bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A CONNECTOME-BASED MODEL OF CONSCIOUS ACCESS IN MONKEY CORTEX

A PREPRINT

Ulysse Klatzmann[1,*], Sean Froudist-Walsh[1,*], Daniel P. Bliss[1], Panagiota Theodoni[1,2], Jorge Mejías[3], Meiqi Niu[4], Lucija Rapan[4], Nicola Palomero-Gallagher[4,5], Claire Sergent[6,7], Stanislas Dehaene[8,9], and Xiao-Jing Wang[1,+]

[1]Center for Neural Science, New York University, New York, NY 10003, USA
[2]Center for Mind, Brain and Consciousness, Department of Philosophy, New York University, New York City 10003, NY, USA
[3]Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands
[4]Research Center Jülich, INM-1, Jülich, Germany
[5]C. and O. Vogt Institute for Brain Research, Heinrich-Heine-University, 40225 Düsseldorf, Germany
[6]Université de Paris, INCC UMR 8002, 75006 Paris, France.
[7]CNRS, INCC UMR 8002, Paris, France
[8]Collège de France, 11 Place Marcelin Berthelot, 75005 Paris, France
[9]Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin Center, 91191 Gif/Yvette, France
[*]co-first authors
[+]corresponding author: xjwang@nyu.edu

May 6, 2022

Abstract

A growing body of evidence suggests that conscious perception of a sensory stimulus triggers an all-or-none activity across multiple cortical areas, a phenomenon called 'ignition'. In contrast, the same stimulus, when undetected, induces only transient activity. In this work, we report a large-scale model of the macaque cortex based on recently quantified structural connectome data. We use this model to simulate a detection task, and demonstrate how a dynamical bifurcation mechanism produces ignition-like events in the model network. Within this framework, the model predicts that feedforward excitatory transmission is primarily mediated by the fast AMPA receptors to ensure rapid signal propagation from sensory to associative areas. In contrast, a large fraction of the inter-areal feedback projections and local recurrent excitation depend on the slow NMDA receptors, to ensure ignition of distributed frontoparietal activity. Our model predicts, counterintuitively, that fast-responding sensory areas contain a higher ratio of NMDA to AMPA receptors compared to association cortical areas that show slow, sustained activity. We validate this prediction using *in-vitro* receptor autoradiography data. Finally, we show how this model can account for various behavioral and physiological effects linked to consciousness. Together, these findings clarify the neurophysiological mechanisms of conscious access in the primate cortex and support the concept that gradients of receptor densities along the cortical hierarchy contribute to distributed cognitive functions.

Keywords consciousness · large-scale brain model · connectome · NMDA · AMPA · feedforward · feedback · computational model · ignition · Global Neuronal Workspace · access consciousness

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

Among the huge flow of information received by our sensory organs, only a fraction of it is consciously perceived (James 1890). In many experiments that probe the access of stimuli to consciousness, subjects (human or non-human) are presented with faint stimuli, and asked to report if they detect them. The relationship between the strength of the stimulus and whether subjects report its presence is highly non-linear (e.g. Del Cul et al. 2007; Van Vugt et al. 2018). Increasing the brightness of a flashlight will increase the probability of detection, but the stimulus will always be either reported or missed. Once a stimulus is consciously perceived, an array of cognitive computations can be performed that are not available for unconsciously perceived stimuli (Trübutschek et al. 2019). Although this behavioral phenomenon has been consistently observed across species, the network, cellular and synaptic mechanisms of conscious access are hotly debated, and largely unresolved (Aru et al. 2020; Baars 2005; Dehaene et al. 1998; Lamme and Roelfsema 2000; Lau and Rosenthal 2011; Mashour et al. 2020; Tononi 2004).

Several studies have shown that, when a stimulus is consciously detected, activity emerges across many brain areas, and is sustained (for a few hundred milliseconds) principally in frontal and parietal cortices (Dehaene et al. 2001; Del Cul et al. 2007; King and Dehaene 2014; Rees et al. 2002; Sadaghiani et al. 2009; Sergent et al. 2005; Van Vugt et al. 2018). This widely distributed, sudden, and sustained activity has been termed 'ignition' (Dehaene et al. 1998, 2003). A common way to investigate ignition is to present a subject with stimuli of different strengths - such as contrast for visual stimuli. By manipulating the visual contrast, experimenters can identify the threshold for detection - the stimulus contrast at which stimuli can be detected on roughly 50% of trials (Green and Swets 1966; Nachmias 1981; Sergent and Dehaene 2003). Many studies have shown that neural activity in the early sensory areas grows approximately linearly with stimulus strength, regardless of whether the stimulus is detected or not (Del Cul et al. 2007; Kouider et al. 2013; Lafuente and Romo 2005, 2006; Romo and Rossi-Pool 2020; Van Vugt et al. 2018). On the other hand, activity in fronto-parietal areas, including the dorsolateral prefrontal cortex (dlPFC), shows a late all-or-none activation (Del Cul et al. 2007; Lafuente and Romo 2005, 2006; Sergent et al. 2005; Van Vugt et al. 2018). The neural dynamics of unconsciously and consciously perceived stimuli differ. Some authors have described how stimuli not consciously perceived elicit a chain of dynamically changing neural activity states, while conscious perception corresponds to the late onset of relatively stable patterns of neural activity (Schurger et al. 2015). Early simulations of a few interconnected brain areas reproduced this stable activity, but were not built on realistic cortical connectivity (Dehaene et al. 2003; Dehaene and Changeux 2005). Other studies describe a reliable, but dynamical trajectory of activity associated with conscious access (Baria et al. 2017; Salti et al. 2015), with the different interpretations possibly due to analysis of different frequency bands (He 2018). More recent modeling studies based on real cortical connectivity data have begun to investigate the dynamical propagation of stimulus information into the fronto-parietal network, but not applied to a detection task (Chaudhuri et al. 2015; Deco et al. 2020; Joglekar et al. 2018). A realistic large-scale model of the ignition phenomenon would create testable predictions for theories of consciousness and provide a concrete link between a subjective psychological phenomenon, large scale neural networks and synaptic interactions.

Several prominent theories of consciousness propose a central role of recurrent synaptic interactions between excitatory neurons (Dehaene et al. 2003; Lamme and Roelfsema 2000). However, the timescale of excitatory synaptic interactions differs drastically depending on the type of postsynaptic glutamatergic receptors. The most widely expressed of these are α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) and N-methyl-D-aspartate (NMDA) receptors. The decay time constant of AMPA receptors is up to 50 times shorter than that of NMDA receptors. The rapid onset and offset of excitatory currents mediated by AMPA receptors could help sensory areas react quickly to changes to external stimuli (Self et al. 2012; Yang et al. 2018). Once a stimulus has disappeared, the slow time-constant of NMDA receptors is crucial for sustaining stimulus-specific activity in prefrontal cortex (Wang et al. 2013; Wang 1999). Differences in receptor expression across cortical areas (Froudist-Walsh et al. 2021b; Goulas et al. 2021; Zilles and Palomero-Gallagher 2017) may contribute to differences in functional capabilities (Froudist-Walsh et al. 2021a), but to replicate distributed cortical activity and function in a large-scale system, we must also consider which type of connections are mediated by each type of receptor. Indeed, from their inception, the first simulations of ignition in a global workspace architecture (Dehaene et al. 2003; Dehaene and Changeux 2005) already included a key role for long-distance recurrent and top-down connections in the maintenance of long-lasting sustained ignition. Furthermore, the hypothesis was made that those connections, to support slow dynamics, must primarily involve synapses dense in NMDA receptors, a proposal that was later supported experimentally (Self et al. 2012). However, it is not known if feedback mediated by NMDA receptors is specific to V1 or can be generalised to the entire cortex. Moreover, it is unclear whether the theoretically hypothesized crucial role of the NMDA receptors is consistent with recent surprising reports that in human brain the ratio of NMDA to AMPA receptors actually decreases along the cortical hierarchy (Goulas et al. 2021; Zilles and Palomero-Gallagher 2017). It is not known whether a similar distribution of glutamatergic receptors exist in macaque monkeys, for which neural recordings are more widely available. Direct and indirect recording studies suggest that conscious access is associated with a mixture of dynamic and stable activity across

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

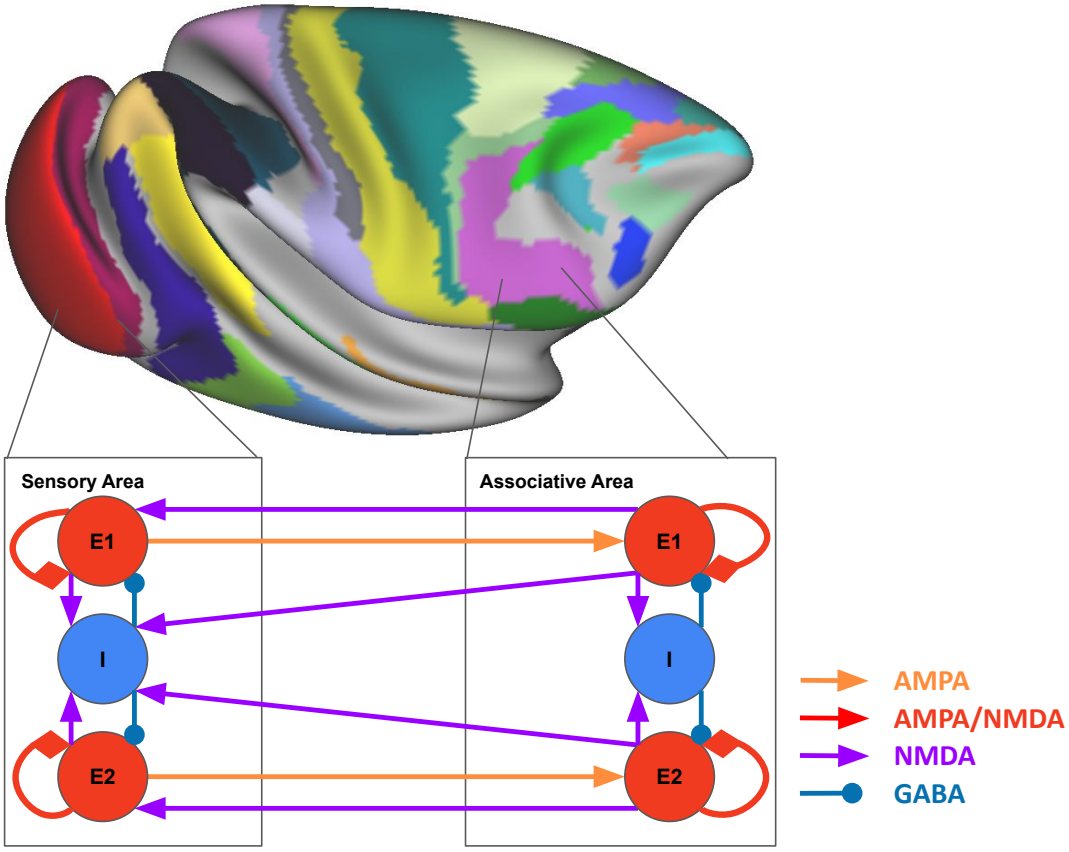

Figure 1: *A large-scale cortical model of stimulus detection in the macaque monkey. Each cortical area contains two populations of excitatory neurons (red) and a single inhibitory population (blue). The local circuit model is placed at each of 40 cortical areas (Top). Cortical areas are connected according to the weighted and directed connectivity data of Markov et al (2012) and Froudist-Walsh et al (2021a). In the model feedforward connections are predominantly via AMPA receptors (orange). Local excitatory-to-excitatory connections are mediated by both NMDA and AMPA receptors (red). Long-range feedback connections and local excitatory-to-inhibitory connections are primarily mediated by NMDA receptors (purple). Feedback connections also target inhibitory cells to a greater degree than feedforward connections. Local inhibitory connections are mediated by GABA receptors (blue).*

many brain areas (Dehaene et al. 2001; Del Cul et al. 2007; King and Dehaene 2014; Rees et al. 2002; Sadaghiani et al. 2009; Sergent et al. 2005; Van Vugt et al. 2018). Any neuroscientific theory of consciousness must account for such physiological phenomena given the constraints of the emerging anatomical data on the connectome and cortical glutamatergic receptor distributions.

In this paper, we develop a connectome-based dynamical model of the macaque cortex with realistic biophysical constraints and assess its behavior during a stimulus detection task, similar to that used experimentally. Secondly, we examined whether the parameter regime necessary for realistic model behavior was consistent with receptor distributions in the macaque cortex. Our model reproduces multiple aspects of monkey behavior and physiology, including aspects that have evaded previous models such as strong propagation of activity to prefrontal cortex, bifurcation dynamics and sustained distributed activity. Furthermore, sufficiently strong stimulus propagation and ignition requires NMDA/AMPA distributions across cortex that closely match those measured experimentally. Therefore our findings shed light on the synaptic and systems-level mechanisms underlying ignition and reconcile seemingly contradictory anatomical, physiological and modeling results.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Results

A large-scale cortical model with feedforward AMPAR and feedback NMDAR connections

We built a large-scale model of the macaque cortex containing 40 different interacting cortical areas (Fig 1, Top). Each cortical area contains a local circuit, with two populations of excitatory neurons and one population of inhibitory neurons (Fig. 1, Bottom) (Mejias and Wang 2022; Wong and Wang 2006). Excitatory connections are mediated by both $NMDA$ and $AMPA$ receptors, and inhibitory connections are mediated by $GABA_A$ receptors. Cortical areas differ in their pattern of inter-areal connections and the strength of excitatory input.

Cortical areas are connected according to weighted and directed inter-areal connections. The inter-areal connectivity data was collected by Kennedy and colleagues in a series of retrograde tract-tracing experiments (Froudist-Walsh et al. 2021a; Markov et al. 2014, 2012; Mejias et al. 2016). The inter-areal connections can be divided into feedforward and feedback connections. In general, feedforward connections emerge from the superficial cortical layers and target areas that are higher in the hierarchy. Feedback connections mostly emerge from the deep layers and target areas that are lower in the hierarchy (Barone et al. 2000; Felleman and Van Essen 1991; Markov et al. 2014). In each long-range connection, the fraction of neurons emerging from the superficial layers in the source area (known as the supragranular labeled neurons, or SLN) is thus a proxy measure of the 'feedforwardness' of a connection (Barone et al. 2000; Markov et al. 2014).

Feedforward and feedback connections may be mediated by different receptors (Dehaene et al. 2003; Self et al. 2012). Except where otherwise stated, feedforward connections in the model are mostly mediated by AMPA receptors. In contrast, feedback connections are principally mediated by NMDA receptors. The proportion of inter-areal excitation mediated by AMPA receptors in the model increases linearly with the SLN (Equations 14, 16, 17 and 18). Similarly, the proportion of inter-areal excitation mediated by NMDA receptors decreases linearly with the SLN. However, this is just one possible way in which feedforward and feedback connections may mediate their excitation. Later we explore alternative possibilities, and the effects of these choices on model behavior.

Feedforward and feedback connections may also target different cell types. If both feedforward and feedback connections mainly target excitatory cells, models often fail to propagate activity, or display global network synchrony and instability (Joglekar et al. 2018; Mejias and Wang 2022). This issue can be alleviated by assuming that feedback connections have a slight net bias towards targeting inhibitory cells in the so-called counterstream inhibition hypothesis (Mejias and Wang 2022). Therefore, in our model, feedforward connections principally target excitatory cells, and feedback connections have a bias towards inhibitory cells (Fig 1). Alternative scenarios are explored and analysed later.

