

Decoding spontaneous changes in the content of consciousness from prefrontal ensembles

**Vishal Kapoor^{1,†*}, Abhilash Dwarakanath^{1,†}, Michel Besserve¹, Shervin Safavi^{1,2},
Joachim Werner¹, Theofanis I. Panagiotaropoulos^{1,3*}, Nikos K. Logothetis^{1,4}**

¹ Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Tübingen 72076, Germany

² International Max Planck Research School, Tübingen 72076, Germany

³ Cognitive Neuroimaging Unit, CEA, DSV/I2BM, INSERM, Université Paris-Sud, Université Paris-Saclay, Neurospin Center, 91191 Gif/Yvette, France

⁴ Division of Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PT, UK

†These authors contributed equally to this work

*Co-correspondence:

vishal.kapoor@tuebingen.mpg.de

theofanis.panagiotaropoulos@tuebingen.mpg.de

One sentence summary

Neural correlates of conscious perception can be detected and perceptual contents can be reliably decoded from the spiking activity of prefrontal populations.

INTRODUCTION

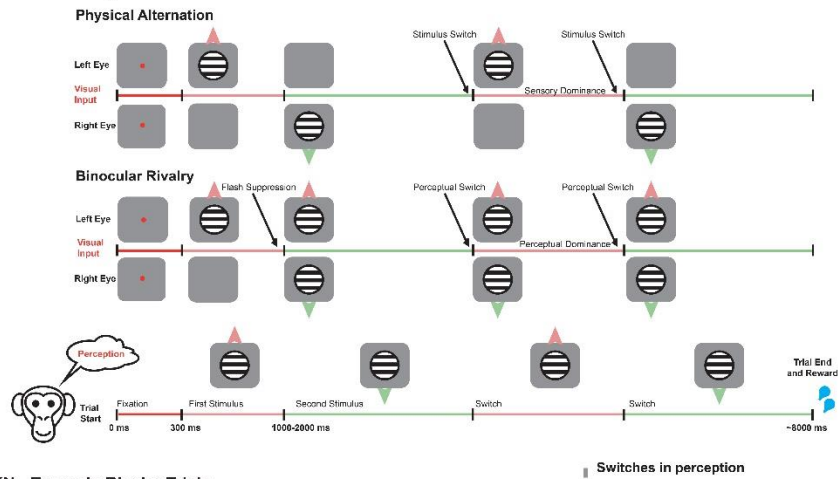
One of the hardest problems in neuroscience is to understand the biological basis of consciousness (Adolphs 2015; Crick and Koch 2006; Miller 2005). A seminal paper almost 30 years ago, incited researchers that “*the time is ripe for an attack on the neural basis of consciousness*” and proposed conscious visual perception as a form of consciousness that is within the reach of neuroscience (Crick and Koch 1990) . Since then, several theoretical treatises (Baars 2005; Crick and Koch 1998; Leopold and Logothetis 1999; Lau and Rosenthal 2011) including the frontal lobe hypothesis (Crick and Koch 1998), the higher order (Lau and Rosenthal 2011) and global workspace theories (Baars 2005; Dehaene and Changeux 2011) suggested a critical role for the brain’s prefrontal cortex (PFC) in mediating conscious perception. Evidence supporting its involvement comes from functional magnetic resonance imaging (fMRI) (Dehaene and Changeux 2011; Rees et al. 2002), experience of visual hallucinations upon electrical stimulation of the region (Vignal et al. 2000; Blanke et al. 2000), impaired perceptual experience following PFC lesions (Odegaard et al. 2017; Del Cul et al. 2009; Fleming et al. 2014; Szczepanski and Knight 2014; Nakamura and Mishkin 1986) as well as intracortical recordings of neural activity (van Vugt et al. 2018; Panagiotaropoulos et al. 2012; Libedinsky and Livingstone 2011; Gelbard-Sagiv et al. 2018). However, an opposing theory of consciousness, information integration theory (IIT) suggests that the contents of consciousness can be traced in a “posterior cortex hot zone” while prefrontal cortex is critical for processing the consequences of conscious perception. In particular, it was suggested that PFC activation during conscious perception paradigms reflects signals related to task demands and monitoring, introspection and motor reports rather than consciousness (Frässle et al. 2014; Knapen et al. 2011; Koch et al. 2016; Brascamp et al. 2015). For example, when BOLD fMRI signal modulation was compared between reported and unreported spontaneous changes in the content of consciousness, frontal cortex was found to be dramatically more active during motor reports (Frässle et al. 2014; Safavi et al. 2014; Zaretskaya and Narinyan 2014), thus casting doubt on the role of PFC in conscious perception (Koch et al. 2016; Brascamp et al. 2018; Odegaard et al. 2017; Boly et al. 2017). Nevertheless, the univariate analysis typically performed in the fMRI studies as well as the indirect nature and limited spatial resolution of BOLD fMRI signal compared to single neuron recordings leaves open the possibility that

prefrontal ensembles do reflect the content of consciousness on a microscopic scale, even without report requirements.

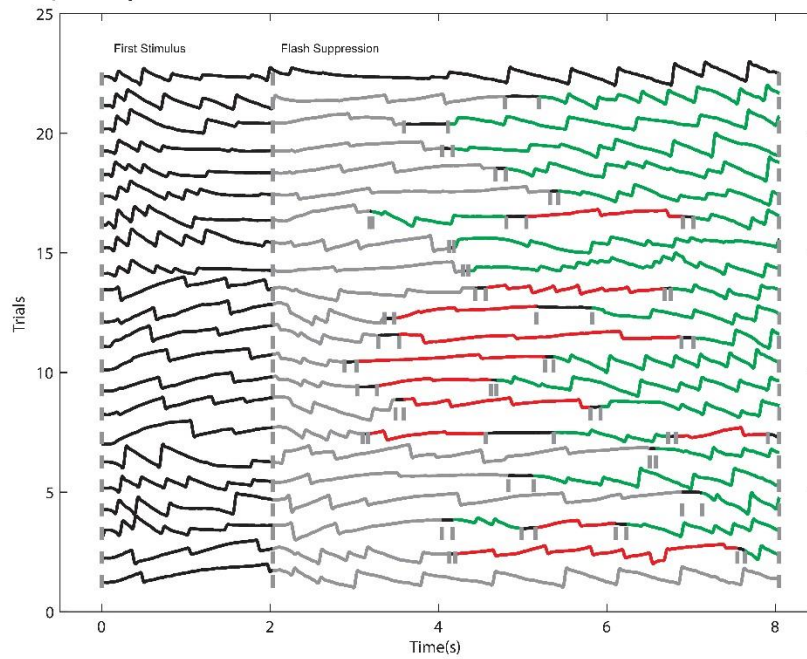
We examined this hypothesis by directly probing large neural populations in the inferior convexity of the macaque PFC with multielectrode arrays during a no-report binocular rivalry (BR) paradigm. BR belongs to the family of multistable perceptual phenomena (Leopold and Logothetis 1999; Wheatstone 1838; Blake and Logothetis 2002), which allow a dissociation of conscious perception from sensory input, by inducing in an observer fluctuations in the content of consciousness across time without a change in sensory stimulation. BR instigates these perceptual fluctuations through presentation of incongruent, dichoptic visual input to corresponding retinal locations of a subject, thereby eliciting stochastic, endogenously driven alternations in subjective perception. For a certain duration, only one of the two images is consciously experienced, while the other is perceptually suppressed (Blake and Logothetis 2002; Wheatstone 1838). The standard practice in BR experiments requires subjects to manually report their percepts, therefore contaminating processes related to conscious perception with signals related to their consequences like voluntary motor reports as well as introspection (Tsuchiya et al. 2015; Pitts et al. 2014; Aru et al. 2012). During rivalry between opposing directions of motion, the polarity of optokinetic nystagmus (OKN) reflex, a combination of smooth pursuit and fast saccadic eye movements, is known to be tightly coupled to the reported direction of motion (Logothetis and Schall 1990; Wei and Sun 1998; Naber et al. 2011). Thus, the reflexive nature of OKN can be exploited as an objective measure of changes in the content of consciousness removing any confounds in the neural activity originating from voluntary motor reports.

