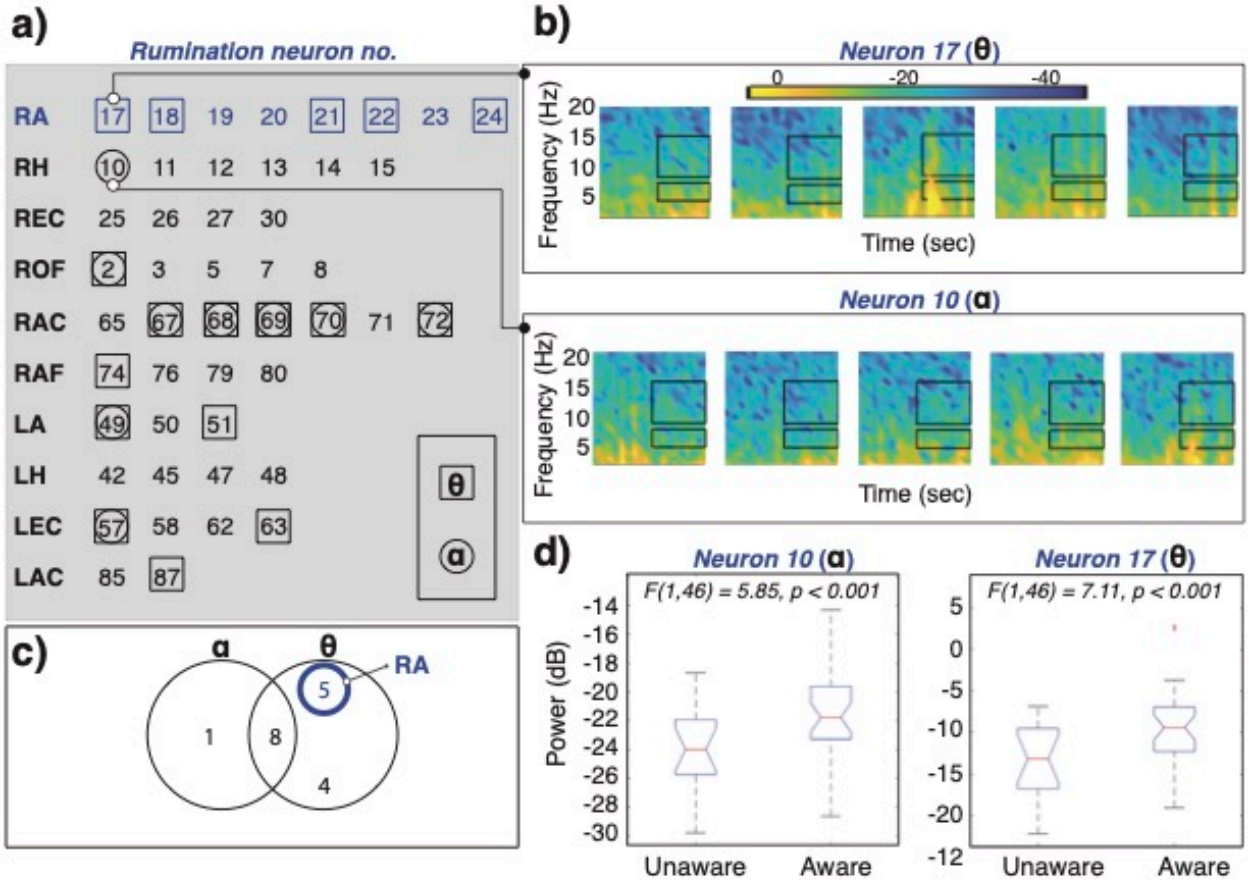
Unit number	Unit label	SU/MU	Site	Spike count
42	Rumination	M		2,430
43		S		30
44	Rumination	M		1,553
45		S		82
46	Rumination	M		112
47	Rumination	M		1,639
48	Rumination	S	**RAC**	67
49	Rumination	M		1,224
50		S		268
51	Rumination	M		2,494
52	Rumination	M		71
53	Rumination	M		4,000
54	Rumination	M		5,574
55	Rumination	M		2,391
56	Rumination	M		393
57	Rumination	M		1,565
58	Rumination	M	**RAF**	159
59	Rumination	M		5,868
60	Rumination	M		5,827
61	Rumination	M		293
62	Rumination	M		792
63		S		48
64	Rumination	M	**LA**	597
65		S		58
66	Rumination	M		386
67		S		61
68		M		27,320
69	Rumination	M		363
70	Rumination	M		2,066
71	Rumination	M		66
72	Rumination	M	**LH**	1,354
73		S		48
74		S		14
75	Rumination	M		521
76		S		38
77	Rumination	M		4,399
78	Rumination	M		2,220
79		S	**LEC**	35
80	Rumination	M		3,472
81	Rumination	M		1,689
82	Rumination	M		777
83	Rumination	S	**LAC**	75
84	Rumination	M		344
85	Rumination	M		722



Supplementary figure 1. **Unit breakdown by location and type.** White background in bottom chart marks single-units, and grey background marks multi-units.



Supplementary figure 2. **Peristimulus time histograms of the clip-gorilla unit for all stimuli in the during-experiment screening.** The stimulus is shown in the top-left corner of the histogram panels. The responsive stimulus (clip-gorilla) is highlighted in grey.



Supplementary figure 3. **LFP modulation between unaware/aware trials in the rumination epoch.** a) of the neurons that exhibited a change in the spiking activity during the rumination epoch, a subset also showed a significant change in alpha/theta band power across all pairs of unaware/aware trials (collapsing the 2 unaware trials and the 3 aware in the comparisons). Note (in blue) that five of the LFP modulations were shown in neurons from the same macro-electrode as the gorilla/clip-gorilla neurons. **b)** we highlight two examples of LFP activities that have shown the modulated activity. Black rectangle highlight the alpha and theta band activity in the rumination epoch investigated. **c)** breakdown of the number of micro-wires showing the modulation in [a]. **d)** The ANOVA for the trials in **[b].**

***Supplementary clip*. A video showing the center of mass during the entire clip in a single "aware" trial for one neuron.** The video shows the CoM in any given moment during the clip (**left**) and the corresponding phase of the neuron's spikes on the theta band LFP, binned as in **figure 8** (**right**). A red line marks the CoM in each bin. We denote in the text the epoch.

***Supplementary code*.** A *Python* code with the analyses script is available on: www.morancerf.com/publications.