Stimulus detection leads to widespread ignition of activity throughout the fronto-parietal circuit

We simulated a stimulus detection task by injecting differing, small amounts of external current to primary visual cortex (V1) (Fig 2A). On some trials no external stimulus was presented.

We begin by analyzing the response to the stimulus in different cortical areas. In V1, activity increased approximately linearly with stimulus intensity, before returning to baseline a few milliseconds after the stimulus was removed (Fig 2B). In contrast, on many trials activity in areas throughout the prefrontal and parietal cortices reached a high activity state at around 200ms (Fig 2B,C). This activity remained stable until the end of the trial, or until the vigilance signal was removed. The distributed network of regions showing late, sustained, activity resembles the core of the mesoscopic connectome (Markov et al. 2013). This pattern of prefrontal activity closely matches the dynamics of neural activity in monkey prefrontal cortex during a similar task (Van Vugt et al. 2018). In *in-vivo* experiments, this late ($\sim$200ms onwards) sustained prefrontal activity is a neural marker for the conscious detection of a stimulus (Van Vugt et al. 2018). We therefore interpret trials with late, sustained activity in area 9/46d of the prefrontal cortex as corresponding to detection of the stimulus. This interpretation allows us to define four trial types based on the combination of a stimulus presence and the detection.

Activity in the parieto-frontal circuit, but not sensory areas distinguishes hit from miss trials

The four trial types are as follows: 1) Hit trials are when a stimulus is presented and detected. 2) Miss trials are when a stimulus is presented but not detected. 3) Correct rejection trials are when no stimulus presented, and none is detected. 4) False alarm trials are when no stimulus is presented, but a detection is incorrectly reported.

We observed all four trial types in model simulations, which demonstrates how identical stimuli can produce dramatically distinct behavior. Activity in sensory areas was very similar for hit and miss trials (Fig 2B). Therefore,

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

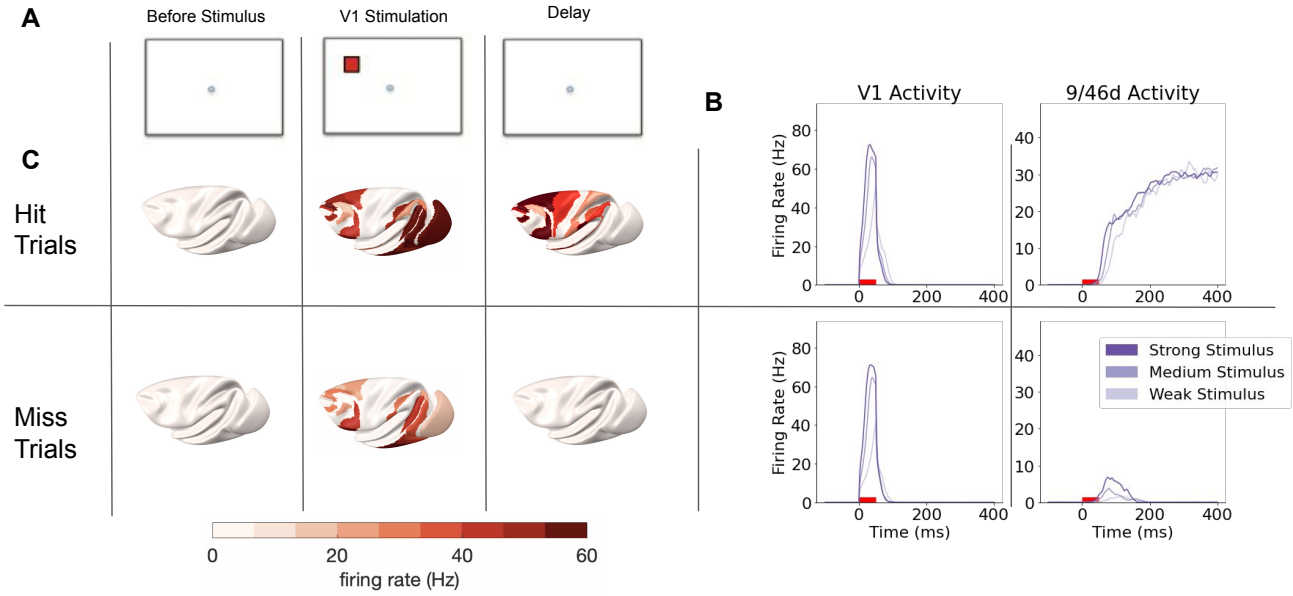

Figure 2: *Approximately linear sensory cortex responses and all-or-none prefrontal dynamics during stimulus detection. A) Task structure. A near-threshold stimulus is presented to the excitatory population in area V1. B) Top: Activity in V1 (primary visual cortex) and area 9/46d (prefrontal cortex) during a hit trial, for differing levels of stimulus intensity (200pA, 250pA and 300pa respectively for Weak, Medium and Strong stimulus strengths). V1 activity rapidly increases to a peak that differs according to stimulus intensity, before falling back to baseline. 9/46d activity in contrast reaches a high sustained activity state after about 200ms, which does not depend on stimulus intensity. Bottom: Activity in V1 and area 9/46d during a miss trial, for differing levels of stimulus intensity. V1 activity is very similar to that on the hit trial. 9/46d activity differs drastically, with a smaller peak of activity followed by a return to a low firing baseline state. C) Firing rates across the cortex during example hit (Top) and miss (Bottom) trials. Hit trials are accompanied by sustained activity throughout much of prefrontal and posterior parietal cortex, which is absent on miss trials. Note that in B) and C) stimulus intensity and network parameters are completely matched between hit and miss trials, which differ only in the random noise.*

regardless of whether the stimulus was detected or not, neural activity in sensory areas reliably tracked the objective stimulus strength. This closely matches experimental reports of sensory neurons responding to stimuli in both visual and somatosensory cortex (Lafuente and Romo 2005; Van Vugt et al. 2018). In contrast, activity throughout the fronto-parietal network differed drastically between hit and miss trials. While hit trials engaged strong, sustained activity throughout frontal and parietal cortices, miss trials led to either a transient increase in activity, which returned to baseline, or no increase in activity (Fig 2B). This closely corresponds to experimental findings (Romo and Rossi-Pool 2020; Van Vugt et al. 2018), validating the model.

Notably, depending on noise, the same stimulus strength could either lead to widespread sustained activity, or a return to baseline. In Figure 2C), we show two trials with the same stimulus strength. In the top example, strong, sustained activity is ignited throughout much of fronto-parietal cortex, and the stimulus is detected. In the bottom example, the same strength stimulus leads only to a transient, and weaker activity throughout the same network, before activity returns to baseline. Our model therefore predicts that the late stimulus-related activity, which was observed by Van Vugt et al. 2018 in dorsolateral prefrontal cortex and proposed as a marker of conscious access, should be detectable throughout a distributed prefrontal and posterior parietal cortical network.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

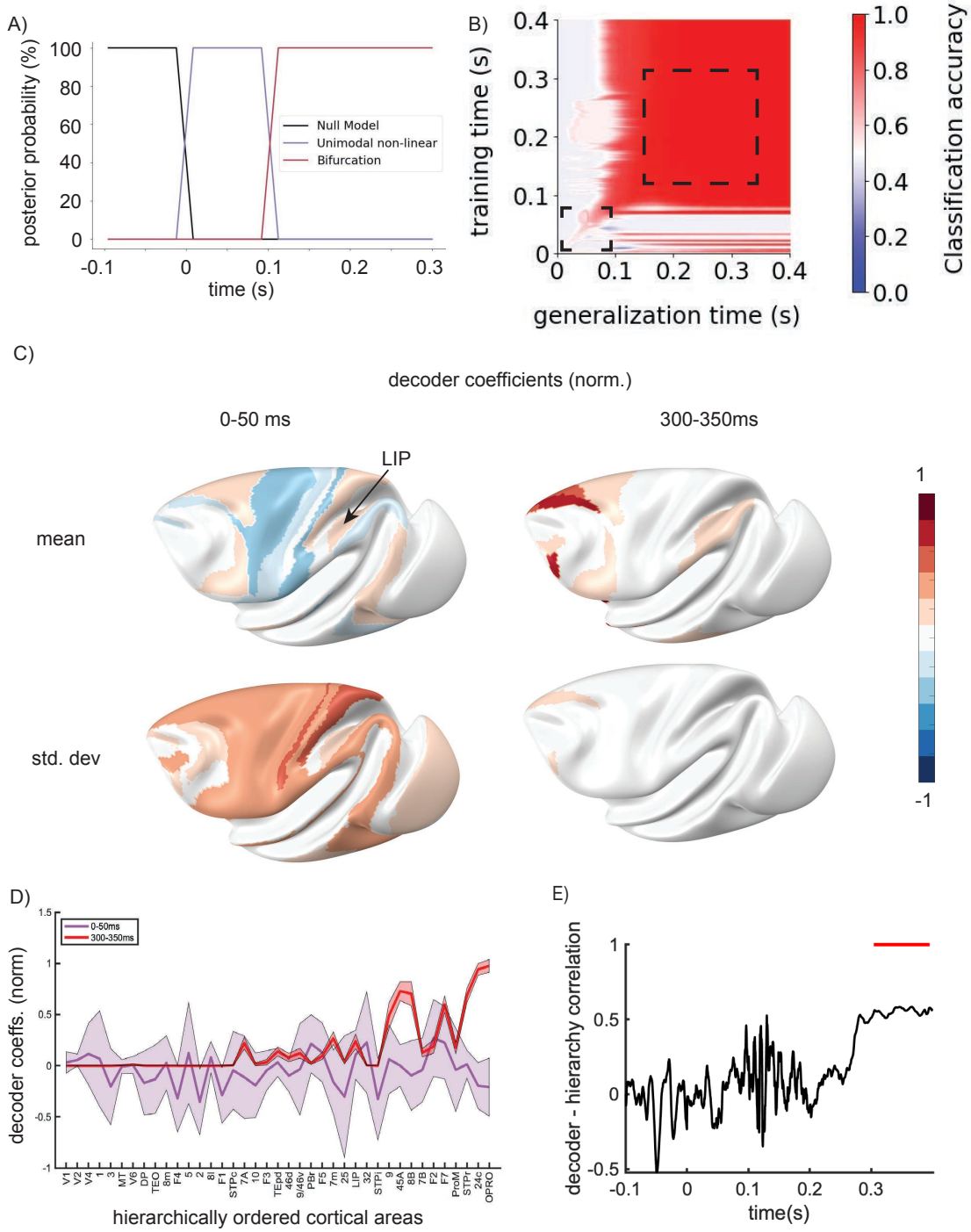

Figure 3: *A dynamic-to-sustained progression of activity states associated with ignition. A) Across-trial statistics of neural activity was used to classify model neural activity as belonging to null (black), unimodal (purple), or bimodal (red) distributions. Activity progresses from a null-distribution to unimodal and finally bimodal across-trial activity distributions, indicative of a bifurcation. B) Temporal generalization matrix. A classifier trained to decode trial outcome (hit/miss) from the activity pattern at each time point in a training data set is used to predict outcome based on the activity at each trial timepoint in held-out data. A diagonal pattern (e.g. in the lower-left dashed box) indicates a quickly-changing dynamical code. A square pattern (e.g. in the upper-right dashed box) indicates a stable code. C) Cortical surface representation of the mean and standard deviation of the normalized decoder coefficients for early (0-50ms) and late (300-350ms) periods of the trial. D) Mean (+,- SD) of the normalized decoder coefficients for early (0-50ms) and late (300-350ms) periods of the trial, as a function of the hierarchical position of each cortical area. E) Correlation (Pearson's r) between the decoder coefficients at each timepoint and the cortical hierarchy. The red bar shows the time-range with a statistically significant correlation.*

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A bifurcation in brain dynamics associated with ignition

The spatial and temporal signatures of conscious access are under intense debate (Block 2019). Sergent and colleagues recently performed a study with human subjects listening to auditory stimuli at various intensities around the threshold for detection, and measured electroencephalography (Sergent et al. 2021). In the first 200ms, they observed that cortical activity increased with increasing stimulus strength, irrespective of whether the stimulus was later detected or missed. When plotted across trials, this early activity created a unimodal distribution, and likely corresponded to pre-conscious processing (Dehaene et al. 2006). After ∼250ms activity either increased to a high activity state, or returned to a low activity state (Sergent et al. 2021). The high activity state corresponded to the conscious detection of the stimulus. In contrast, the low activity state occurred on trials in which the stimulus was not detected. The statistics of late activity (as a combination of high and low late activity trials) displayed a bimodal distribution. This bimodal distribution was particularly notable for stimulation around threshold values, where the same stimulus leads to hits or misses with roughly equal probability. Thus, Sergent and colleagues suggest that trial activity proceeds from a dynamic sequence of early states to one of two possible late activity states. Therefore the number, and/or the stability of internal states of neural activity appears to change over time. A change in the number or stability of internal states is known in non-linear dynamics as a bifurcation. The neural mechanisms underlying the dynamical emergence of such a bifurcation are not yet fully understood.

Following the methods of Sergent et al., we analyzed model activity at each timepoint across several trials, and examined whether activity across trials was best described by a null distribution (neural activity independent of the stimulus), unimodal distribution or a bimodal distribution. We estimated the posterior probability that each statistical model underlay the the trial-by-trial simulated activity of the connectome-based dynamical model. As expected, and seen experimentally, the null model best described activity before the stimulus. Shortly following the stimulus, activity was best described by a unimodal distribution. After approximately 100ms, the data were best described by a bimodal distribution (Fig 3A), which suggests the appearance of a bifurcation. This temporal progression of the across-trial activity from null, to unimodal, to bimodal distributions matches the experimental observations in humans detecting auditory stimuli (Sergent et al. 2021). In our model, we detect a bifurcation after about 100ms, which is earlier than in the human data. This may reflect the relatively faster propagation of activity to prefrontal cortex in the monkey compared to human brain. This may also reflect a lack of the retina-to-V1 pathway, explicit axonal delays, and synaptic transmission latencies in the model, which together could add up to a few tens of milliseconds. However, in our dynamical model the late activity reaches its peak in prefrontal cortex after 200ms (see Fig 2B). This broadly matches the timing of stimulus-induced prefrontal activity observed in monkey experiments, which is observable from $\sim 60ms$ following stimulus onset and peaks after $\sim 150-200ms$ (Bellet et al. 2022; Thorpe and Fabre-Thorpe 2001; Van Vugt et al. 2018). Therefore, our connectome-based dynamical model accounts for the temporal progression of activity states observed in the brain during stimulus detection tasks.