We combined electrophysiology with no-report motion BR and found that prefrontal neurons robustly represent the content of consciousness during spontaneous perceptual switches. Furthermore, monitoring simultaneously large neural populations allowed us to observe for each spontaneous perceptual transition the dynamics among neuronal ensembles representing two percepts that compete for access to consciousness.

A Binocular Rivalry Task



B OKN - Example Rivalry Trials



C Percept duration distributions

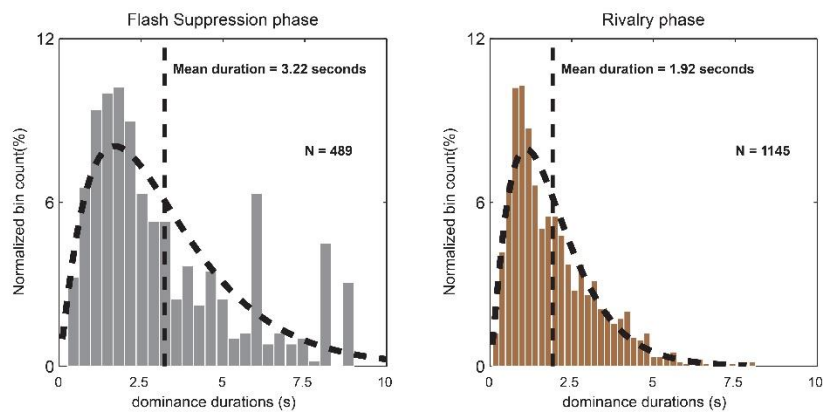


FIGURE 1

(A) The binocular rivalry paradigm. The task consisted of two trial types, namely, the physical

alternation (PA) trials and binocular rivalry (BR) trials. Both trial types started with the presentation of a fixation spot, cueing the animal to initiate fixation. Upon successful fixation for 300 milliseconds, a drifting sinusoidal grating was monocularly presented. After 1 to 2 seconds, the first stimulus was removed and a second grating drifting in the opposite direction was presented in the contralateral eye during PA trials. During BR trials, the second stimulus was added to the contralateral eye without the removal of the first stimulus. This results in perceptual suppression of the first stimulus and is denoted by flash suppression. After this period, the visual input alternated between upward and downward moving gratings in the physical alternation condition. In the binocular rivalry condition, the percept of the animal could randomly switch between the discordant visual stimuli. Note that the perception of the animal displayed in the bottom row is identical, even though the underlying visual input is monocular in PA trials, while its dichoptic during BR.

(B) OKN elicited during example BR trials from one recording session. The first dashed line denotes the beginning of the flash suppression phase. Subsequent dominance phases are color coded and their beginning and end are marked with a grey dashed line. Note that on certain trials the flash suppression resulted in a prolonged continuous suppression of the previously presented direction of motion (example trial denoted with a red arrow), while on other trials, the previously presented direction of motion remained dominant.

(C) Perceptual dominance distributions during flash suppression and rivalry phase displayed a gamma distribution, a psychophysical feature of binocular rivalry specifically (Levelt 1967) and multistable stimuli in general.

RESULTS

no-report Binocular rivalry paradigm and behavior

We exposed two rhesus macaques to a no report BR paradigm, which consisted of two trial types, physical alternation (PA) and BR trials (see Methods). Each trial started with a fixation spot cueing the animal to initiate fixation, which lasted ~300 milliseconds, followed by an upward or downward drifting grating that was presented monocularly for 1-2 seconds. After this initial phase, during BR trials, a second grating drifting in the opposite direction was added to the contralateral eye, typically inducing the perceptual suppression of the first stimulus, a phenomenon known as binocular flash suppression (BFS) (Wolfe 1984; Lansing 1964). Following this, visual competition ensued, resulting in spontaneous perceptual switches between the two opposing directions of motion. PA trials consisted of exogenous alternating monocular presentations of the same stimuli with a temporal distribution of presentation times similar to BR alternations.

The rivaling stimuli contained oppositely drifting gratings in order to elicit optokinetic nystagmus (OKN), (see Methods), which served as a reliable surrogate for phenomenal perception (Logothetis and Schall 1990; Naber et al. 2011; Wei and Sun 1998). Example OKN traces are presented in Figure 1B. In most trials, the OKN switched polarity, indicating perceptual dominance of the newly presented direction of motion (marked with red and green). Following this initial period of externally induced perceptual suppression OKN could be observed switching multiple times indicating spontaneous switches in the content of conscious perception. These well-defined epochs of stable perceptual dominance during both BFS and BR periods displayed a gamma distribution (Figure 1C), typical of multistable perception dynamics (Levelt 1967) with an average dominance duration of 3.22 and 1.92 seconds respectively.

Neuronal activity reflects spontaneous changes in the content of conscious perception

Content-specific neuronal activity during conscious perception is directly related to feature selectivity. If neurons during BR reliably increase their firing rate each time their preferred stimulus is perceived and suppress their firing when it is perceptually suppressed, then they are thought to explicitly represent conscious content. Figure 2B displays simultaneously recorded discharge activity from two prefrontal sites and OKN during two single PA and BR trials. In these two recorded locations activity was stronger for different directions of motion in the PA trial resulting in opposing modulation depending on the presented direction of motion. Neuronal activity recorded in the same sites was also reliably correlated with subjective changes in conscious perception during the BR trial, for both the externally induced perceptual suppression phase of BFS and the spontaneous perceptual switches during BR.

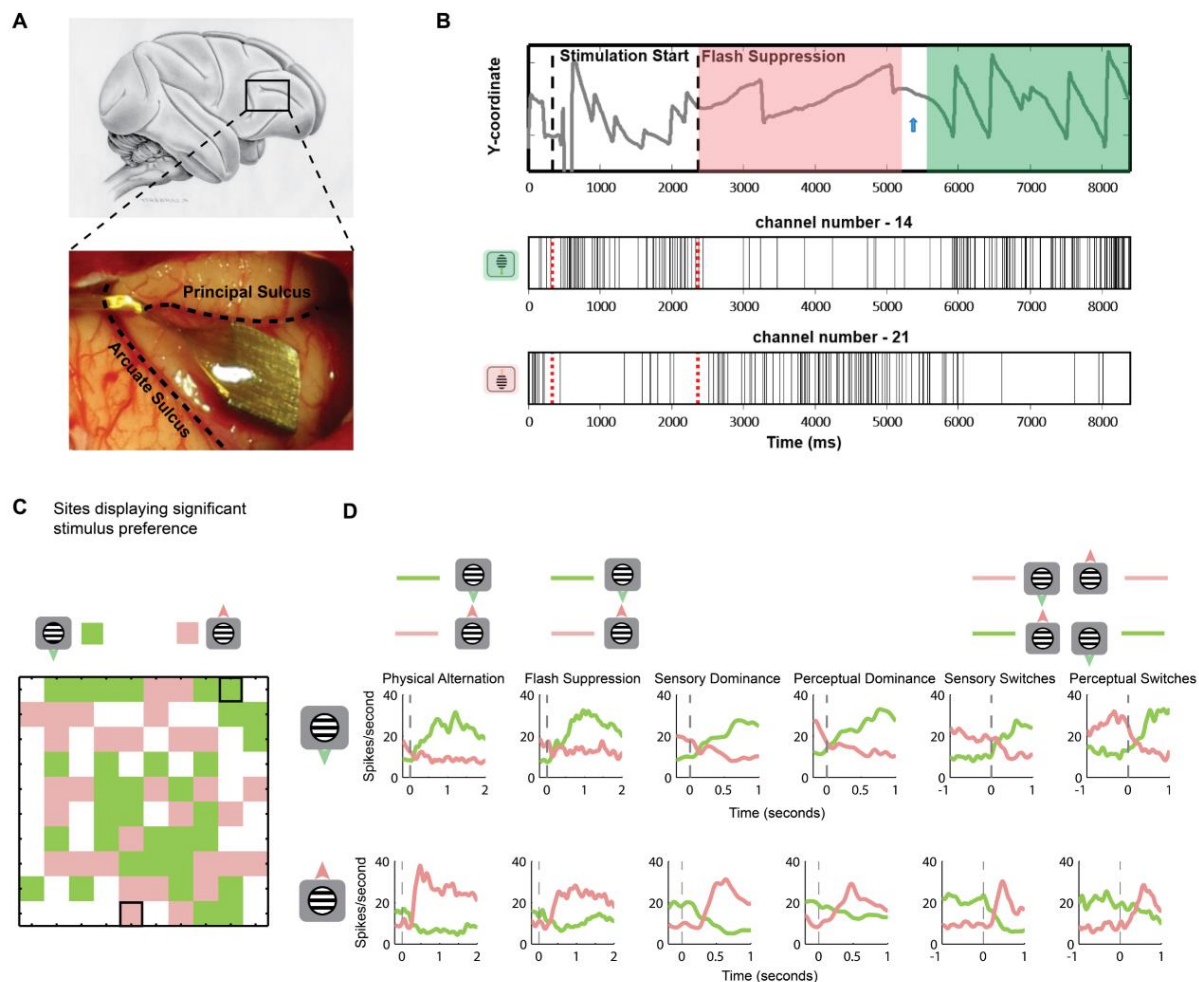


FIGURE 2

(A) The figure displays the location of the implanted Utah array in the prefrontal cortex on a schematic macaque brain and during actual surgery. We specifically targeted the ventrolateral prefrontal cortex, which has strong anatomical connectivity with the temporal lobe and has previously been demonstrated to display neural signals selective to object stimuli.

(B) OKN and the corresponding spiking activity during one example trial. The trial started with monocular presentation of downward drifting gratings. An upward drifting grating was added to the contralateral eye ~2000 ms later, which resulted in perceptual suppression. A spontaneous switch (blue arrow) in perception happens at ~5000 ms, and can be detected as a change in the OKN direction. The corresponding spiking activity as rasters from two simultaneously recorded channels is presented below. While channel number 14 displays strong spiking activity when downward drifting gratings is perceptually dominant, channel number 21 responds stronger when upward drifting gratings is perceived.

(C) Sites which display significant stimulus preference have been displayed on the array. Green and pink pixels reflect sites which preferred upward or downward drifting gratings respectively. The selective sites were distributed throughout the array.

(D) Displayed are spike density functions of two perceptually modulated sites recorded simultaneously in the vIPFC during PA and BR trials of the task. Pink and green colors in the first four columns correspond to the response elicited by presentation or perception of downward and upward drifting grating respectively during different task conditions. In the last two columns, pink curves depict the activity elicited during a stimulus or perceptual switch from downward drifting grating to an upward drifting grating while green denotes switches in the opposite direction. Displayed in the upper row is activity of an example site, which was significantly modulated by a downward drifting grating, while the lower row displays psth of a site strongly modulated by an upward drifting grating. The activity of these units is very similar during the two trial types, thus displaying robust perceptual modulation. The recording locations are outlined by black squares in (C).

We focused our analysis of the spiking activity underlying conscious perception during distinct phases of the task. These were perceptual dominance of a given stimulus (i) externally induced because of a stimulus change during flash suppression, or (ii) due to an endogenous spontaneous switch, typical of BR and (iii) around such a perceptual switch. Corresponding temporal phases from PA trials were also assessed for selectivity and compared with activity elicited during BR trials (see methods). The activity for two example units (their recording locations marked by black boxes in Figure 2C) are displayed in Figure 2D. The first neuron's activity is robustly modulated by the presentation of a downward drifting grating, while the other is modulated strongly by the grating drifting upwards. These two units were simultaneously recorded and located on different locations on the multielectrode array and displayed robust modulation both during PA as well as BR trials. Moreover, the spiking activity switched reliably during stimulus as well as perceptual switches.

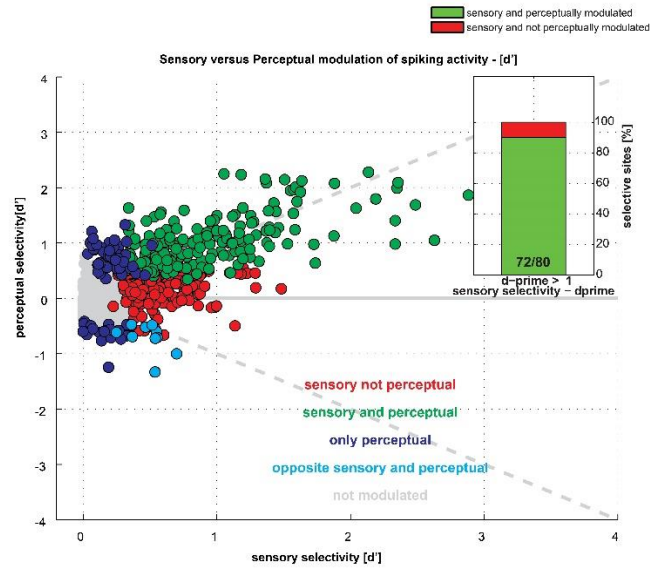
Furthermore, when stimulus preference was projected back to the multielectrode array we found that the recorded sites displaying similar stimulus preference were not randomly distributed but clustered throughout the 16mm² recorded area of the inferior prefrontal convexity (Figure 2C) suggesting that competition in PFC happens at a columnar level.

Modulation of prefrontal units during subjective perception

We quantified the modulation of all (n=987) recorded single and multi-units during subjective perception and compared it with the modulation of the same units during purely sensory stimulus presentations in PA trials, using a d' index as a measure of stimulus selectivity (Panagiotaropoulos et al. 2012; Keliris et al. 2010). On average, we observed that the proportion of feature selective units which were perceptually modulated in the perceptual

dominance phases of BFS and BR trials was 53.8 % and 40.38 %, respectively. In general, the recorded neural population displayed considerable heterogeneity in their selectivity, as evident from the scatter of d-primes (Figure 3 A), during the perceptual dominance phases in BFS and BR. For units with strong sensory selectivity in PA ($d\text{-prime} > 1$), perceptual modulation was more likely to remain significant during the perceptual dominance phases of BFS (90 %) and BR (84.78 %). This result indicates that the proportion of perceptually modulated neurons in the PFC is remarkably similar to the proportion previously observed in the temporal lobe (Sheinberg and Logothetis 1997), thus suggesting that there are at least two cortical regions where neuronal activity correlates strongly with subjective (Panagiotaropoulos et al. 2012) changes in conscious perception. Furthermore, it seems that the activity of prefrontal units correlates with subjective perception of more simple visual features like direction of motion, in addition to faces and more complex stimuli as previously described.

A Binocular Flash Suppression



B Perceptual Dominance

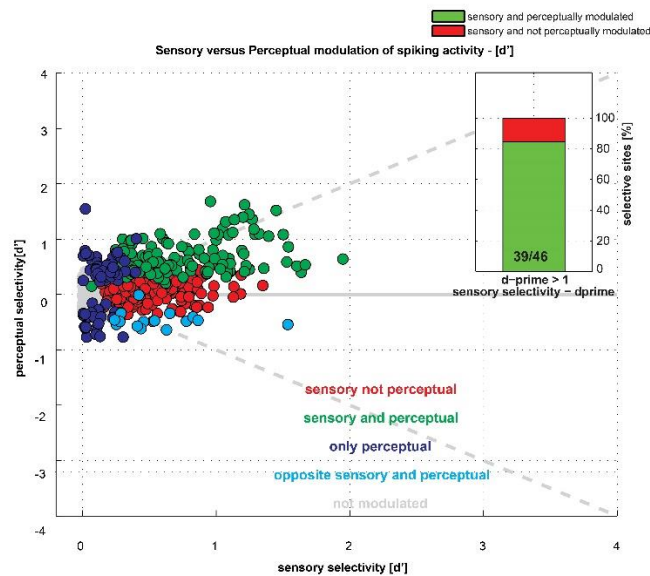


FIGURE 3

Sensory versus perceptual modulation of spiking activity during PA and BR trials. (A) A scatter plot of sensory versus perceptual preference (d') for all units across all datasets is displayed. Units showing no significant modulation in PA or BR trials are displayed as grey dots while those with significant modulation during both conditions are displayed as green dots. In red color, are dots which display significant preference only during PA trials. Units displaying significant modulation only during BR trials are displayed as blue dots, while in cyan are units which changed their preference across the two conditions. It is evident from the scatter plot, that the proportion of PA modulated units which are also significantly modulated during FS increase as a function of the strength of sensory selectivity (d').