A dynamic-to-sustained progression of activity states associated with ignition

Previous studies of conscious perception have reported that neural dynamics evolve from a dynamic to a relatively stable activity pattern (Dehaene and King 2016; Schurger et al. 2015). To test the stability of a neural code, many authors use the temporal generalization method (Cavanagh et al. 2018; King and Dehaene 2014; Meyers et al. 2008; Wasmuht et al. 2018). As with many decoding methods, the trials are first separated into training and test sets. The classifier aims to predict some trial feature (such as whether a participant perceived the stimulus) based on neural activity. Particular to this method, a separate classifier is trained for each timepoint in the trial. The classifier trained at a particular timepoint in the training set is then tested using the activity at each timepoint in the test set. Thus, there are T classifiers (equal to the number of timepoints in every trial in the training set) and each classifier is associated with T different accuracies, one for each timepoint in the test set. The overall classifier performance can therefore be represented in a $T \times T$ temporal generalization matrix. We aimed to decode the trial outcome (hit/miss) from activity at each timepoint. We defined trial outcome based on activity in area 9/46d, and predicted this outcome using activity in all other cortical areas. Here high classifier accuracy is represented in red, low in white, and below-chance accuracy in blue. High accuracy that is restricted to a narrow band around the diagonal is indicative of a succession of different activity patterns that predicts stimulus visibility. In contrast, a square pattern indicates a stable activity pattern associated with stimulus visibility. We observed a succession of patterns coding for stimulus visibility in the early trial stages, with high classification accuracy remaining close to the diagonal (Fig 3B, lower left box). In the later trial period we observed a stable pattern, with the decoders trained between $\sim 100-400ms$ generalizing to all other timepoints within that range (Fig 3B, upper right box). Classifiers trained on some early timepoints had below-chance accuracy at decoding later timepoints (blue patches in Fig 3B). This indicates that early activity patterns are effectively reversed later in the trial. Similar results have been reported in the human experimental literature (King et al. 2016; Sergent et al. 2021). Our model suggests that the below-chance generalization from early to late timepoints may be due to higher associative areas sending net inhibitory feedback to areas that are lower in the visual hierarchy. Put

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

another way, the stable, ignited activity pattern can lead to a reversal of the activity patterns that occur during stimulus propagation.

We then examined the decoder coefficients in the early (0-50ms) and later (300-350ms) trial periods. We first rescaled coefficients of all decoders to the same (-1,1) range (by dividing each area's coefficient by the absolute value of the coefficient with the greatest magnitude for each decoder). In the early trial period (Fig 3C left, Fig 3D), we see that the standard deviation of decoder coefficients is high. This confirms that the activity patterns predictive of trial outcome change rapidly throughout the first 50ms. No regions in early sensory cortex have a high mean decoding accuracy, despite them having high activity at this stage (see Fig 2 B,C). This is due to stimulus having a very similar impact on activity in early sensory cortex in both hit and miss trials (See Fig 2B). Therefore, there is no sustained pattern of early post-stimulus sensory activity that is predictive of stimulus detection for trials of matched stimulus strength. Mean decoder strength is relatively high in parts of the later dorsal visual stream, such as the lateral intraparietal area (LIP) of parietal cortex. One possible interpretation is that stronger propagation along the later dorsal visual stream may separate hit from miss trials, even early in the trial. Another possible explanation is that pre-stimulus activity in these brain areas contributes to stimulus recognition (Baria et al. 2017; Podvalny et al. 2019; Van Vugt et al. 2018). The standard deviation of the decoder coefficients is high throughout a distributed network of regions early in the trial. This suggests a highly dynamic, widely distributed succession of activity patterns that is predictive of trial outcome, from very early in the trial.

In the late trial period (Fig 3C, right, Fig 3D), the mean decoder coefficients are relatively high throughout a distributed network of frontal, parietal and some temporal regions, while the standard deviation of decoder coefficients is low. This confirms that the activity patterns predictive of trial outcome are stable late in the trial, and identifies them as belonging to the fronto-parietal network. The standard deviation is moderate only in regions of frontal cortex to which activity propagates last. This demonstrates how a stable code throughout fronto-parietal cortex can coexist with a dynamic activity in some areas of cortex. This prediction can be tested experimentally.

The coefficients of the late decoder are higher in areas that are high in the cortical hierarchy (Fig 3D). A significant, and stable, correlation between decoder coefficients and the cortical hierarchy emerges late in the trial, after about 300ms (Fig 3E). This suggests that late in the trial activity in areas that are higher in the cortical hierarchy, rather than early sensory areas, is most important for stimulus detection.

The probability of detecting a stimulus increases nonlinearly with stimulus intensity

Due to the stochastic single-trial behavior, it is possible to analyze how the proportion of hit trials varies with stimulus intensity. Note that late activity always proceeded to either a high or low activity state, as seen in monkey and human experiments (Sergent et al. 2021; Van Vugt et al. 2018). We analyzed the proportion of trials for which strong sustained activity was engaged in area 9/46d of the prefrontal cortex (i.e. hits) for each stimulus intensity. The proportion of hits increased with the stimulus intensity, with a sigmoidal curve accurately fitting the data (Fig. 4). The sigmoidal relationship between stimulus intensity and detection has frequently been observed in monkeys and humans (Del Cul et al. 2007; Lafuente and Romo 2006; Van Vugt et al. 2018), further justifying the use of sustained activity in area 9/46d as a marker of detection in our model.

Non-linear dynamics underlying stimulus detection, false alarms and misses

To better understand the dynamics determining whether individual trials would result in a hit, a miss, a false alarm or a correct rejection, we built a simplified local model with a single area made of a single excitatory and a single inhibitory population (Fig. 5A). The equations are the same as in the full model, only the connectivity is different. Additionally, we focused on excitation mediated by the NMDA receptors. This reduces the system to two dynamical variables, corresponding to the synaptic variables S_{NMDA} and S_{GABA}. This simplification enables us to analyze the dynamics of individual areas by looking at their phase portraits.

We first analyze the dynamics for hit and miss trials (Fig 5Bi). If the system is in a particular activity state, the phase portraits show us how the activity will evolve over time. The phase portraits show the excitatory synaptic activity on the horizontal axis, and the inhibitory synaptic activity on the vertical axis. We show the nullclines for the excitatory (red curved line) and inhibitory (blue line) synaptic dynamics. The nullclines indicate the activity states for which a small increase in time will lead to no change in the excitatory (red curved line) or inhibitory (blue line) synaptic activity. The steady states of the system correspond to the points where the nullclines intersect. The system initially has two stable steady states, corresponding to low and high activity states (Excitation close to 0 and Excitation close to 0.5) and an unstable steady state (at about Excitation = 0.2). The stable steady states attract activity towards them, while the unstable steady state repels nearby activity to move towards one of the stable steady states.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

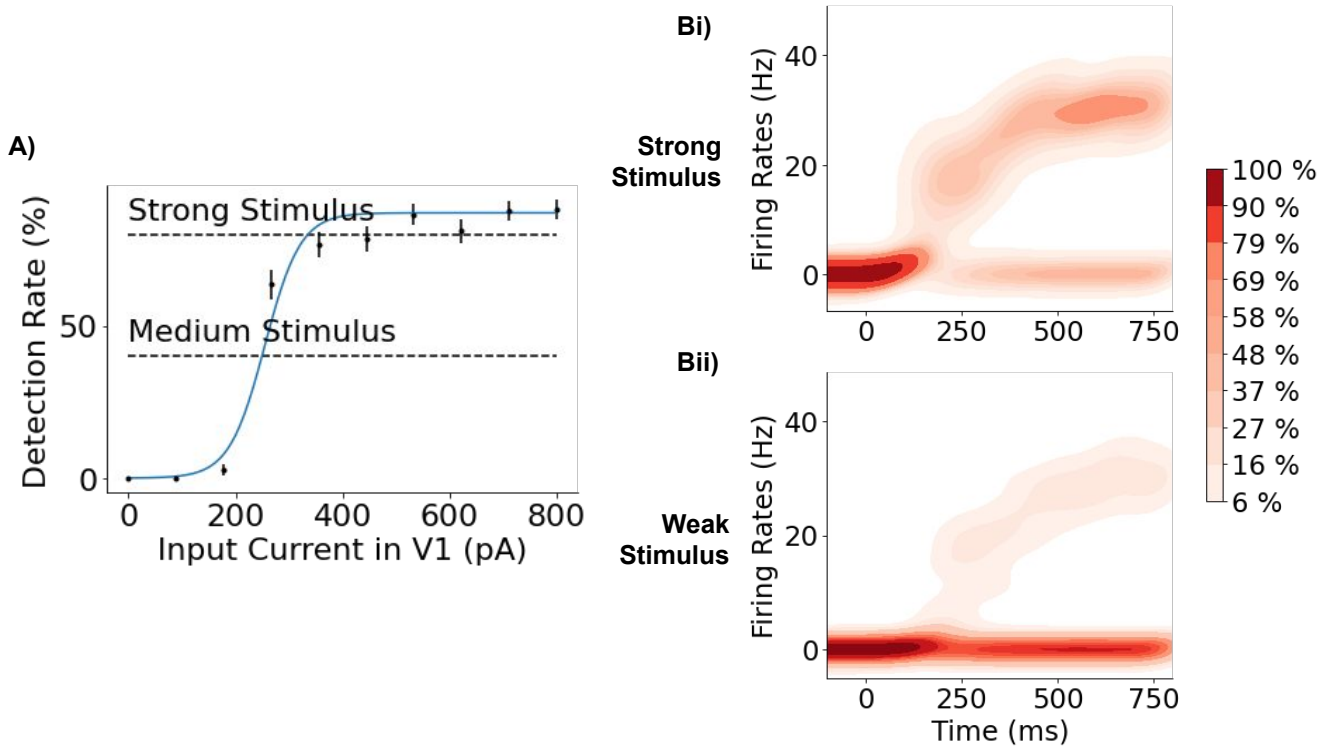

Figure 4: *A sigmoidal relationship between stimulus intensity and detection. A) The rate at which the large-scale cortical model detects the stimulus (engages sustained activity over 15Hz in area 9/46d) increases non-linearly with the stimulus intensity (input current in V1). B) The distributions of firing rates across trials for area 9/46d in the large-scale model for strong (Top) and weak (Bottom) stimuli. The outcome of individual trials is stochastic, depending on the noise in the system, but the system always ends in either a high activity state, corresponding to stimulus detection, or a low activity state, corresponding to a miss. A higher percentage of trials with a strong stimulus end in the high activity state compared to trials with a weak stimulus (as seen by the darker red in the high activity branch after 200ms). Stimulus is presented at 0ms for 50ms.*

The dynamics of single trials are illustrated in green and purple for example hit and miss trials, respectively. Before the stimulus, the system begins at the low steady state. The stimulus to the excitatory neural population shifts the excitation nullcline up. This reduces the number of nullcline crossings from three to one. The single remaining crossing represents a stable steady state, and activity is attracted towards this high activity state during the stimulus. This can be seen as both green and purple trajectories move towards the single nullcline crossing during the 'Stimulus on' period. Due to noise, the speed at which the activity increases towards the nullcline crossing differs. Here the green trajectory moves further towards the nullcline crossing in the same amount of time. When the stimulus is removed ('Early Post-Stimulus'), the nullclines rapidly shift back to their original position. As the unstable steady state (the middle nullcline crossing) repels activity away from itself, this effectively acts as a threshold. When the stimulus is removed, any activity to the left of the unstable steady state (such as the purple trajectory in Fig 5Bi) is repelled from the unstable steady state (middle crossing) and attracted back to the low activity steady state (left crossing), resulting in a miss. A firing pattern corresponding to these dynamics is shown in purple, in an example trial from the large-scale network (Fig 5Bii). At the time when the stimulus is removed, if activity is to the right of the unstable steady state (i.e higher excitation), the activity is again repelled from the unstable steady state (middle crossing), but now is attracted towards the high activity steady state (right crossing), leading to a hit. A stronger stimulus will lead to a larger shift in the excitatory nullcline, which increases the possibility of trajectories reaching the basin of attraction of the high activity state by the time the stimulus is removed. An example of this is shown by the green trajectory in Fig 5Bi and 5Bii.

The stimulus represents a strong, rapid change in input to the system. This strong input causes a large change in the nullclines, and a rapid increase in activity towards the high activity state. However, subtle shifts in the baseline neural firing rate also occur from trial to trial (Van Vugt et al. 2018). This is sometimes interpreted as representing different levels of 'vigilance'. We now analyze the effect of changes to the vigilance, implemented in the model as a small

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

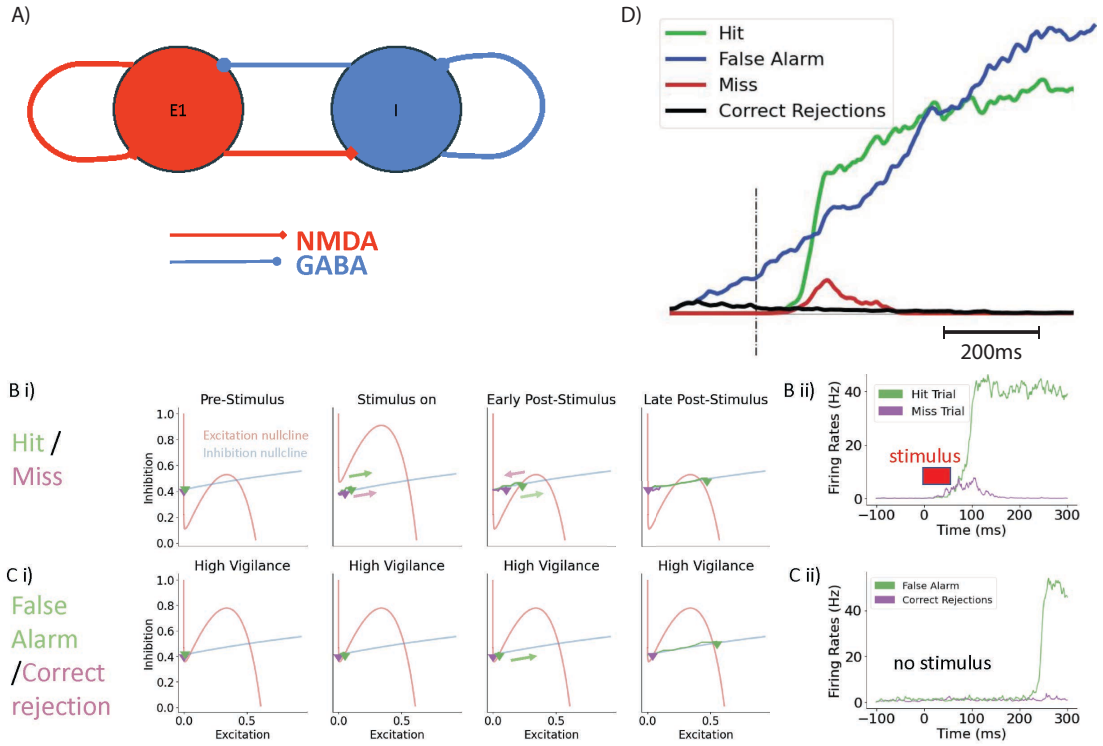

Figure 5: *Nonlinear dynamics underlying stimulus detection, misses, false alarms and correct rejections. A) Simplified circuit for analysis of dynamics, containing a single excitatory and a single inhibitory population, interacting via NMDA and GABA receptors. Bi) Phase portrait of the simplified network at different trial stages. Example dynamics for individual hit and miss trials are shown in green and purple, respectively. The stimulus causes the excitation nullcline (red) to move up, reducing the number of crossings with the inhibition nullcine (blue). Removal of the stimulus moves the nullclines back to the original positions. In the hit trial, by the time the stimulus has been removed, activity has reached the basin of attraction of the high activity steady state and progresses towards it (right nullcline crossing). In the miss trial, activity remains in the basin of attraction of the low activity steady state (left nullcline crossing), and returns towards it after removal of the stimulus. Bii) Firing rates from area 9/46d in the large-scale system on individual hit (green) and miss (purple) trials, corresponding to the dynamics shown in i). Ci) Phase portrait of the network with a strong vigilance signal in the absence of a stimulus. Two example individual trials are shown. The vigilance causes the excitation nullcline to shift up, to a smaller degree than during a stimulus. Cii) This can lead to an increase of activity to the high activity state, in the absence of a stimulus, corresponding to a false alarm (green). Activity on a correct rejection trial is shown in purple. D) Average activity across trials in dorsolateral prefrontal cortex for each trial type in the model (area 9/46d). Compare with the experimental data from dlPFC in Figure 2 of Van Vugt et al., 2018.*

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

increase in current to the excitatory population. In the brain, this could represent input from subcortical structures (such as the thalamus or sources of neuromodulatory transmitters) or perhaps top-down attention. In the following analysis, no stimulus is presented. For low vigilance, as in the pre-stimulus example above, the network is at the low activity steady state. An increase in the vigilance input gradually shifts the excitation nullcline up, to a lesser degree than the stimulus (Fig 5Ci). This reduces the basin of attraction for the resting state, and could even eliminate it. In this case the excitation nullcline remains extremely close to the inhibition nullcline, near the location of the previous low activity steady state. Activity that remains close to the low attractor state throughout the trial represents a correct rejection (Fig 5C, purple trajectories and activity trace). When the low steady state attractor no longer exists, we are very close in parameter space to its location. An attractor that is 'no longer with us' but still affects dynamics is known as a ghost attractor (Strogatz 2018; Vohryzek et al. 2020). Due to the proximity (in parameter space) of the attractor, it can still have an effect by slowing down the dynamics (Lo et al. 2015). On some of these trials without a stimulus, activity increases past the threshold for stimulus detection. This represents a false alarm (Fig 5Ci-ii, Green). The average activity in area 9/46d on false alarm trials appears to be a slow ramping up (Fig 5D i), in part because the simulations escape the ghost attractor at different times. This closely matches experimental data showing that the average activity in dorsolateral prefrontal cortex on false alarm trials appears as a slow upward ramp (see data on dlPFC in Figure 2 of Van Vugt et al. 2018). Therefore, the dynamic activity patterns of all four trial types can result from noise and transient bifurcations induced by the external stimulus or an internal vigilance signal.

Increased spontaneous activity increases probability of detection and false alarms

Would variations in the vigilance signal also induce false-alarm behavior in the connectome-based model? Vigilance was implemented as an additional current (0-95pA) to the excitatory populations of the associative areas of cortex (top 75% of cortical areas in the hierarchy). As before, the target stimulus was presented to one of the excitatory populations in V1. Again, the detection rate increased sigmoidally with the strength of the stimulus. The strength of the vigilance signal increased the slope of the sigmoid function (Fig 6A). This increase in slope results in a greater sensitivity to detect weak stimuli. For a very high vigilance signal (95pA) the intercept of the sigmoid with the vertical axis was raised (Fig. 6A). This means that for very high vigilance, the system is also more susceptible to false alarms (increased 'detection' rate at 0pA input current). This is consistent with the analysis of the simplified system, showing that vigilance signals can bring the system closer to a threshold for detection. Thus, with higher vigilance, weak stimuli are more likely to be detected, but the system is also vulnerable to erroneously detecting a stimulus when it is actually absent.

The baseline firing rate also rose with the increased vigilance signal input. We tested whether, for a constant noise level, shifts in the baseline firing rate were associated with an increase in false alarm trials. Several experimental studies have reported a relationship between baseline brain activity and either stimulus detection (Boly et al. 2007; Busch et al. 2009; Podvalny et al. 2019; Sadaghiani et al. 2009; Wyart and Tallon-Baudry 2009) or false alarms (Podvalny et al. 2019; Van Vugt et al. 2018). Indeed, we found a strong positive relationship between baseline firing rate and the false alarm rate. We understand this as follows, in a bistable system, if the low activity steady state is sufficiently far from the threshold (unstable steady state), then noise is unlikely to make activity jump across the threshold. However, if the low activity steady state (i.e. the baseline firing rate) is high enough to be close to the threshold, then the same amount of noise can enable it to occasionally cross the threshold into the attractor basin of the high activity attractor (Lo et al. 2015). Once it has crossed this threshold, it will inevitably continue to increase until it reaches the high activity steady state, which constitutes a false alarm.

We also examined how model behavior depended on the magnitude of the noise injected in both excitatory pools of all areas. The most noticeable effect of increasing the standard deviation of the noise is a large corresponding upward shift in the sigmoid function relating input current to detection. This also led to a more subtle decrease in the slope of the sigmoid (Fig 6C). In particular, this corresponds to a dramatic increase in the false alarm rate. The false alarm rate grew with the standard deviation of the noise, with a roughly linear increase between 5 and 7Hz, until it reached a plateau (Fig 6D). Thus, for a given low activity steady state (i.e. baseline firing rate), the probability of jumping to the attractor basin of the high activity steady state (i.e. a false alarm) increases with the standard deviation of the noise. Therefore, the simplified local cortical model with NMDA and GABA receptors provides a good intuition for some aspects of neural dynamics in the large-scale cortical model, and experimental data.

Fast propagation of stimulus information to prefrontal cortex depends feedforward excitation mediated by AMPA receptors

There is an abundance of fast AMPA receptors throughout the cortex. We investigated the role of AMPA versus NMDA receptors in the ignition phenomenon observed here. A previous modeling work has suggested that feedback connections targeting inhibitory neurons may be important for stabilizing a strongly recurrent cortical system (Mejias

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

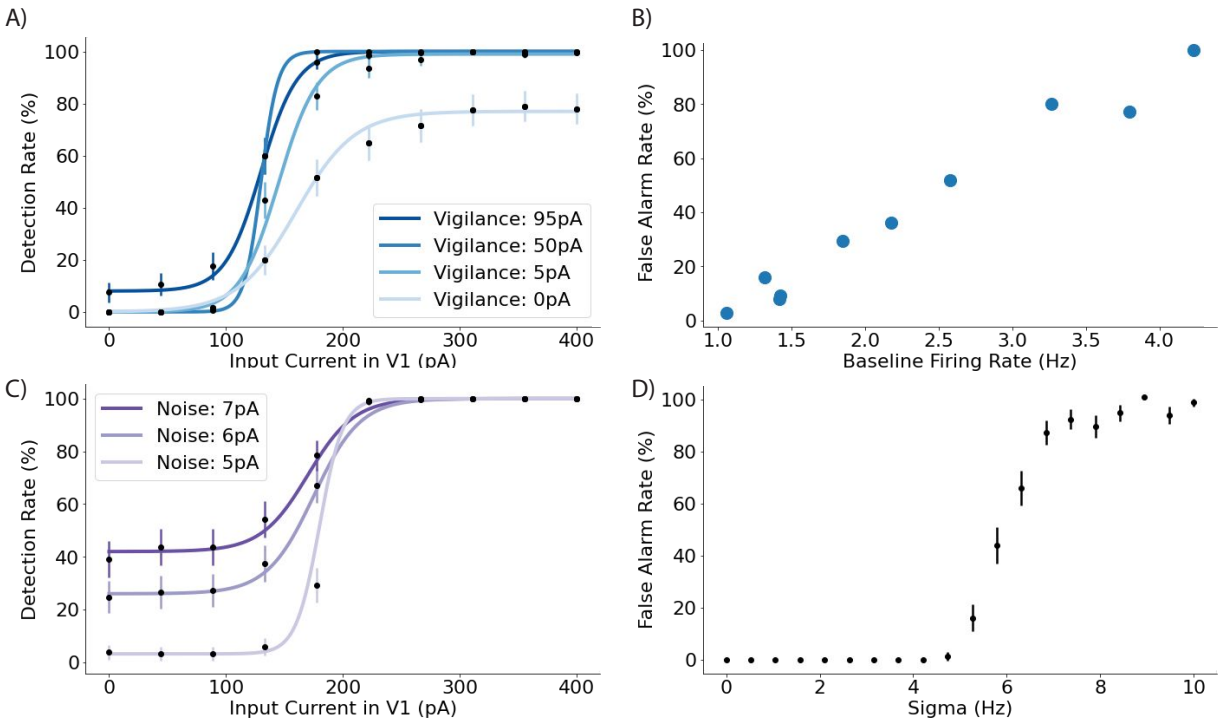

Figure 6: *Increased pre-stimulus activity increases probability of detection and false alarms. A) An increase in the vigilance signal in the model increases the sensitivity of the model to detecting weak stimuli. Very high levels of vigilance (95pA) also increase erroneous detection of non-existent (0pA) stimuli. B) The false-alarm rate increases approximately linearly with the pre-stimulus baseline firing rate. C) An increase in the standard deviation of the noise shifts up the stimulus-response curve, leading to a large increase in false alarms, and increased detection of weak stimuli. D) The false-alarm rate increases with the standard deviation of the noise.*

and Wang 2022). We investigated how model behavior changes according to whether inter-areal feedforward and feedback connections are mediated by AMPA or NMDA receptors on excitatory versus inhibitory cells (Fig 7A).

We found three distinct regimes of model behavior in response to a brief, strong 50ms stimulus. In one regime (Fig 7B, blue), activity in response to the stimulus is transient, before returning to baseline. The distribution of this activity across trials would be represented by a unimodal distribution, and is thus not consistent with the dynamics of conscious access (Sergent et al. 2021). The second regime (Fig 7B, red) corresponds to the entire cortex, including sensory and association areas showing sustained high activity. This is inconsistent with experimental results, showing that sensory cortical activity rapidly returns to baseline following the removal of a stimulus, with only a small positive modulation on correctly detected trials (Romo and Rossi-Pool 2020; Van Vugt et al. 2018). The final regime (Fig 7B, green) corresponds to a bistable state in which sensory areas show a transient response to the stimulus, and a distributed sustained response is seen across posterior parietal and frontal cortices. This is consistent with both the cross area temporal dynamics observed in the monkey cortex (Van Vugt et al. 2018), and the bimodal distribution of activity across hit and miss trials observed in the human brain (Sergent et al. 2021). Due to the close correspondence of this regime to theoretical proposals of the Global Neuronal Workspace theory, we name it the 'workspace bistability' regime.

The identified parameter blocks in which the workspace bistability is present suggest that feedforward connections should be principally mediated by AMPA receptors (Fig 7B, the workspace bistability regime appears to a greater extent in lower rows). If we increase the contribution of NMDA-mediated excitation to feedback connections (Fig 7B, moving from left to right columns), the workspace bistability regime grows to encompass a larger fraction of parameter space, and is thus more robust to small parameter changes. When looking within individual blocks of parameter space (Fig 7B, individual squares), it is clear that the workspace bistability regime depends on feedforward connections targeting mainly excitatory cells (Fig 7B, the workspace bistability regime is near the top of individual squares). In contrast, a large proportion, but not necessarily all, of the feedback connections should target inhibitory cells. This

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

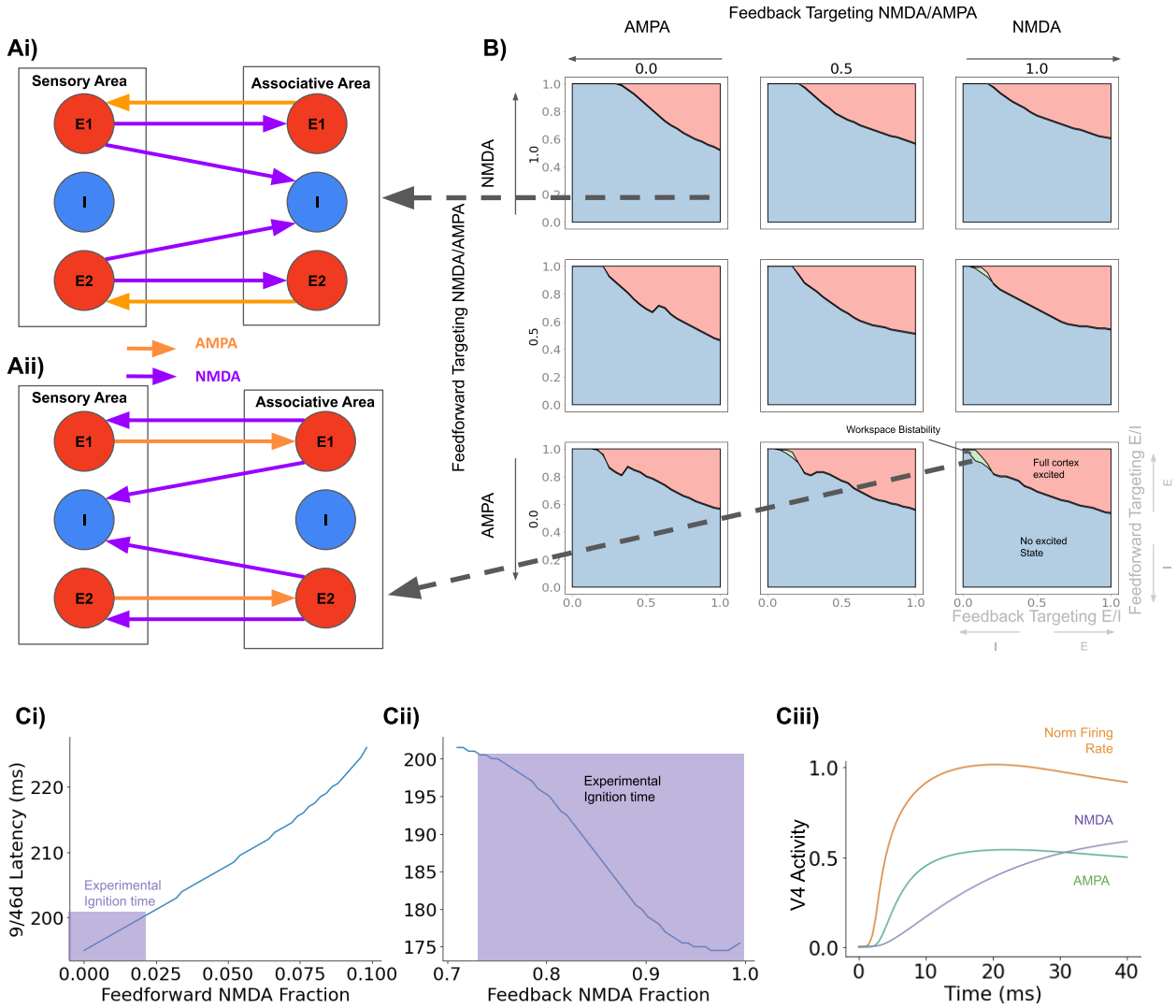

Figure 7: *Fast propagation of stimulus information to prefrontal cortex depends on AMPA receptors at long-range feedforward connections. A) Schematics showing two alternative inter-areal connectivity schemes. i) In this example, corresponding to the parameter set indicated in the upper-left box of B, feedforward connections are via NMDA receptors, and target a mix of excitatory and inhibitory cells. Feedback connections are via AMPA receptors, and target excitatory cells. ii) In this example, corresponding to the parameter set indicated in the lower right box of B, feedforward connections are via AMPA receptors, and target excitatory cells. Feedback connections are via NMDA receptors and mostly target inhibitory cells. This is close to the reference parameter set used throughout the majority of the paper. Note that local connections are omitted from this schematic for simplicity. B) Results of a 4-parameter search. The principal receptors mediating feedforward (top to bottom) and feedback connections (left to right) are shown across boxes. The cell-type targets of feedforward (top to bottom) and feedback (left to right) connections are shown within boxes. The resulting dynamical behaviors of the models are classified into three regimes when exposed to a 50ms strong stimulus input in V1, no excited state (blue), full-cortex excited (red) and workspace bistability (green). The workspace bistability regime corresponds most closely to experimental findings. C) Effects of AMPA and NMDA receptors on ignition times. i) The ignition time increases steadily, and rapidly passes the experimentally observed range, as the fraction of feedforward connections mediated by NMDA receptors increases. ii) In contrast, an increase of NMDA receptors at feedback connections is associated with a decrease in ignition times. iii) The dynamics of activity in area V4 (normalized firing rate) reflect the distinct dynamics of AMPA and NMDA receptors at feedforward and feedback connections.*

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

is in line with recent electrophysiology studies showing that feedback signals in the primate brain are net inhibitory, despite a focal excitatory component (Huang et al. 2019; Yoo et al. 2021).