The inset displays the proportion of PA modulated units with a d' greater than 1, which were also

significantly modulated during the flash suppression phase in BR trials (green). (B) Same as in (A) for perceptual dominance phase of the task. Almost 80 (BR) to 90 % (BFS) of units with a d' greater than one in PA were also perceptually modulated during BR trials.

Population spiking activity in the PFC reflects conscious perception

Plotting the spiking activity averaged across all units which displayed significant selectivity (see methods) during the first monocular switch phase of PA trials (corresponding temporally with the flash suppression phase of the BR trials), revealed robust modulation when their preferred stimulus was presented (Figure 4A). There was a strong peak response followed by a long sustained response. A dramatic reduction in activity was observed, during the presentation of the non-preferred stimulus. When the average activity of these units was plotted during the flash suppression phase of BR trials, robust perceptual modulation was observed. The population displayed not only a strong activation when their preferred stimulus was perceived, but also a robust reduction in response, when their preferred stimulus was suppressed by the presentation of a non-preferred stimulus in the contralateral eye. Moreover, remarkably similar perceptual modulation in population activity was observed when stimuli were perceived during spontaneous changes in subjective perception in BR trials as well as around perceptual switches (Figure 4B and see methods).

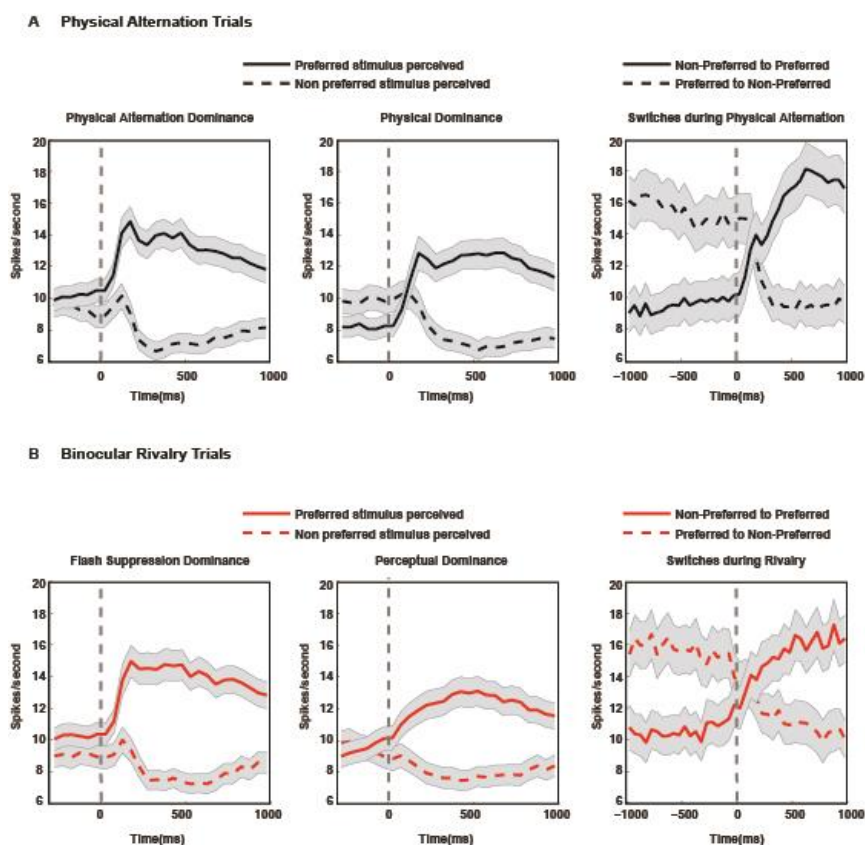


FIGURE 4

Displayed across three columns is the average population spiking activity during (A) PA and (B) BR trials. The population activity pooled and averaged across all units which were significantly modulated during both PA and BR trials for the same stimulus is plotted here during different temporal phases of the task, namely, the flash suppression phase, the perceptual dominance phase and switches. A remarkable similarity in population activity across the two trial types indicates strong and robust perceptual modulation in the units recorded in the vIPFC. We observe this both when a perceptual transition is brought about exogenously as is the case with binocular flash suppression as well as endogenous transitions which occur during binocular rivalry.

Decoding the content of conscious perception from prefrontal ensembles

Probing the PFC with multi-electrode arrays allowed us to monitor simultaneously content specific neuronal ensembles, displaying preferential responses to gratings drifting in opposite directions. We could therefore examine the population code for single instances of different types (upward to downward and downward to upward motion) of exogenous stimulus and endogenous, spontaneous perceptual transitions. Figure 5A plots the average normalized responses of selective neuronal ensembles (see Methods) around stimulus and perceptual switches for the two different switch types.

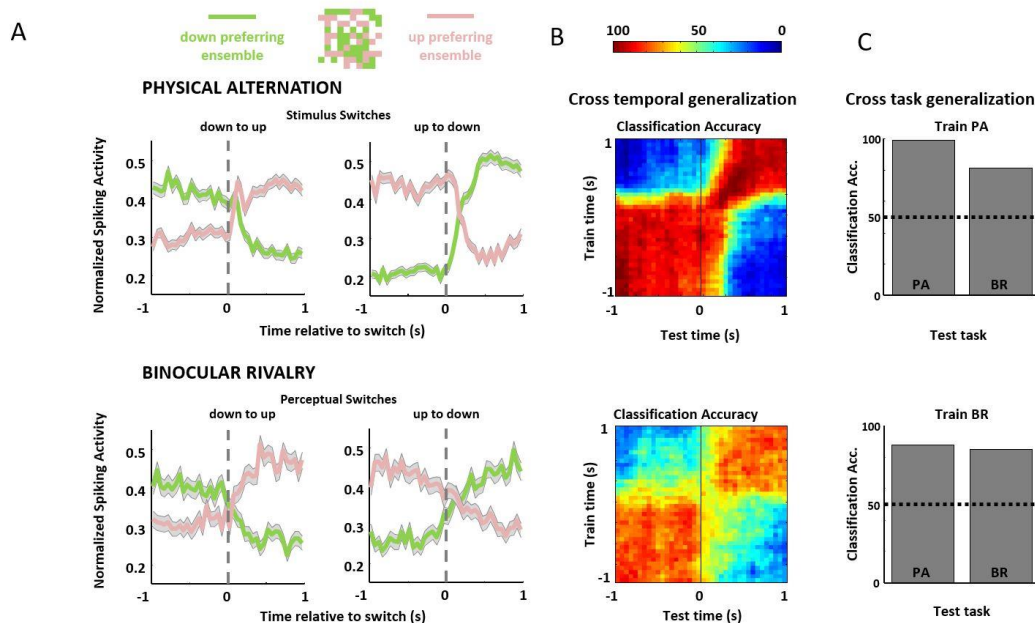


FIGURE 5

Decoding the contents of conscious perception from simultaneously recorded prefrontal ensembles. Recording with multi-electrode arrays allowed us to simultaneously monitor the activity of competing

neuronal ensembles which display significant modulation to the two stimuli. (A) Plotted is the normalized spiking activity of down preferring ensemble (green) and up preferring ensemble (pink) of units during up to down or down to up stimulus and perceptual switches. We observe a reliable modulation of in activity of neuronal ensembles during both PA as well as BR trials. (B) Cross-temporal decoding of stimulus contents around switches in perception during PA and BR trials. Decoding accuracy was tested for each pair of train/test time windows around a switch. (C) The cross task invariance of the population code was tested by training a classifier on activity in one task and testing on the other, for a single bin of 400 ms starting 200 ms after the switch. We observed significant cross-task generalization accuracy, thus suggesting that the underlying code is invariant to the task, and encoding perceptual contents.