In monkey experiments, frontal cortical activity ignites in response to sensory information in the macaque after approximately 130-200ms (Lafuente and Romo 2006; Thorpe and Fabre-Thorpe 2001; Van Vugt et al. 2018). In the model, we estimated the 'ignition time' as the time at which activity in dlPFC (area 9/46d) reached 95% of its maximum firing rate. We found that the ignition time was sensitive to the receptor type that mediated feedforward excitation. With an increase in the proportion of feedforward excitation mediated by NMDA instead of AMPA receptors, the ignition time rose to unrealistically slow levels (Fig 7Ci, left panel). Therefore, our model suggests that AMPA receptors at feedforward connections are crucial for the fast access of stimuli to conscious processing.

Surprisingly, an increase in the fraction of feedback connections mediated by AMPA receptors *increased* the ignition time. If the proportion of feedback connections mediated by AMPA instead of NMDA receptors was increased, again we began to see unrealistically slow ignition times (Fig 7Cii). To understand this further, we examined the normalized firing rate and currents through AMPA and NMDA receptors in area V4, which receives both feedforward and feedback inter-areal input (Fig 7Ciii). The firing rate increases rapidly with AMPA-dependent excitatory input. The firing rate gradually declines as NMDA receptors are engaged. This is due to a large proportion of feedback connections mediated by NMDA receptors being on inhibitory neurons. If feedback connections to inhibitory cells are mediated by AMPA receptors, the time window in which excitation exceeds inhibition is drastically shortened, effectively shutting down stimulus-related activity. Therefore, fast AMPA-dependent excitation and slow NMDA-dependent effective inhibition allows the feedforward excitation to transiently 'escape' the inhibition and successfully propagate stimulus-related activity along the cortical hierarchy.

The NMDA/AMPA ratio decreases along the cortical hierarchy

Near the top of the cortical hierarchy, areas receive a greater proportion of feedforward than feedback inter-areal connections. Conversely, areas near the bottom of the hierarchy (such as early sensory cortex) receive a greater proportion of inter-areal feedback connections. This would predict that the proportion of NMDA receptors (as a fraction of glutamate receptors) may actually decrease along the cortical hierarchy. We calculated the NMDA fraction (as a fraction of NMDA and AMPA receptor-mediated excitation) in the model for the reference parameter set used throughout the paper. We compared the model NMDA fraction to a cortical hierarchy of 40 areas derived from laminar patterns of inter-areal connections (Froudist-Walsh et al. 2021a; Markov et al. 2014). In the model, we observed a strong decrease in the NMDA fraction along the cortical hierarchy (Fig 8A, $r = -0.71, p = 3 \times 10^{-7}$). Therefore, in apparent contrast to previous physiological and computational studies that have emphasized the importance of AMPA receptors in sensory cortex (Self et al. 2012; Yang et al. 2018) and NMDA receptors in prefrontal cortex (Van Vugt et al. 2020; Wang et al. 2013; Wang 1999; Yang et al. 2018), our model predicts a decrease in the NMDA fraction along the macaque monkey cortical hierarchy.

We tested this prediction by analyzing *in-vitro* receptor autoradiography data from 109 regions of macaque cortex (Froudist-Walsh et al. 2021b; Impieri et al. 2019; Niu et al. 2020, 2021; Rapan et al. 2021). By dividing the receptor density by the neuron density in each area (Collins et al. 2010; Froudist-Walsh et al. 2021a,b), we were able to estimate the NMDA and AMPA density per neuron in each area (Fig 8B). We found that both the NMDA ($r = 0.70, p = 1 \times 10^{-5}$) and the AMPA ($r = 0.80, p = 3 \times 10^{-8}$) receptor densities per neuron increased along the cortical hierarchy (Fig. 8C). We defined the 'NMDA fraction' in each area as the NMDA receptor density divided by the sum of the NMDA and AMPA receptor densities. Despite the increases in both NMDA and AMPA densities along the hierarchy, there was a strong negative correlation between the NMDA fraction from the experimental data and the cortical hierarchy (Fig 8D,E, $r = -0.81, p = 2 \times 10^{-8}$), confirming our model prediction.

The decreasing NMDA/AMPA gradient supports ignition

We then adjusted the model parameters to match the experimentally observed NMDA and AMPA receptor densities. In Figure 7 we demonstrated how the NMDA fraction at feedforward and feedback connections critically determines ignition time. We therefore maintained the NMDA fraction at feedforward and feedback connections from the reference model. However, we allowed the NMDA fraction at local connections to vary across areas, so that the overall NMDA fraction of each area in the model closely matched that observed in the receptor autoradiography data. Without changing any other parameters, we observed that this receptor data-based simulation displayed workspace-bistability dynamics (Fig 8F). Specifically, in the receptor data-based model, transient activity was observed in response to the stimulus in visual areas, and distributed sustained activity was observed in posterior parietal and prefrontal cortex. Therefore, a counterintuitive decrease in the NMDA fraction along the hierarchy may have evolved to enable the dynamic inter-areal interactions required to support ignition-like dynamics.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

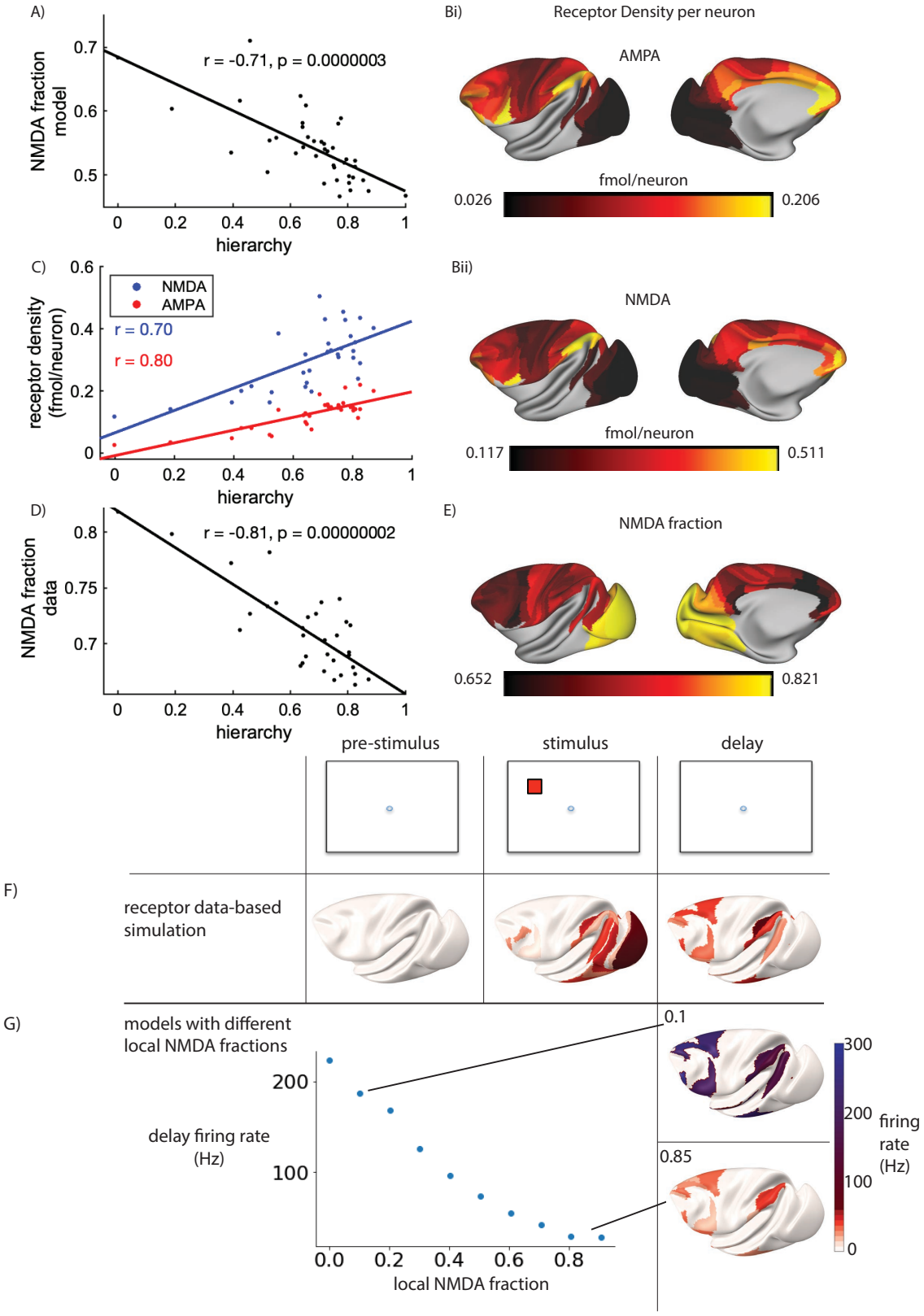

Figure 8: *The NMDA/AMPA ratio decreases along the cortical hierarchy and supports ignition. A) The fraction of excitatory inputs via NMDA receptors (compared to total NMDA+AMPA inputs) in the model decreases along the hierarchy. B) The density of i) AMPA and ii) NMDA receptors across 109 regions of macaque cortex. Receptor density was measured using in-vitro receptor autoradiography, and divided by the neuron density data from Collins et al., 2010. C) AMPA and NMDA densities per neuron both increase along the cortical hierarchy. D, E) The fraction of NMDA receptors (compared to total NMDA+AMPA receptors) in the macaque receptor autoradiography data decreases along the hierarchy. F) The model was adjusted to matched the receptor densities observed in the autoradiography data. The receptor data-based model shows ignition of fronto-parietal activity in response to a visual stimulus. G) Left. Average firing rate in areas showing delay activity for models with different local NMDA fractions. Right. Delay period activity across the cortex for models with different local NMDA fractions (Top: local NMDA fraction = 0.1, Bottom: local NMDA fraction = 0.85). Only the models with a relatively high fraction of local excitation mediated by NMDA receptors can reproduce realistic fronto-parietal activity levels.*

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Ignition depends on NMDA receptor activation in local excitatory connections

Cortical dynamics depend on both inter-areal and within-area (local) interactions. Does ignition also depend on the receptors that mediate local excitation? We adjusted the model so that the NMDA fraction of local excitatory connection varied. For a very low NMDA fraction, we see sustained activity in the fronto-parietal network, but at unrealistically high firing rates (area 9/46d delay activity = 213Hz for local NMDA fraction = 0.1, Fig 8G, Right Top). Only models with a relatively high fraction of local excitatory connections mediated by NMDA receptors showed workspace bistability and sustained activity in the fronto-parietal network at a reasonable firing rate (area 9/46d delay activity = 35Hz for local NMDA fraction = 0.85, Fig 8G, Right Bottom). However, if local connections were completely mediated by NMDA receptors, bistability was lost. This suggests that a small contribution of AMPA-mediated excitation is helpful to engage NMDA-mediated excitation and sustained activity. Therefore, in addition to feedforward AMPA and feedback NMDA connections, our model suggests that NMDA receptors at local excitatory connections are crucial for ignition of cortical activity in response to a stimulus, in support of the previous theoretical (Wang 1999) and experimental (Wang et al. 2013) findings.

Discussion

We developed a novel large-scale model of monkey cortex and used it to simulate a detection task. The present work builds upon a previously proposed global neuronal workspace (GNW) architecture (Dehaene et al. 2003; Dehaene and Changeux 2005), but now incorporating newly available weighted and directed cortical connectivity data (Froudist-Walsh et al. 2021a; Markov et al. 2014, 2012; Markov et al. 2013; Mejias et al. 2016) and receptor data (Froudist-Walsh et al. 2021b; Impieri et al. 2019; Niu et al. 2020, 2021; Rapan et al. 2021). Using this model, we replicated multiple spatial and temporal features of neural activity that are observed during conscious perception. The close correspondence of our simulations to experimental results enabled us to identify candidate synaptic and network level mechanisms underlying conscious perception. Our model predicts that the rapid ignition of fronto-parietal activity that accompanies conscious perception depends on feedforward inter-areal excitation being primarily mediated by AMPA receptors. However, AMPA receptors are not sufficient to reproduce experimental activity patterns, as networks that rely on AMPA receptors for local excitation cannot reproduce realistic firing rates in fronto-parietal cortex. We show that sustained fronto-parietal activity depends critically on the NMDA receptors. Our modeling results led to the surprising prediction that the NMDA/AMPA ratio should decrease along the cortical hierarchy. We confirmed this model prediction by analyzing *in-vivo* receptor autoradiography data across dozens of cortical regions. Finally, we show how this decreasing NMDA/AMPA ratio supports the fast propagation of stimulus-related information along the hierarchy, and ignition of a distributed fronto-parietal network, as is seen during conscious perception in both human and nonhuman animals.

Network, cellular, synaptic mechanisms of conscious access

Our model displayed behavior and cortical activity patterns that fit well with several results from the experimental literature. Notably, early sensory cortex displayed a transient, approximately linear response to the stimulus. This contrasted with the late, sustained, all-or-none response in prefrontal cortex, reproducing the experimental findings in a monkey experiment specifically designed to test the GNW theory (Van Vugt et al. 2018). Our model suggests that similar transient signals should be detected throughout the early visual system, while an all-or-none response should depend on the fronto-parietal circuit. This all-or-none distributed activity broadly lends support to the GNW framework (Dehaene et al. 1998, 2003), in which a strongly interconnected network of prefrontal and parietal neurons broadcasts information widely throughout the brain. Some models of the GNW theory suggest that this fronto-parietal network would excite and sustain the original stimulus representation in sensory neurons after the stimulus has disappeared (Dehaene et al. 2003). While such delayed activity in early sensory cortex has been observed in some studies (Harrison and Tong 2009; Kerkoerle et al. 2017; Supèr et al. 2001), the consistency with which it occurs is not clear, and many studies have not observed sustained neural activity (Leavitt et al. 2017), or only a small modulation (e.g. Van Vugt et al. 2018). To reproduce these latter experimental activity patterns, we found that feedback connections must largely target inhibitory neurons. This is consistent with recent reports that top-down attention has a net-inhibitory effect on sensory cortex (Huang et al. 2019; Yoo et al. 2021) and with previous models predicting the need for feedback inhibition to sustain distributed persistent activity (Mejias and Wang 2022). A parsimonious explanation of the contrasting data on early sensory activity is that many tasks can be performed by maintaining a representation of only higher-order aspects of the stimuli, and therefore sustained activity is not required in early sensory cortex. This frees early sensory cortex to encode new sensory information. However, a focal top-down excitatory connection onto excitatory neurons in the target sensory area may be engaged, for example when attention to precise sensory details is required. The mixed excitatory and inhibitory effects of top-down connections can be addressed in more detail in future large-scale models with more neural populations per area (Ardid et al. 2007). Such models could shed light

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

on how different cognitive states may affect the dynamics (Melloni et al. 2011) and activity patterns across cortical regions, and indeed subcortical structures.

The network mechanisms of conscious perception themselves depend on cellular and synaptic mechanisms. Our model suggests that at the synaptic level, NMDA receptors are crucial for the ignition phenomenon. NMDA-dependent excitation is affected by various neuromodulators, which can enable (Yang et al. 2013), enhance (Arnsten et al. 2020; Galvin et al. 2020; Vijayraghavan et al. 2007), or counteract (Arnsten et al. 2020, 2019; Vijayraghavan et al. 2007) its effects. NMDA-spikes are relatively localized within pyramidal cell dendrites, but they increase the probability of plateau-like calcium spikes, which propagate a much greater distance towards the soma (Aru et al. 2020). This dendrite-soma coupling fades under anaesthesia (Suzuki and Larkum 2020). Moreover, activation of the apical dendrites of subcortically-projecting layer 5 pyramidal cells in the mouse somatosensory cortex is robustly associated with stimulus detection (Takahashi et al. 2020). Activation of these layer 5 pyramidal cells could trigger the subcortical vigilance signal required for ignition of the distributed cortical network. Therefore, our finding of the importance of NMDA receptors and a (potentially subcortical) vigilance signal to ignition is consistent with multiple converging lines of evidence, which are illuminating the network, cellular and synaptic mechanisms of conscious processing. In the future, the development of large-scale network simulations with spiking neurons (Brette et al. 2007; Joglekar et al. 2018), linking from ion channels to dendritic, laminar, areal and brain-wide dynamics (Goldman et al. 2019) across species (Sanz Leon et al. 2013) should help link across the scales of neuroscience and provide a platform to test various theories of consciousness (Doerig et al. 2021; Seth and Bayne 2022; Signorelli et al. 2021).

A dynamic-to-stable transition of cortical activity during conscious perception

Although more complex models may be valuable, we show that even a reduced two-variable model can provide useful intuitions to explain conscious and unconscious neural dynamics. The dynamics of neural activity in response to stimuli consistently follow one of two patterns. Sergent and colleagues analyzed the inter-trial variability of cortical activity while humans performed an auditory stimulus detection task (Sergent et al. 2021). Regardless of whether the stimulus was detected or not, in the first 250ms, neural activity followed a reliable trajectory. Following this point activity either proceeded towards a high, or a low state, which corresponded to conscious or unconscious processing of the stimuli (Sergent et al. 2021). The transition in time could be appreciated by activity across trials at each timepoint in the first 250ms being described by a unimodal distribution, whereas after that point the two distributions could be observed, separating trials in which the stimulus was missed from those in which it was consciously perceived. Our model showed the same transition in time from unimodal to bimodal across-trial distributions. Previous studies in humans have also shown how conscious perception is accompanied by an initial dynamic chain of activity states, before a later more stable encoding phase (Dehaene and King 2016; Del Cul et al. 2007; Schurger et al. 2015; Sergent et al. 2005). Our model provides a simple explanation for both findings. Early stimulus processing, in both unconscious and conscious mode, involves propagation of stimulus-specific information along the sensory hierarchy, in which individual sensory areas are transiently engaged. However, this dynamic activity probably remains pre-conscious. As the stimulus propagation is very similar on hit and miss trials, activity across trials can be well described by a unimodal distribution at this stage. Conscious processing then differs from unconscious processing by engaging sustained activity throughout the fronto-parietal network. This activity is relatively stable across time, and due to its all-or-none nature, trial-to-trial late activity follows a bimodal distribution. While several studies have reported an association between stable late activity and conscious access, others have reported that stable activity is only seen in relatively high frequencies (β and γ), and is accompanied by a reliable dynamic trajectory in very slow frequencies (δ and sub-δ, Baria et al. 2017; He 2018). It is possible that stable and dynamic activity coexist, if they belong to distinct subspaces (Murray et al. 2017). We do not explicitly try to model this very slow cortical dynamic activity here. Our model is agnostic as to whether the entire late distributed activity pattern represents conscious access alone, or whether some of the activity could prepare a conscious report, or other post-access conscious processing (Cohen et al. 2020; Dellert et al. 2021; Pitts et al. 2018, 2012, 2014; Schlossmacher et al. 2020; Sergent et al. 2021). Simulation of activity patterns during cognitive tasks in connectome-based dynamical models is in its infancy, and the types of tasks that have been simulated are particularly simple (Froudist-Walsh et al. 2021a; Mejias and Wang 2022). Future models may attempt to tackle more complex task structures and disambiguate patterns of activity related to conscious access and distinct post-access processes. On some missed trials we could observe transient activity throughout the fronto-parietal network, that died back to baseline. This is consistent with recent reports in humans that the spatial activity pattern for unrecognised and recognised stimuli is similar, and differs mainly by the magnitude of activation (Levinson et al. 2021), and monkey physiology studies showing transient prefrontal activity on some miss trials (Van Vugt et al. 2018).

Previous large-scale models have tackled stimulus propagation and cortical ignition (Chaudhuri et al. 2015; Deco et al. 2020; Dehaene et al. 2003; Dehaene and Changeux 2005; Joglekar et al. 2018; Schmidt et al. 2018). However, the earliest of these models (Dehaene et al. 2003; Dehaene and Changeux 2005), while mimicking the bimodal dynamics

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

of conscious access, relied on a highly abstracted view of thalamo-cortical circuits. Indeed, they included only 4 simulated areas, and while they modeled AMPA, NMDA and GABA receptors with realistic dynamics, their connectivity was not constrained by actual monkey or human data. Thus, while several of the ignition phenomena reported here were already present, these more abstract models did not easily make a connection to empirically measured physiological parameters. The second generation of previous models, on the other hand, did take into account empirical data on the connectivity of the brain (Chaudhuri et al. 2015; Deco et al. 2020; Joglekar et al. 2018; Schmidt et al. 2018). However, unlike the present work, these models all produce transient, and relatively weak responses in prefrontal cortex. It is likely none of these previous large-scale models would be capable of reproducing the unimodal-to-bimodal transition pattern of temporal dynamics observed in this model and the data of Sergent and colleagues (Sergent et al. 2021) and previous experimental work (Del Cul et al. 2007; Sergent et al. 2005). Our model shows multistability (Golos et al. 2015; Strogatz 2018), in which a stimulus induces a dynamic bifurcation, which leads activity to evolve to one of two steady states. This bears some resemblance to recent large-scale models of working memory (Froudist-Walsh et al. 2021a; Mejias and Wang 2022). While working memory signals generally persist for a period of a few seconds, conscious perception can be much shorter, on the order of 200ms. Our results suggest a shared mechanism between ignition and persistent activity associated with working memory. Behaviorally, conscious access can be brief and incessantly switch to different percepts or thoughts, whereas working memory storage is longer lasting. However, the difference between the two may not be fundamental. Indeed, a classifier trained to decode the visibility of a stimulus in a conscious perception task from MEG activity can generalise to decoding visibility from the early stages of activity during a working-memory delay period (Trübutschek et al. 2017). The differences in later activity patterns could be explained by how a sustained signal is turned off in the brain, such as by a corollary discharge signal during a motor response. This theoretical hypothesis remains to be tested experimentally. Other aspects of the neural dynamics that separate conscious from unconscious processing, including the complexity of pre- and post-stimulus neural activity (Goldman et al. 2019), and transient synchronisation (Melloni et al. 2007) were not investigated here. These types of dynamics may be related to attractors other than steady states, such as chaotic attractors and limit cycles (Wang 2021). In the working memory models, excitatory interactions are dominated by NMDA receptors. Our analysis of the current model supports a previous proposition that feedforward connections should principally be mediated by AMPA receptors, while feedback connections should principally be mediated by NMDA receptors (Dehaene et al. 2003), and suggests that this arrangement is crucial for realistically fast ignition of fronto-parietal activity during conscious perception. This led to the additional counterintuitive prediction that the proportion of NMDA receptors (compared to AMPA receptors) should decrease along the cortical hierarchy.

The NMDA/AMPA receptor ratio decreases along the macaque monkey cortical hierarchy.

We found, using *in-vitro* receptor autoradiography data across the macaque cortex, that the ratio of NMDA to AMPA receptors decreased along the cortical hierarchy, in agreement with our modeling prediction. This anatomical finding is corroborated by recent findings of a similar reduction in the NMDA/AMPA ratio along the cortical hierarchy in the human brain (Goulas et al. 2021; Zilles and Palomero-Gallagher 2017). However, many aspects of the NMDA and AMPA receptor distribution in the cortex remain to be characterised. For example, the *in-vitro* receptor autoradiography technique is blind to the subcellular location of the receptors, and whether they are synaptic, or extra-synaptic. It is also possible that distinct gradients of expression may emerge in different cortical layers. We recently described how cortical receptor distributions in the macaque are dominated by a principal receptor gradient, which is aligned with the cortical hierarchy (Froudist-Walsh et al. 2021b). A similar receptor gradient has also recently been described in human cortex (Goulas et al. 2021). We show that both NMDA and AMPA receptors per neuron increase along this gradient, but variations in the intercept and slope of the gradient across receptors can have important functional implications. In particular, our model suggests that the reduction in the NMDA/AMPA ratio along the hierarchy reflects an asymmetry in how feedforward and feedback excitatory connections are mediated by these receptors, and that this asymmetry is crucial for fast access of stimuli to consciousness.

Asymmetric feedforward and feedback excitation via AMPA and NMDA receptors reconciles contrasting anatomical and physiological findings

We show that ignition of distributed fronto-parietal activity can occur rapidly, when inter-areal feedforward excitation is predominantly mediated by AMPA receptors. In the model, when feedforward excitation was mediated by AMPA receptors, ignition occurred within about 200ms. This figure closely matches the timescale of ignition of prefrontal activity in monkeys performing a stimulus detection task (Van Vugt et al. 2018), and is similar to (but slightly quicker than) proposed neural signals of conscious perception in the larger human brain (Del Cul et al. 2007; Salti et al. 2015; Sergent et al. 2005; Sergent et al. 2021). Realistically fast access of stimuli to consciousness therefore necessitates a relative increase of AMPA receptors in cortical areas that receive feedforward connections from the visual system.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Physiological and computational studies of recurrent excitation in local circuits have emphasized the role of AMPA receptors in primary visual cortex (Yang et al. 2018) and NMDA receptors in dlPFC (Wang et al. 2013; Wang 1999). Reliance on the fast-acting AMPA receptors may facilitate the rapid encoding of stimulus information in V1, while the slow dynamics of the NMDA receptor are thought to be crucial for the characteristic persistent activity seen during working memory tasks. By going beyond local circuit considerations, we showed in this work that the density of both NMDA and AMPA receptors per neuron increases along the cortical hierarchy (Fig 8B), which does not contradict previous local area studies. The increasing density of NMDA receptors along the hierarchy, may account for the commonly observed persistent activity in dlPFC, that is largely absent in V1 (Leavitt et al. 2017). This parallels the increases in dendritic spines on layer 3 pyramidal cells, which is the most common sight of local recurrent excitatory connections (Elston 2007; González-Burgos et al. 2019). We show that NMDA at local connections are crucial for sustained activity to emerge. While much work has contrasted the response of V1 neurons during stimulus presentation to dlPFC neuron activity during the delay period of working memory tasks, some neurons in dlPFC also react transiently to a stimulus, although with a delayed onset. The transient response of these 'Cue' neurons in dlPFC is largely abolished by AMPA receptor antagonists (Wang et al. 2013; Yang et al. 2018). This is consistent with feedforward excitation of 'Cue' neurons in dlPFC being mediated by AMPA receptors. Persistent activity may be more reliant on NMDA receptors. However, NMDA receptors cannot act on their own in this function, as excitation via AMPA or nicotinic receptors is also required to depolarize the cell and remove the magnesium block to engage NMDA receptors (Van Vugt et al. 2020; Yang et al. 2013). NMDA receptors in V1 are largely responsible for modulating late responses via inter-areal feedback connections, rather than the initial feedforward pass (Self et al. 2012). As all the inter-areal cortico-cortical connections that V1 receives are essentially feedback, this may account for the larger NMDA/AMPA ratio in V1, even if relatively few local recurrent connections are via NMDA receptors. Our simulations show that a model that closely matches the experimentally-observed NMDA and AMPA distributions could produce robust stimulus propagation and ignition of fronto-parietal activity. The receptor-based model produced ignition-like dynamics without the need for adjustment to any parameters in the model. This is because the requirement for a decreasing NMDA/AMPA ratio can be naturally fulfilled with feedforward excitation mediated by AMPA receptors, and feedback excitation mediated by NMDA. Therefore, far from being incompatible, we suggest that the same asymmetry in feedforward and feedback connections produces both the decreasing anatomical NMDA/AMPA gradient, and the dynamic physiological activity that produces rapid responses to stimuli and sustained firing in prefrontal cortex.

The brain mechanism of false alarms

While subcortical structures such as the thalamus and neuromodulatory centres are known to play an important role in conscious access (Schiff et al. 2007; Suzuki and Larkum 2020), here for simplicity their contribution was summarized by a single scalar vigilance signal (similar to Dehaene et al. 2003; Dehaene and Changeux 2005). A sufficiently high level of this vigilance signal was critical to enable the model to accurately detect incoming stimuli. Recent evidence in humans suggests that several subcortical structures show increased activation when stimuli are perceived, but may not contain content-specific information (Levinson et al. 2021). This is also consistent with our recent work, which suggests that content-independent dopamine release in the cortex enables the engagement of distributed, sustained activity in working memory (Froudist-Walsh et al. 2021a). Facilitation of sustained fronto-parietal activity by subcortical nuclei could be a shared network mechanism for working memory encoding and conscious perception. Increased tonic dopamine levels, albeit in the mouse striatum, are seen before false alarms (Schmack et al. 2021). We found that a high vigilance signal led to an increase in the pre-stimulus cortical firing rate, and an increase in the false alarm rate. Increased pre-stimulus activity is also predictive of false alarms in humans and non-human primates (Podvalny et al. 2019; Van Vugt et al. 2018). A subcortical vigilance signal may also signify a contribution of these subcortical structures to the level, rather than the contents of consciousness (Solt et al. 2014). This suggests that subcortical activity may not be sufficient to encode the specific content of consciousness, but may be necessary for any conscious state to exist. The necessary brain areas for consciousness are still under debate. In the future, simulated lesions to different brain areas and networks in the model could create testable predictions for experiments on consciousness. In principle, small cortical lesions to peripheral areas (outside of primary sensory cortex) in our model may result in little-to-no change in the ability of the model to produce ignition, while larger lesions, particularly to core regions of the fronto-parietal network (Mejias and Wang 2022) or regions with a high density of neuromodulatory receptors (Froudist-Walsh et al. 2021a) could impair the ability of the whole network to sustain the bistable dynamics associated with conscious access.