Similar to the average unit PSTH's which were constructed after averaging the mean unit responses, the average ensemble activity, constructed from averaging the activity of feature selective populations across transitions, displayed changes concomitant with both exogenous stimulus changes in PA and subjective changes in perceptual states during BR trials (Figure 5A). We utilized a multivariate decoding approach to assess the reliability with which we could predict conscious perceptual contents from prefrontal spiking activity on single cases of perceptual transitions. We first decoded monocularly presented stimuli during PA trials (Figure 5B). The classifier discriminated between the two stimuli strongly above chance (50%) and generalized across temporal windows for the total duration of stimulus presentation (Figure 5C). Interestingly, a classifier trained on BR activity displayed remarkably similar temporal generalization around perceptual switches as was observed for stimulus switches. Lastly, we assessed the generalization of the classifier across the two trial types, in order to test the similarity of the population code representing the sensory input and its phenomenal experience. Strongly significant generalization of the classifier across PA and BR suggests that the prefrontal population code for sensory input and subjective perception, is not only similar but also reliable and robust.

Controlling for eye movement patterns

Given that the OKN is tightly linked to perceptual content, it was critical to dissociate neural activity related to oculomotor signals from activity related to subjective visual perception. Therefore, in four of six recording sessions, we probed their responses during monocular presentation of gratings drifting in eight different directions, in two control tasks,

with (fixation off) and without (fixation on) eye movements. During the fixation off task, the fixation window was the entire stimulus thus eliciting OKN, while in the fixation on task, an overlaid fixation spot (and a smaller fixation window) indicated that the monkeys must fixate to get reward, thus suppressing the OKN. To examine the influence of oculomotor signals on decoding motion direction from prefrontal populations we utilized a multivariate decoding approach and investigated if a classifier trained on activity elicited in response to stimuli which elicited OKN could reliably predict the same stimuli when they were viewed with the eye movements suppressed. We observed strongly significant and above chance decoding accuracy of the classifier both with and without eye movements (Figure 6 A and B). Furthermore, decoding could be generalized above chance across conditions (Figure 6C) thus suggesting that the spiking activity contained information related to stimulus content, and was not just driven by the eye movements. This result is in line with previous findings suggesting that the frontal cortex is activated strongly in response to visual motion both in the presence and absence of OKN (Dieterich et al. 1998; Zaksas and Pasternak 2006) and neurons in the region reflect a mixture of perceptual and oculomotor signals (Kim and Shadlen 1999).

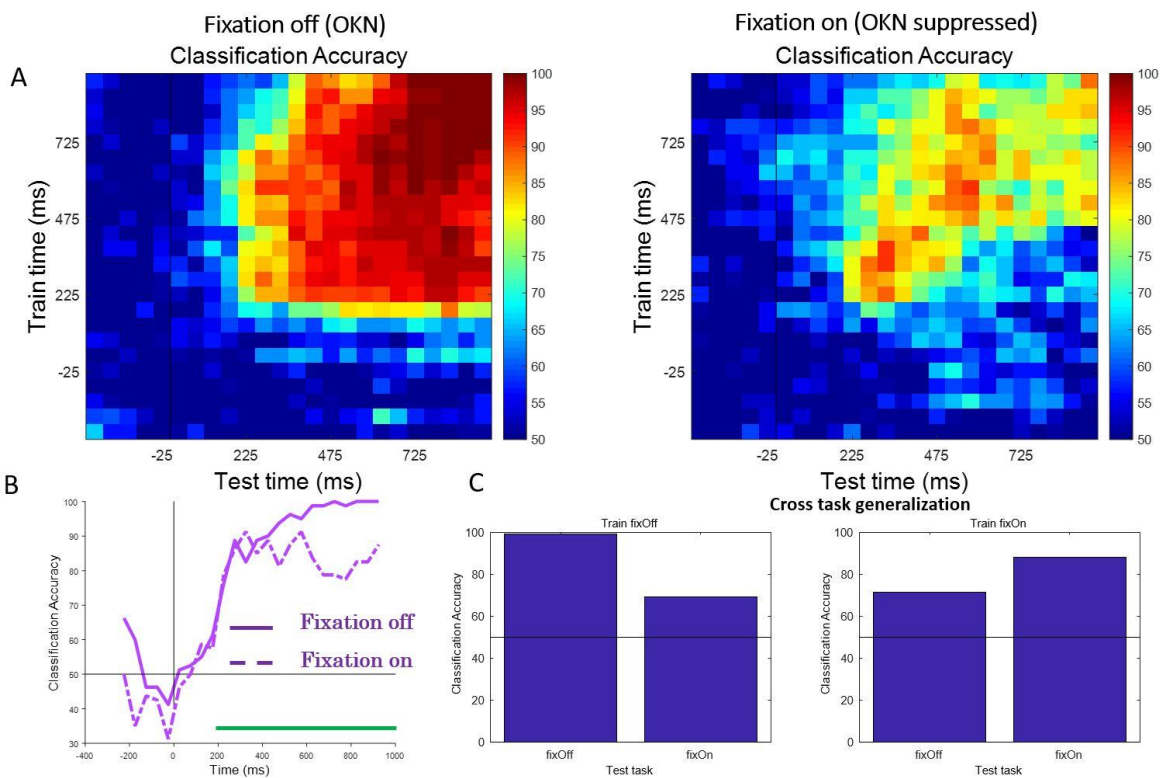


FIGURE 6

Decoding as an approach to understand the invariance of neural activity to the presence and absence of eye movements. Cross-temporal decoding of stimulus contents during fix off and fix on tasks

(Supplementary figure 1). Decoding accuracy was tested for each pair of train/test time windows. (B) Plotted is the classification accuracy as a function of time during the two tasks. Strong classification is observed during both paradigms (C) The cross task invariance of the population code was tested by training a classifier on activity in one task and testing on the other, for a single bin of 400 ms during the presentation of the visual stimulus. We observed significant cross-task generalization accuracy, thus suggesting that the underlying code is invariant to the task, and encodes stimulus contents in the presence and absence of eye movements.

DISCUSSION

Present results suggest that feature selective neural responses in the primate frontal cortex reflect subjective perception during a no report paradigm. While addressing the ongoing controversy about the neural correlates of conscious perception in the PFC (Koch et al. 2016; Brascamp et al. 2018; Odegaard et al. 2017; Boly et al. 2017; Mashour 2018), we additionally demonstrate that the contents of subjective experience can be reliably decoded from the activity of prefrontal ensembles during BR. Thus far, the paradigm of BR has been instrumental to investigating the neural correlates of conscious perception (Logothetis 1998). It offers a distinct advantage over BFS, since the perceptual switches are stochastic, internally driven and without any concomitant changes in the sensory input. Hence, it confers a unique opportunity to observe intrinsic neural dynamics contemporaneous with changes in the contents of subjective experience without any conflation from changes in the retinal feedforward signal. When paired with electrophysiological investigation of different regions of the primate visual system (Schmid and Maier 2015; Panagiotaropoulos et al. 2014), it has helped reveal that the proportion of feature selective neurons which correlate with conscious perception increase as one progresses in the visual cortical hierarchy from early visual areas (Leopold and Logothetis 1996; Logothetis and Schall 1989; Keliris et al. 2010) to later temporal regions (Panagiotaropoulos et al. 2012; Sheinberg and Logothetis 1997; Libedinsky and Livingstone 2011). The present study investigated a sub-region of PFC, where cells display selective responses to object stimuli (Pigarev et al. 1979; O Scalaidhe et al. 1997; Scalaidhe et al. 1999; Wilson et al. 1993) and it has reciprocal anatomical connections with the inferotemporal cortex (Webster et al. 1994). Previous studies which probed PFC neural activity during paradigms investigating conscious perception either utilized a motor report (van Vugt et al. 2018; Gelbard-Sagiv et al. 2018; Libedinsky and Livingstone 2011), observed stimulus unspecific activation

before a perceptual change (Gelbard-Sagiv et al. 2018), and therefore likely conflated by prerequisites of conscious perception, or investigated perceptual modulation among PFC neurons with a BFS paradigm, which lacks endogenous changes in perception (Panagiotaropoulos et al. 2012). Hence, a plausible resolution of the debate on PFC's contribution to conscious perception necessitated an electrophysiological investigation of neurons encoding stimulus content in the region during a no report perceptual paradigm.