Therefore, going forward, our model can provide a computational platform to connect research on the level and content of consciousness. Among open questions are: what is the nature of neural activity underlying false alarm, is it really a vigilance signal or prior expectation of a specific upcoming stimulus (Podvalny et al. 2019; Powers et al. 2017; Schmack et al. 2021)? Is such a signal localized in specific brain regions; if so, where? What might be a neural circuit mechanism for combining a prior expectation signal with an actual stimulus to generate perception, as proposed by predictive coding (Bastos et al. 2012)?

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Integration of the model in the consciousness literature

The present work proposes a neural mechanism for conscious access, the cognitive function that lets a stimulus enter in the current stream of consciousness (James 1890) and makes it reportable, verbally or non-verbally (Baars 2005; Dehaene et al. 1998). The other cognitive functions associated with consciousness, such as metacognition, self-awareness, or any form of attention, are not addressed. Two major current theories of consciousness, among many (Seth 2007), are Global Neuronal Workspace theory (Dehaene et al. 2003) and Integrated Information Theory (IIT) (Tononi 2004). Our model fits in the GNW literature, as it possesses the major characteristics of the Global Workspace, namely independent sensory modules competing to pass their information to a widely distributed set of areas that broadcast the information to vast parts of the cortex. It differs from previous computational models of the Global Neuronal Workspace in that it is built explicitly on mesoscopic connectome data, and (for Figure 8) incorporates anatomical gradients of glutamatergic receptors. The core of associative areas responsible for ignition in the model, heterogeneously connected by the FLN matrix, looks similar to the specialized majority network taken as an example of a high Phi complex (Oizumi et al. 2014) as stated by the IIT for the origin of Phenomenological Consciousness (Tononi 2004). However the location of the cortical areas, predominantly in frontal and parietal cortex seems to fit more precisely with GNW theory than IIT, which attributes conscious perception primarily to a posterior cortical "hot zone" (Koch et al. 2016). Further anatomically-constrained large-scale modelling, or analysis of our model, could make explicit the areas of agreement and disagreement between GNW and IIT. Future work addressing aspects of metacognition, attention or predictive processing could aim to further bridge GNW with other prominent theories of consciousness (Fleming and Lau 2014; Graziano and Webb 2015; Lau and Rosenthal 2011; Seth and Hohwy 2021; Shea and Frith 2019), and provide much-needed testable predictions about behaviour and neural dynamics to distinguish between such theories (Demertzi et al. 2019; Melloni et al. 2021; Yaron et al. 2021).

Conclusions

We built a connectome-based dynamical model of primate cortex, that successfully accounted for salient results on the spatiotemporal activity and behavior of primates performing tasks designed to assess conscious access. Our model predicts that feedforward excitatory connections should be dominated by AMPA receptors for rapid propagation of stimulus-related activity, while NMDA receptors in local recurrent connections and feedback projections are required for the ignition and sustained activity that accompanies conscious access. Our model reconciles seemingly contradictory anatomical and physiological data on the relative proportion of AMPA and NMDA receptors along the cortical hierarchy, and takes a step towards a cross-level (bridging network, cellular and synaptic mechanisms) theory of consciousness.

Acknowledgements

This work was funded by the NIH grant R01MH062349, ONR grant N00014, NSF NeuroNex grant 2015276, ONR grant N00014, James Simons Foundation grant 543057SPI, the Swartz Foundation to XJW; NIH/NSF CRCNS grant R01MH122024 to XJW and NPG. SD was supported by Collège de France, INSERM and CEA. NPG, LR and MN were supported by the Federal Ministry of Education and Research (BMBF) under project number 01GQ1902. We thank Guillermo González-Burgos for comments on an earlier draft of the manuscript.

Materials and Methods

Model overview

We developed a connectome-based dynamical model of the macaque cortex to investigate the synaptic and network mechanisms underlying the ignition of distributed neural activity that accompanies conscious perception. We simulated local cortical circuits at each of 40 cortical areas and set the existence and strength of directed connections between areas using retrograde tract-tracing data. Cortical areas differed based on their inter-areal connectivity and dendritic spine count on pyramidal cells. As a starting point, we adapted a recently developed model of distributed working memory in 30 cortical areas (Mejias and Wang 2022).

Retrograde tract-tracing data

The inter-areal connectivity data in this paper was acquired by Henry Kennedy and colleagues as part of an ongoing effort to map the cortical connectome of the macaque using retrograde tract-tracing (Markov et al. 2014, 2012; Markov

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

et al. 2013; Mejias et al. 2016). Here we use the directed, weighted connectivity data between 40 cortical areas, which was recently released (Froudist-Walsh et al. 2021a).

A few details of how the connectivity data was collected and processed will help the reader understand the connectome-based dynamical model. For each target area, a retrograde tracer was injected into the cortex. The tracer was taken up in the axon terminals in this area, and retrogradely transported to the cell bodies of neurons that projected to the target. The cortical areas (l) that send axons to the target area (k) are called source areas. For a given injection, the total number of marked cell bodies in the cortex outside of the injected (target) area was counted as a labeled neuron. The number of labeled neurons (LN) within a source cortical area was then divided by the number of labeled neurons in the whole cortex (excluding the target area), to give a fraction of labeled neurons (FLN). The FLN was averaged across all injections in a given target area. For this calculation, we include all cortical areas ($n^{areas} = 91$) defined in the Lyon atlas (Markov et al. 2012).

$$FLN_{[k,l]} = \frac{LN_{[k,l]}}{\sum_{l=1}^{n^{areas}} LN_{[k,l]}} \tag{1}$$

Note that there are 91 cortical areas in the Lyon atlas, and currently 40 areas have been injected with retrograde tracers. This gives the connection strength from all 91 areas to the 40 injected areas, and the full bidirectional connectivity of a subgraph of 40 areas. We use this 40-area subgraph as an anatomical basis for the dynamical model.

In addition, for each inter-areal connection we defined the supragranular labeled neurons (SLN) as the fraction of neurons in the source area whose cell bodies were in the superficial (aka supragranular) layers.

$$SLN_{[k,l]} = \frac{LN^{supra}_{[k,l]}}{LN^{supra}_{[k,l]} + LN^{infra}_{[k,l]}} \tag{2}$$

The subiculum (SUB) and piriform cortex (PIR) have a qualitatively different laminar structure to the neocortical areas, and therefore supra- and infra-laminar connections (and thus the SLN) from these areas are undefined. We removed all connections from these areas from the following calculations ($n^{areas,SLN} = 89$). These connectivity data are available on the core-nets website (register, click the "Download" button, and select the data associated with Froudist-Walsh et al. 2021a).

Dendritic spine data

The spine count data were taken from a series of studies by Elston and colleagues (Elston 2007) and mapped onto the Yerkes19 cortical surface (Donahue et al. 2016), as described in (Froudist-Walsh et al. 2021a,b). Locations on the Yerkes19 cortical surface are represented by 32,492 vertices. The spine count data was obtained by Elston and colleagues from 27 injection sites across the cortex. For each injection site we estimated the number of vertices overlapping with each area in the Lyon atlas. If a cortical area contained only one injection site, the mean spine count from pyramidal cells in that site was taken as the spine count for the area. If a cortical area contained multiple injection sites, we performed a weighted average of the spine counts, according to the number of vertices of overlap. In this way we estimated the spine counts on pyramidal cells in 24 of the 40 injected regions in the Lyon atlas. Based on the strong positive correlation between spine count and cortical hierarchy (r = 0.61, p = 0.001), and following previous work (Chaudhuri et al. 2015; Froudist-Walsh et al. 2021a; Mejias and Wang 2022), we inferred the spine count for the remaining regions based on the hierarchy using linear regression.

Local cortical circuit architecture

In each cortical area we simulated a local circuit, with two interacting excitatory populations (E_1 and E_2), and one population of inhibitory (I) neurons. This is based on a mean-field reduction of a spiking neural network model of cortex (Wang 2002; Wong and Wang 2006).

Description of dynamical variables

The neural populations interact via synapses that contain NMDA, AMPA and GABA receptors. Each receptor has its own dynamics, governed by the following equations.

The synaptic variables are updated as follows (Wang 1999; Wong and Wang 2006)

$$\frac{ds^{NMDA}}{dt} = -\frac{s^{NMDA}}{\tau^{NMDA}} + (1 - s^{NMDA})\gamma_{NMDA} r_E \tag{3}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

$$\frac{ds^{AMPA}}{dt} = -\frac{s^{AMPA}}{\tau^{AMPA}} + (1 - s^{AMPA})\gamma_{AMPA} r_E \tag{4}$$

$$\frac{ds^{GABA}}{dt} = -\frac{s^{GABA}}{\tau^{GABA}} + (1 - s^{GABA})\gamma_{GABA} r_I \tag{5}$$

where s is the fraction of open receptor channels, τ is the time constant of decay of that receptor and γ_{NMDA}, γ_{AMPA} and γ_{GABA} are constants. r_E and r_I are the firing rates of the presynaptic excitatory and inhibitory cells that stimulate the NMDA, AMPA and GABA receptors, calculated below.

NMDA/AMPA ratio

We explored the effects of different NMDA/(NMDA+AMPA) fractions, κ, at local and long-range feedforward and feedback connections. The values used for the main simulations, unless otherwise stated, are in Table 2.

Modulation of excitatory connections by dendritic spines

Approximately 90% of excitatory synapses on neocortical pyramidal cells are on dendritic spines (Nimchinsky et al. 2002). On this basis, we modulate the strength of excitatory connections according to the dendritic spine count.

$$\chi_{[k]} = \frac{\chi^{raw}_{[k]} - \chi^{raw}_{min}}{\chi^{raw}_{max} - \chi^{raw}_{min}}$$

for all cortical areas $[k]$. $\chi^{raw}_{[k]}$ is the spine count for area k, and χ^{raw}_{min} and χ^{raw}_{max} are the minimum and maximum spine counts observed in the data. $\chi_{[k]}$ is therefore the spine count of area k rescaled to lie in the $[0, 1]$ range.

We then apply the gradient of excitation as follows.

$$z_{E,[k]} = z^{min} + \chi_{[k]}(1 - z^{min}) \tag{6}$$

where z^{min} sets the lower bound for the modulation of excitatory connections by the spine count, χ. $z_{E,[k]}$ therefore defines how spine count modulates excitatory connections in area k.

Description of local currents

The local NMDA current onto each population $Ei \in \{E_1, E_2\}$ in area $[k]$ is calculated as follows

$$I^{NMDA,local}_{Ei,[k]} = z_{E,[k]} \kappa^{local} G^{NMDA,loc}_{E,E} s^{NMDA}_{Ei,[k]} \tag{7}$$

Where $z_{E,[k]}$ is the dendritic spine count, κ^{local} the NMDA receptor fraction of the postsynaptic population, $G^{NMDA,loc}_{E,E}$ the local NMDA coupling from the population to itself.

Local connections tend to target the perisomatic area (soma and proximal dendrites) of pyramidal cells (Kalisman et al. 2005; Markram et al. 1997; Petreanu et al. 2009). The soma and proximal dendrites act as a single functional compartment that is separate from a distal dendritic compartment (Yuste et al. 1994). As our dendritic function F (described below) models this distal dendritic compartment, we do not pass local excitatory connections through F.

Similarly local excitatory connections via the AMPA receptors are scaled by the AMPA receptor fraction $1 - \kappa^{local}$, the dendritic spine count $z_{E,[k]}$, and $G^{AMPA,loc}_{E,E}$ the local AMPA coupling from the population to itself.

$$I^{AMPA,local}_{Ei,[k]} = z_{E,[k]}(1 - \kappa^{local}) G^{AMPA,loc}_{E,E} s^{AMPA}_{Ei,[k]} \tag{8}$$

Local inhibitory connections are not directly modulated by the dendritic spine count (as spines indicate excitatory synapses on pyramidal cells, Nimchinsky et al. 2002).

$$I^{GABA}_{Ei,[k]} = G^{GABA}_{E,I} s^{GABA}_{[k]} \tag{9}$$

Where $G^{GABA}_{E,I}$ is the connection strength from the inhibitory pool to the excitatory pools.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

In order to keep the spontaneous activity level similar across brain areas, the local NMDA input to the I population increases with the spine count, and is defined as followed (Mejias and Wang 2022)

$$I_{I,[k]}^{NMDA,local} = z_{I,[k]} G_{I,E}^{NMDA,loc} \sum_{Ei \epsilon \{E_1,E_2\}} s_{Ei,[k]}^{NMDA} \tag{10}$$

with

$$z_{I,[k]} = z_I^{min} + \chi_{[k]}(1 - z_I^{min}) \tag{11}$$

There is no local AMPA current targeting the inhibitory population.

Description of noise, background and vigilance currents

Noise is modeled as an Ornstein-Uhlenbeck process, separately for each population.

$$\tau^{AMPA} \frac{dI_i^{noise}(t)}{dt} = -I_i^{noise}(t) + \eta(t)\sqrt{\tau^{AMPA}\sigma_{i,noise}^2} \tag{12}$$

where $\sigma_{i,noise}$ is the standard deviation of the noise and η is Gaussian white noise with zero mean and unit variance.

A constant background current I_i^{bg} was also added to each population (Table 2). This represents input from brain areas that are not explicitly modeled.

In addition, we examined the effect of an extra, weak excitatory current, I^{vig}, to each unit in associative areas (top 75% of areas ranked according to the hierarchy), which simulated the effect of vigilance on the model (Dehaene et al. 2003; Dehaene and Changeux 2005).

As a simplification, each of these currents targets the perisomatic compartment (i.e, it is not passed through the distal dendritic function F).

Large-scale connectivity structure

In the model, cortical areas are connected using connectivity strengths derived from the retrograde tract-tracing data. The long-range connectivity matrices are built from the FLN matrix. However, as noted in (Markov et al. 2014, 2012; Mejias et al. 2016), the FLN matrix spans 5 orders of magnitude. The relationship between anatomical and physiological connectivity strengths is not clear, but if we were to use the raw FLN values in the large-scale model, many of the weaker connections would become irrelevant. To deal with this, we rescale the FLN matrix in order to increase the influence of smaller connections while maintaining the topological structure (Mejias et al. 2016; Mejias and Wang 2022). We also divide by the total number of connections into each area, so the the sum of connections into each area in the model equals one, as in the original definition of FLN.

$$w_{[k,l]} = \frac{FLN_{[k,l]}^{b_1}}{\sum_{l=1}^{n^{sub}} FLN_{[k,l]}^{b_1}} \tag{13}$$

Here we restrict calculations to the injected cortical areas i, j, which allows us to simulate the complete bidirectional connectivity structure within the subgraph ($n^{sub} = 40$). We use the same parameter value $b1$ as in (Mejias et al. 2016; Mejias and Wang 2022) (Table 2) to construct our inter-areal connectivity matrix W .

Calculation of long-range currents

Excitatory cells in different cortical areas with the same receptive fields are more likely to be functionally connected (Zandvakili and Kohn 2015). This is reflected in our model as follows. In the source area, there are two excitatory populations, 1 and 2, each sensitive to a particular feature of a visual stimulus (such as a location in the visual field). Likewise in the target area, there are two populations 1 and 2, sensitive to the same visual features. We assume that the output of population 1 in the source area goes to population 1 in the target area, and the output of population 2 in the source area goes to population 2 in the target area.