Our findings are in contrast to the results obtained with recent imaging studies which have indicated a reduced involvement of the PFC in conscious perception (Frässle et al. 2014). However, a lower spatial resolution of the BOLD signal as well as its complicated relationship with neural activity limits the interpretations from imaging data, especially so, when null findings are reported (Michel and Morales n.d.; Morales and Lau n.d.). Moreover, the contrastive univariate analytical approach, typically utilized for investigating BOLD signal modulations in PFC, remains limited in sensitivity in comparison to the resolution which intracortical recordings confer. Importantly, the neurons in the frontal cortex often display a high degree of mixed selectivity (Rigotti et al. 2013; Mante et al. 2013) and distinct patterns of activity during perceptual paradigms (Kapoor et al. 2018). We find that the units displaying preferential responses to opposite directions of motion were distributed throughout the array. Finally, imaging studies typically utilize a subtractive approach, across conditions. Given present findings of a similarity in responses among feature selective neuronal ensembles during the PA and BR conditions, such a subtractive approach would result in a null finding.

In summary, the results detailed above show that the activity of a large proportion of visually selective neurons in the PFC is modulated by conscious content. They, therefore lend support to theoretical approaches such as the global workspace hypothesis and the higher order theory of consciousness, which hypothesize a critical role to the PFC in mediating consciousness in general and conscious perception in particular (Lau and Rosenthal 2011; Brown et al. 2019; Anon n.d.). While these theories have recently received criticism because of this region's functional relevance to other cognitive processes, we address with this study, one such confound, namely the motor report (Tsuchiya et al. 2015). Future work aimed at elucidating the neural correlates of conscious perception not just in the PFC, but the primate brain in general could greatly benefit from disentangling consciousness from other cognitive processes such as attention, decision making or cognitive control (Aru et al. 2012; Block 2015).

METHODS

Binocular rivalry task and stimulus presentation

The task consisted of two trial types, namely, the physical alternation (pa) trials and binocular rivalry (br) trials. Both trial types started with the presentation of a red fixation spot (subtending 0.2 degree of visual angle), cueing the animal to initiate fixation. Upon successful fixation for 300 milliseconds within a fixation window, a drifting sinusoidal grating (Grating Parameters - Size: 8 degrees, Speed: 12-13 cycles per degree, Spatial Frequency: 0.5 cycles per degree, gratings were typically drifting vertically up or down) was monocularly presented. After about 2 seconds, the first stimulus was removed and a second grating drifting in the opposite direction was presented in the contralateral eye during pa trials. During br trials, the second stimulus was added to the contralateral eye without the removal of the first stimulus. This results in perceptual suppression of the first stimulus and is denoted by flash suppression (Lansing 1964; Wolfe 1984; Keliris et al. 2010; Panagiotaropoulos et al. 2012) in Figure 1. After this period, the visual input alternated between oppositely drifting gratings in the pa condition according to a distribution acquired from br trials presented during the training phase of the task. In the br condition, the percept of the animal could randomly switch between the discordant visual stimuli, whose temporal histogram could be approximated with a gamma distribution. A single trial duration was between 10-12 seconds. Note that the perception of the animal displayed in C is identical in the two conditions, even though the underlying visual input is monocular in pa trials, while its dichoptic in case of br. The eye (where the first stimulus was presented), the grating (which was presented first) and the different trial types (pa or br) were pseudorandomized and balanced in a single dataset. During the entire period of a trial, animals fixated within a fixation window, which typically was the same size ($\pm 8^\circ$) as the size of the stimulus. The animals were given a liquid reward upon successful maintenance of fixation within the window for the entire trial duration (typically, 10-12 seconds).

Dichoptic visual stimulation was carried out with the aid of a stereoscope and displayed at a resolution of 1280X1024 on the monitors (running at a 60 Hz refresh rate) using a dedicated graphics workstation. The visual stimuli and the task were designed with an in-house software written in C/Tcl. A QNX real-time operating system (QNX Software Systems) managed the precise temporal presentation of visual stimuli, and sent digital pulses to the Blackrock system. An infrared camera captured eye movements (1kHz sampling rate) with the software iView (SensoriMotoric Instruments GmbH, Germany). Besides monitoring eye movements online, they were also stored for offline analysis in both QNX-based acquisition system as well as

Blackrock data acquisition system. All behavioral training and electrophysiological recordings were carried out while the animals sat in a custom designed chair.

Surgical procedures

Two healthy rhesus monkeys (*Macaca mulatta*), H'07 and A'11 participated in behavioral and electrophysiological recordings. All experiments were approved by the local authorities (Regierungspräsidium) and were in full compliance with the application guidelines of the European community (Application number - KY-6/12). Each animal was implanted with a cranial headpost (material: titanium) custom designed to fit the skull based upon a high resolution MR scan collected using a 4.7 tesla scanner (Biospec 47/70c; Bruker Medical, Ettlingen, Germany). The headpost implantation was carried out while the animal was under general anesthesia and prior to the beginning of behavioral training on the binocular rivalry paradigm. The MR scan further aided in localizing the inferior convexity of the LPFC. Post behavioral training in the task, the animals underwent another surgery, where a Utah microelectrode array (Blackrock Microsystems, Salt Lake City, Utah USA; (Maynard et al. 1997)) was implanted in the inferior convexity of the prefrontal cortex. The array (4x4mm with a 10 by 10 electrode configuration and inter-electrode distance of 400 μ m) was placed 1 - 2 millimeters anterior to the bank of the arcuate sulcus and below the ventral bank of the principal sulcus, thus covering a large part of the inferior convexity in the ventrolateral PFC (Figure 1).

Electrophysiology data collection

Broadband neural signals (0.1 - 30 kHz) were recorded with the Neural Signal Processors (Blackrock Microsystems). Data were filtered between 0.3 - 3 kHz using a 2nd order Butterworth filter. Spikes were detected with an amplitude threshold set at five times the median absolute deviation (Quiroga et al. 2004). Any spike events with an inter-spike interval of less than a refractory period of 0.5 ms were discarded. Events satisfying the aforementioned criterion of threshold and the refractory period were kept for further analysis. Collected spike events were aligned to their minima. For spike sorting, 45 samples (1.5 milliseconds) around the peak were extracted. An automatic clustering procedure identified putative single neurons via a Split and Merge Expectation-Maximisation algorithm which fits a mixture of Gaussians on the spike feature data consisting of the first three principal components of the spike waveforms. Inspection and manual cluster cutting was carried out in Klusters (Lynn Hazan,

Buzsáki lab, Rutgers, Newark NJ). The details of the spike sorting algorithms have been described elsewhere (Tolias et al. 2007).

Selectivity of single unit and multi-unit activity

For judging selectivity, a given unit during BR trials, its spiking response was aligned to the onset of two events, invoking a perceptual change (i) flash suppression phase and (ii) perceptual dominance. Corresponding temporal phases during PA trials, that is the presentation of a second stimulus (corresponding to flash suppression phase) and every subsequent stimulus presentation (corresponding to perceptual dominance phase) was analyzed for assessing selectivity using a Wilcoxon rank-sum test. Thus any comparisons made across a given unit's activity during the two conditions were from temporally matched phases in PA and BR trials. Further, we analyzed all observation periods consisting of an uninterrupted perception (BR trials) or monocular presentation (PA trials) of one given motion direction for at least one second. The perception during BR trials was assessed via manual inspection and the phases depicting OKN direction were marked. With respect to perceptual switches, we analyzed transitions, which consisted of at least one second of clear dominance (as judged by the eye movement pattern), before and after a transition. Further, in order to make the comparisons to PA condition as close as possible, we analyzed transitions, which had a maximum of 250 milliseconds between the end of the preceding dominance, and the beginning of a new percept.

D-prime calculation

For every unit, we calculated a preference index - d' , by quantifying the strength of its selectivity during PA and BR trials during both the PA and BR trials. It was calculated as follows:

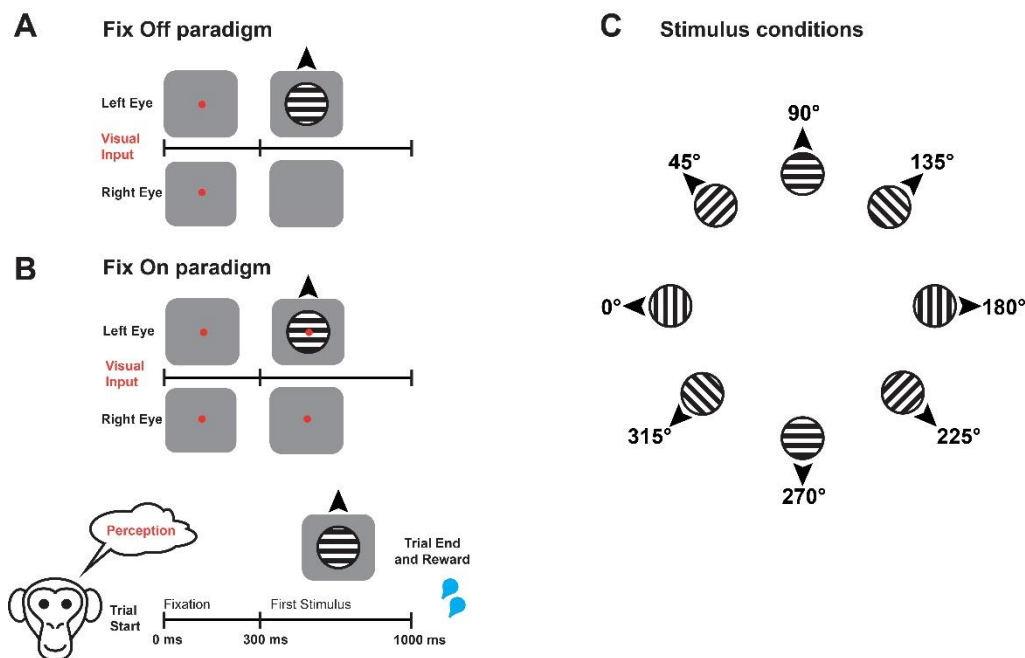
$$d' = \frac{\mu_{preferred} - \mu_{non\ preferred}}{\sqrt{(\text{var}_{preferred} + \text{var}_{non\ preferred})/2}}$$

where, $\mu_{preferred}$ and $\mu_{non\ preferred}$ refer to the average spiking response of a given unit during the presentation of its preferred and non-preferred stimulus, calculated over a duration of 1000 milliseconds after a stimulus or a perceptual change. The difference between these two quantities is normalized by the variance in the response distributions.

Decoding Analysis

A decoding approach was utilized in order to estimate if the spiking activity of neuronal ensembles in the prefrontal cortex contained information about the stimulus on a trial by trial basis. We utilized the maximum correlation coefficient classifier implemented as a part of the neural decoding toolbox for performing the analysis (Meyers 2013). We used 20 cross-validation splits, implying that 19 presentations of each stimulus was used for training and one of each for testing. 10 resample runs were carried out for estimating the decoding accuracy within a given task condition and for estimating the cross task generalization of the classifier trained on the two different task conditions (Figure 5 and 6). All decoding accuracy estimates are zero-one-loss results. For calculating the classification accuracy as a function of time, we utilized firing rates, calculated in 150 ms bins, sampled every 50 ms and used this for generating the plots in figure 6B, which plots the classification accuracy as a function of time. Each time bin is independently classified.

Supplementary Figure 1



Control tasks, fix off and fixon. Animals were exposed to two control tasks. Both paradigms started with cueing the animal to fixate for 300 ms, after which a drifting grating was presented monocularly. (A) During fix off paradigm, the fixation spot was removed at the onset of the stimulus, thus inducing optokinetic nystagmus eye movements. (B) During fix on paradigm, the stimulus was presented without removal of the fixation spot, and the animal was required to fixate until the end of the trial, in order to receive a juice reward. (C) During

both tasks, on each trial, a gratings drifting in one of eight different directions was presented.

Bibliography

Adolphs, R. 2015. The unsolved problems of neuroscience. *Trends in Cognitive Sciences* 19(4), pp. 173–175.

Anon Comparing the major theories of consciousness. - PsycNET [Online]. Available at: <https://psycnet.apa.org/record/2009-19897-077> [Accessed: 30 July 2019].

Aru, J., Bachmann, T., Singer, W. and Melloni, L. 2012. Distilling the neural correlates of consciousness. *Neuroscience and Biobehavioral Reviews* 36(2), pp. 737–746.

Baars, B.J. 2005. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in Brain Research* 150, pp. 45–53.

Blake, R. and Logothetis, N.K. 2002. Visual competition. *Nature Reviews. Neuroscience* 3(1), pp. 13–21.

Blanke, O., Landis, T. and Seeck, M. 2000. Electrical cortical stimulation of the human prefrontal cortex evokes complex visual hallucinations. *Epilepsy & Behavior* 1(5), pp. 356–361.

Block, N. 2015. CONSCIOUSNESS, BIG SCIENCE, AND CONCEPTUAL CLARITY. In: MARCUS, G. and FREEMAN, J. eds. *The Future of the Brain*. Essays by the World’s Leading Neuroscientists. Princeton University Press, p. 161.

Boly, M., Massimini, M., Tsuchiya, N., Postle, B.R., Koch, C. and Tononi, G. 2017. Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? clinical and neuroimaging evidence. *The Journal of Neuroscience* 37(40), pp. 9603–9613.

- Brascamp, J., Blake, R. and Knapen, T. 2015. Negligible fronto-parietal BOLD activity accompanying unreportable switches in bistable perception. *Nature Neuroscience* 18(11), pp. 1672–1678.
- Brascamp, J., Sterzer, P., Blake, R. and Knapen, T. 2018. Multistable perception and the role of the frontoparietal cortex in perceptual inference. *Annual review of psychology* 69, pp. 77–103.
- Brown, R., Lau, H. and LeDoux, J.E. 2019. Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Sciences*.
- Crick, F.C. and Koch, C. 2006. *23 problems in systems neuroscience*. van Hemmen, J. L. and Sejnowski, T. J. eds. Oxford University Press.
- Crick, F. and Koch, C. 1998. Consciousness and neuroscience. *Cerebral Cortex* 8(2), pp. 97–107.
- Crick, F. and Koch, C. 1990. Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*.
- Dehaene, S. and Changeux, J.-P. 2011. Experimental and theoretical approaches to conscious processing. *Neuron* 70(2), pp. 200–227.
- Del Cul, A., Dehaene, S., Reyes, P., Bravo, E. and Slachevsky, A. 2009. Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain: A Journal of Neurology* 132(Pt 9), pp. 2531–2540.
- Dieterich, M., Bucher, S.F., Seelos, K.C. and Brandt, T. 1998. Horizontal or vertical optokinetic stimulation activates visual motion-sensitive, ocular motor and vestibular cortex areas with right hemispheric dominance. An fMRI study. *Brain: A Journal of Neurology* 121 (Pt 8), pp. 1479–1495.
- Fleming, S.M., Ryu, J., Golfinos, J.G. and Blackmon, K.E. 2014. Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain: A Journal of Neurology* 137(Pt 10), pp. 2811–2822.
- Frässle, S., Sommer, J., Jansen, A., Naber, M. and Einhäuser, W. 2014. Binocular rivalry: frontal activity relates to introspection and action but not to perception. *The Journal of Neuroscience* 34(5), pp. 1738–1747.
- Gelbard-Sagiv, H., Mudrik, L., Hill, M.R., Koch, C. and Fried, I. 2018. Human single neuron activity precedes emergence of conscious perception. *Nature Communications* 9(1), p. 2057.
- Kapoor, V., Besserve, M., Logothetis, N.K. and Panagiotaropoulos, T.I. 2018. Parallel and functionally segregated processing of task phase and conscious content in the prefrontal cortex. *Communications Biology* 1, p. 215.
- Keliris, G.A., Logothetis, N.K. and Tolias, A.S. 2010. The role of the primary visual cortex in perceptual suppression of salient visual stimuli. *The Journal of Neuroscience* 30(37), pp. 12353–12365.
- Kim, J.N. and Shadlen, M.N. 1999. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience* 2(2), pp. 176–185.

- Knapen, T., Brascamp, J., Pearson, J., van Ee, R. and Blake, R. 2011. The role of frontal and parietal brain areas in bistable perception. *The Journal of Neuroscience* 31(28), pp. 10293–10301.
- Koch, C., Massimini, M., Boly, M. and Tononi, G. 2016. Neural correlates of consciousness: progress and problems. *Nature Reviews. Neuroscience* 17(5), pp. 307–321.
- Lansing, R.W. 1964. Electroencephalographic correlates of binocular rivalry in man. *Science* 146(3649), pp. 1325–1327.
- Lau, H. and Rosenthal, D. 2011. Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences* 15(8), pp. 365–373.
- Leopold, D.A. and Logothetis, N.K. 1996. Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379(6565), pp. 549–553.
- Leopold, D.A. and Logothetis, N.K. 1999. Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences* 3(7), pp. 254–264.
- Levelt, W.J. 1967. Note on the distribution of dominance times in binocular rivalry. *British Journal of Psychology* 58(1), pp. 143–145.
- Libedinsky, C. and Livingstone, M. 2011. Role of prefrontal cortex in conscious visual perception. *The Journal of Neuroscience* 31(1), pp. 64–69.
- Logothetis, N.K. 1998. Single units and conscious vision. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 353(1377), pp. 1801–1818.
- Logothetis, N.K. and Schall, J.D. 1990. Binocular motion rivalry in macaque monkeys: eye dominance and tracking eye movements. *Vision Research* 30(10), pp. 1409–1419.
- Logothetis, N.K. and Schall, J.D. 1989. Neuronal correlates of subjective visual perception. *Science* 245(4919), pp. 761–763.
- Mante, V., Sussillo, D., Shenoy, K.V. and Newsome, W.T. 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503(7474), pp. 78–84.
- Mashour, G.A. 2018. The controversial correlates of consciousness. *Science* 360(6388), pp. 493–494.
- Maynard, E.M., Nordhausen, C.T. and Normann, R.A. 1997. The Utah intracortical Electrode Array: a recording structure for potential brain-computer interfaces. *Electroencephalography and Clinical Neurophysiology* 102(3), pp. 228–239.
- Meyers, E.M. 2013. The neural decoding toolbox. *Frontiers in Neuroinformatics* 7, p. 8.
- Michel, M. and Morales, J. MINORITY REPORTS: CONSCIOUSNESS AND THE PREFRONTAL CORTEX.
- Miller, G. 2005. What is the biological basis of consciousness? *Science* 309(5731), p. 79.

- Morales, J. and Lau, H. The neural correlates of consciousness.
- Naber, M., Frässle, S. and Einhäuser, W. 2011. Perceptual rivalry: reflexes reveal the gradual nature of visual awareness. *Plos One* 6(6), p. e20910.
- Nakamura, R.K. and Mishkin, M. 1986. Chronic “blindness” following lesions of nonvisual cortex in the monkey. *Experimental Brain Research* 63(1), pp. 173–184.
- Odegaard, B., Knight, R.T. and Lau, H. 2017. Should a few null findings falsify prefrontal theories of conscious perception? *The Journal of Neuroscience* 37(40), pp. 9593–9602.
- O Scalaidhe, S.P., Wilson, F.A. and Goldman-Rakic, P.S. 1997. Areal segregation of face-processing neurons in prefrontal cortex. *Science* 278(5340), pp. 1135–1138.
- Panagiotaropoulos, T.I., Deco, G., Kapoor, V. and Logothetis, N.K. 2012. Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. *Neuron* 74(5), pp. 924–935.
- Panagiotaropoulos, T.I., Kapoor, V. and Logothetis, N.K. 2014. Subjective visual perception: from local processing to emergent phenomena of brain activity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 369(1641), p. 20130534.
- Pigarev, I.N., Rizzolatti, G. and Scandolara, C. 1979. Neurons responding to visual stimuli in the frontal lobe of macaque monkeys. *Neuroscience Letters* 12(2–3), pp. 207–212.
- Pitts, M.A., Metzler, S. and Hillyard, S.A. 2014. Isolating neural correlates of conscious perception from neural correlates of reporting one’s perception. *Frontiers in psychology* 5, p. 1078.
- Quiroga, R.Q., Nadasdy, Z. and Ben-Shaul, Y. 2004. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation* 16(8), pp. 1661–1687.
- Rees, G., Kreiman, G. and Koch, C. 2002. Neural correlates of consciousness in humans. *Nature Reviews. Neuroscience* 3(4), pp. 261–270.
- Rigotti, M., Barak, O., Warden, M.R., et al. 2013. The importance of mixed selectivity in complex cognitive tasks. *Nature* 497(7451), pp. 585–590.
- Safavi, S., Kapoor, V., Logothetis, N.K. and Panagiotaropoulos, T.I. 2014. Is the frontal lobe involved in conscious perception? *Frontiers in psychology* 5(1063), p. 1063.
- Scalaidhe, S.P., Wilson, F.A. and Goldman-Rakic, P.S. 1999. Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. *Cerebral Cortex* 9(5), pp. 459–475.
- Schmid, M.C. and Maier, A. 2015. To see or not to see--thalamo-cortical networks during blindsight and perceptual suppression. *Progress in Neurobiology* 126, pp. 36–48.

- Sheinberg, D.L. and Logothetis, N.K. 1997. The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences of the United States of America* 94(7), pp. 3408–3413.
- Szczepanski, S.M. and Knight, R.T. 2014. Insights into human behavior from lesions to the prefrontal cortex. *Neuron* 83(5), pp. 1002–1018.
- Tolias, A.S., Ecker, A.S., Siapas, A.G., Hoenselaar, A., Keliris, G.A. and Logothetis, N.K. 2007. Recording chronically from the same neurons in awake, behaving primates. *Journal of Neurophysiology* 98(6), pp. 3780–3790.
- Tsuchiya, N., Wilke, M., Frässle, S. and Lamme, V.A.F. 2015. No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends in Cognitive Sciences* 19(12), pp. 757–770.
- Vignal, J.P., Chauvel, P. and Halgren, E. 2000. Localised face processing by the human prefrontal cortex: stimulation-evoked hallucinations of faces. *Cognitive neuropsychology* 17(1), pp. 281–291.
- van Vugt, B., Dagnino, B., Vartak, D., et al. 2018. The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science* 360(6388), pp. 537–542.
- Webster, M.J., Bachevalier, J. and Ungerleider, L.G. 1994. Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cerebral Cortex* 4(5), pp. 470–483.
- Wei, M. and Sun, F. 1998. The alternation of optokinetic responses driven by moving stimuli in humans. *Brain Research* 813(2), pp. 406–410.
- Wheatstone, C. 1838. Contributions to the physiology of vision. part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London* 128(0), pp. 371–394.
- Wilson, F.A., Scalaidhe, S.P. and Goldman-Rakic, P.S. 1993. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* 260(5116), pp. 1955–1958.
- Wolfe, J.M. 1984. Reversing ocular dominance and suppression in a single flash. *Vision Research* 24(5), pp. 471–478.
- Zaksas, D. and Pasternak, T. 2006. Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *The Journal of Neuroscience* 26(45), pp. 11726–11742.
- Zaretskaya, N. and Narinyan, M. 2014. Introspection, attention or awareness? The role of the frontal lobe in binocular rivalry. *Frontiers in Human Neuroscience* 8(527), p. 527.