The total long-range connections mediated by the NMDA receptors on the excitatory population E_di in area $[k]$ is calculated as follows:

$$I_{E_i[k]}^{NMDA,LR} = F\left(G_E^{NMDA} z_{E,[k]}\left(\sum_{l=1}^{n^{sub}} w_{[k,l]}(SLN_{[k,l]}\kappa_N^{FF}\rho_E^{FF} + (1 - SLN_{[k,l]})\kappa_N^{FB}\rho_E^{FB}) s_{Ei[l]}^{NMDA}\right)\right) \tag{14}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The amount of long-range current onto the excitatory population E_i in area k that comes via NMDA receptors depends on the fraction of open NMDA receptor channels and the anatomical strength of inter-areal connections from all source areas l that target area k. This is scaled by the global excitatory NMDA coupling strength and the amount of dendritic spines per pyramidal cell (i.e. excitatory synapses) in area k. We separate the feedforward from the feedback projections as they may be mediated by different receptor types and target different cell types. All of the excitatory long-range inputs are filtered by the dendrites of the excitatory cells in the target area. Equations (16 - 18) can be understood similarly.

To be precise, G_E^{NMDA} is the global coupling for NMDA-mediated inter-areal connections, $z_{E,[k]}$ is the dendritic spine count for area k (as defined above), $w_{[k,l]}$ is the anatomical connection strength from area l to area k, $SLN_{[k,l]}$ is the feedforward fraction of the connection from area l to area k (as above), κ_N^{FF} is the fraction of feedforward excitation that is mediated by NMDA receptors, ρ_E^{FF} is the fraction of feedforward inter-areal connections targeting excitatory cells and $s_{[l]}^{NMDA}$ is the NMDA synaptic gating variable from the corresponding excitatory population in source area l. Similarly, κ_N^{FB} is the fraction of feedback excitation mediated by NMDA receptors, and ρ_E^{FB} is the fraction of feedback inter-areal connections targeting excitatory cells. $i = 1, 2$ denotes the excitatory population.

Long-distance connections tend to target more distal parts of the dendrites (Petreanu et al. 2009), which act as a functionally separate compartment from the perisomatic area (Yuste et al. 1994). For this reason, we pass the long-distance connections through the dendritic function F before they reach the soma.

The function F is a simplification of a dendritic function used in previous local and large-scale models (Froudist-Walsh et al. 2021a; Yang et al. 2016). It helps the network stabilize, and avoid epileptic behaviours.

$$F(X) = \left\{ \begin{array}{ll} 0pA & \text{for } X \leq 0pA \\ 300pA, & \text{for } X \geq 300pA \\ X, & \text{otherwise} \end{array} \right\} \tag{15}$$

Similarly, the total long-range connections of the excitatory population in area $[k]$ mediated by AMPA receptors is calculated as follows:

$$I_{Ei[k]}^{AMPA,LR} = F\left(G_E^{AMPA} z_{E,[k]} \left(\sum_{l=1}^{n^{sub}} w_{[k,l]}(SLN_{[k,l]}(1-\kappa_N^{FF})\rho_E^{FF} + (1-SLN_{[k,l]})(1-\kappa_N^{FB})\rho_E^{FB}) s_{Ei[l]}^{AMPA}\right)\right) \tag{16}$$

where $(1-\kappa_N^{FF})$ and $(1-\kappa_N^{FB})$ are the fraction of feedforward and feedback inter-areal connections mediated by $AMPA$ receptors. This is scaled by the global excitatory AMPA coupling strength (G_E^{AMPA}).

The total long-range connections targeting inhibitory population in area $[k]$ that are mediated by NMDA receptors is calculated as follows:

$$I_{I[k]}^{NMDA,LR} = G_I^{NMDA} z_{I,[k]} \left(\sum_{l=1}^{n^{sub}} w_{[k,l]}(SLN_{[k,l]}\kappa_N^{FF}(1-\rho_E^{FF}) + (1-SLN_{[k,l]})\kappa_N^{FB}(1-\rho_E^{FB})) s_{[l]}^{NMDA}\right) \tag{17}$$

where $(1-\rho_E^{FF})$ and $(1-\rho_E^{FB})$ are the fraction of feedforward and feedback inter-areal connections targeting inhibitory cell populations. We assume different effective strengths for long-range connections targeting excitatory and inhibitory pools, captured by G_I^{NMDA} and G_I^{AMPA}. Although cortical inhibitory interneurons do not contain dendritic spines, we assume that the level of excitation onto inhibitory scales similarly with the spine count. This has been shown to be an effective way of maintaining spontaneous activity levels across areas (Mejias and Wang 2022).

The total long-range connections targeting the inhibitory population in area $[k]$ that are mediated by AMPA receptors is calculated as:

$$I_{I[k]}^{AMPA,LR} = G_I^{AMPA} z_{I,[k]} \left(\sum_{l=1}^{n^{sub}} w_{[k,l]}(SLN_{[k,l]}(1-\kappa_N^{FF})(1-\rho_E^{FF}) + (1-SLN_{[k,l]})(1-\kappa_N^{FB})(1-\rho_E^{FB})) s_{[l]}^{AMPA}\right) \tag{18}$$

with all variables as described above.

Application of external stimuli for tasks

In all simulations, the stimulus is applied for 50ms to excitatory population 1 in area V1. In the brain, visual input from LGN to V1 targets layer IV excitatory interneurons, which then excite the perisomatic areas of layer III pyramidal

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

cells. For this reason we model external input to the perisomatic compartment of excitatory neurons in V1 (i.e., it is not passed through the dendritic function F). In all equations, the stimulus is designated by the term I^{stim}.

Total current in large-scale model

The total current for each neural population i in each area k equals the sum of all long-range, local and external inputs, and intrinsic currents,

$$I_{i[k]}^{total} = I_{i[k]}^{NMDA,LR} + I_{i[k]}^{AMPA,LR} + I_{i[k]}^{NMDA,local} + I_{i[k]}^{AMPA,local} + I_{i[k]}^{GABA,local} + I_{i[k]}^{noise} + I_{i}^{bg} + I_{i[k]}^{stim} + I_{i[k]}^{vig} \quad (19)$$

where $i\epsilon\{E_1, E_2, I\}$, (E1: excitatory population 1; E2: excitatory population 2; I: inhibitory population).

Description of f-I curves

The f-I (current to frequency) curve of the excitatory population is

$$f(I_E^{total}) = \frac{aI_E^{total} - b}{1 - e^{-d(aI_E^{total} - b)}} \quad (20)$$

where I_E^{total} is the total input to the population, a is a gain factor, d determines the shape of $f(I_E^{total})$, such that if d is large, $f(I_E^{total})$ acts like a threshold-linear function, with threshold b (Abbott and Chance 2005).

The f-I curves for the inhibitory neuron populations are modeled using a threshold-linear function

$$f(I_I^{total}) = \left\{ \begin{array}{ll} \beta_i(I_I^{total} - I_{th}) & \text{for } I_I^{total} \geq I_{th} \\ 0, & \text{otherwise} \end{array} \right\} \quad (21)$$

where I_I^{total} is the total input to the population, β_i is the gain and I_{th} is the threshold.

See Table 2 for parameter values.

The firing rates are updated as follows

$$\tau^r \frac{dr}{dt} = -r + f(I^{total}) \quad (22)$$

for all cell types.

Classification of model dynamics

Model 0: Null Model
In this model, the external input has no effect on the activity. Irrespective of whether a stimulus was presented or not, and irrespective of its strength, activity follows a Gaussian distribution centered on μ with a standard deviation of σ.

$$p_0(activity = x | I_{stim} = I) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (23)$$

There are two free parameters: μ and σ

Model 1: Unimodal Non-Linear
In this model, the activity evoked for each stimulus strength I follows a Gaussian distribution centered on a mean μ - following a sigmoid function of I - and a standard deviation σ, following a linear function of the mean μ. For this model, the probability to reach activity level x for an input I is given by:

$$p_1(activity = x | I_{stim} = I) = \frac{1}{\sigma(I)\sqrt{2\pi}} e^{-\frac{(x-\mu(I))^2}{2\sigma(I)^2}} \quad (24)$$

with

$$\mu(I) = \frac{\mu_{max} - \mu_{min}}{1 + e^{-k(I-I_0)}} + \mu_{min} \quad (25)$$

and

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

$$\sigma(I) = \sigma_{slope}\mu(I) + \sigma_{intercept} \tag{26}$$

There are six free parameters: μ_{max}, μ_{min}, k, I_0, σ_{slope} and $\sigma_{intercept}$

Model 2: bifurcation

In this model, the activity evoked for each stimulus strength I has a probability $\beta(I)$ to belong to a high state (Gaussian distribution centered on μ_{high} of variance σ_{high}), and a probability $(1 - \beta(I))$ to belong to a low state (Gaussian distribution centered on μ_{low} of variance σ_{low}, the baseline activity observed in the absence of stimulation). For this model, the probability to reach activity level x for an input I is given by:

$$\begin{aligned} p_2(activity = x|I_{stim} = I) =& \beta(I)\left(\frac{1}{\sigma_{high}\sqrt{2\pi}}e^{-\frac{(x-\mu_{high})^2}{2\sigma_{high}^2}}\right) \\ &+ (1-\beta(I))\left(\frac{1}{\sigma_{low}\sqrt{2\pi}}e^{-\frac{(x-\mu_{low})^2}{2\sigma_{low}^2}}\right) \end{aligned} \tag{27}$$

with

$$\beta(I) = \frac{1}{1 + e^{-k(I-I_0)}} \tag{28}$$

There are six free parameters: μ_{high}, μ_{low}, σ_{high}, σ_{low}, k and I_0

Bayesian model comparison

We compared the performance of the different models simulating 100 simulations for four different input current values (800 trials in total). The activity was sampled every 40ms and the activity was averaged over all 40 areas. The best parameters for each model were estimated by maximum likelihood, i.e., by finding the parameters maximizing the product of the likelihoods across the different trials (or, equivalently, maximizing the sum of the log likelihoods). The parameter search was achieved using the scipy.optimize function. In order to compare our different models, we used the following formula, where $P(M_i|x(t))$ is the posterior probability of the model $i \in \{0, 1, 2\}$ at the time step t (Lebarbier and Mary-Huard 2004).

$$P(M_i|x(t)) = \frac{e^{\frac{BIC_i(t) - MIN_{i\in\{0,1,2\}}(BIC_i(t))}{2}}}{\sum_{i\in\{0,1,2\}} e^{\frac{BIC_i(t) - MIN_{i\in\{0,1,2\}}(BIC_i(t))}{2}}} \tag{29}$$

where $BIC_i(t)$ correspond to the Bayesian Information Criterion of model i at time step t for the best parameter set of this model.

Temporal generalization of stimulus detection decoders

To decode the trial outcome from instantaneous trial activity patterns, we first separated the data from 400 trials into a training set (300 trials) and a test set (100 trials). All trials received a near-threshold (50% detection rate) stimulus input to population E1 of area V1. The combined training and test set contained 200 hit and 200 miss trials, and these were randomly shuffled and allocated to the training and test sets.

As activity in region 9/46d was used to read out the trial outcome, we trained the classifier on activity in all other areas. Trials were considered a 'Hit' if the mean activity in area 9/46d in the last 500ms before the vigilance signal was removed was greater than 15Hz, and a 'Miss' otherwise. We trained each support vector classifier using scikit-learn in Python and standard parameter settings (Pedregosa et al. 2011). A separate classifier was trained for each timepoint in the training data. We then used each of these classifiers to predict the trial outcome based on activity at each time point in a separate test set. Finally, we compared these predictions to the actual trial outcome (defined according to the late sustained activity in 9/46d).

To estimate whether the coding pattern is similar between times t and t', we can train a classifier at time t (across trials) cand test it at time t'. When applied across all pairs of timepoints, this leads to a square $T \times T$ temporal generalization matrix, where T is the number of timepoints (King and Dehaene 2014; King et al. 2016; Meyers et al. 2008).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We assessed the strength of correlation between decoder coefficients (for the decoder trained at each timepoint) and the cortical hierarchy using Pearson correlations (Fig 3E). We conservatively judged the correlation at a particular timepoint to be significant only if the p value was less than 0.001 for all timepoints within a 10ms period centered on the timepoint.

In-vitro receptor autoradiography

Quantitative in-vitro receptor autoradiography was applied to determine the densities of NMDA and receptors in cytoarchitectonically identified cortical areas of the macaque monkey brain (Impieri et al. 2019; Niu et al. 2020, 2021; Rapan et al. 2021). We provide a short description of the method to help the reader understand how receptor data serves to enrich the connectome-based dynamical model. Brain tissue was shock frozen at -40°C in isopentane, hemispheres serially sectioned in the coronal plane at 20µm by means of a cryomicrotome, and sections thaw mounted onto glass slides. Alternating sections were processed for the visualization of cell bodies (Merker 1983) or of receptor densities with previously published protocols (Palomero-Gallagher and Zilles 2018; Table 1). In short, receptor incubation protocols consisted of a preincubation to rehydrate sections and remove endogenous ligands, a main incubation, and a washing step to stop the binding process and remove surplus ligand and buffer salts. The main incubation encompassed parallel experiments to identify the total and non-specific binding of each ligand, whereby sections were incubated with the radiolabeled ligand alone or with the radiolabeled ligand in conjunction with a non-labelled displacer, respectively. Radiolabelled sections were co-exposed with plastic standards with known concentrations of radioactivity against tritium (3H) sensitive films, and the ensuing autoradiographs digitized for densitometric analysis. Thus, for each receptor type, the mean (averaged over all cortical layers) of the grey values contained in 3-5 sections of each cytoarchitectonically identified area were transformed into a receptor concentration per unit protein (fmol/mg protein) by means of a calibration curve which was computed from the grey value images of the plastic standards (Palomero-Gallagher and Zilles 2018).

Table 1: Incubation protocols

Table 1	AMPA	NMDA
[^{3}H]-Ligand	AMPA (10 nM)	MK-801 (3.3 nM)
Displacer	quisqualate (10 µM)	MK-801 (100 µM)
Incubation buffer	50mM Tris-acetate (pH 7.2) + 100 mM KSCN*	50mM Tris-acetate (pH 7.2) + 50 µM glutamate + 30 µM glycine* + 50 µM spermidine*
Preincubation	3 x 10 min, 4°C	15 min, 4°C
Main incubation	45 min, 4°C	60 min, 22°C
Final rinsing	1. 4 x 4 sec, 4°C 2. 2 x 2sec in 100/2.5 acetone/glutaraldehyde, 4°C	1. 2 x 5 min, 4°C 2. rinse in distilled water, 22 °C

* substance only included in buffer for the main incubation

Receptor data-based model

For the receptor data-based model, we matched the total NMDA fraction to that seen in the data, adjusting for a constant mean shift between the model and receptor data.

We calculate the overall NMDA fraction $X_{[k],model}$(fraction of NMDA receptors over total number of excitatory receptors) in each area of the model.

$$X_{[k],model} = NMDA_{[k]}/(NMDA_{[k]} + AMPA_{[k]}) \tag{30}$$

where $NMDA_k$ and $AMPA_k$ are the total local and inter-areal connections mediated by each receptor type.

Here

$$NMDA_k = N^{loc}_{[k]E,E} + N^{loc}_{[k]I,E} + N^{lr}_{[k]} \tag{31}$$

$$NMDA_k = N^{loc}_{[k]E,E} + N^{loc}_{[k]I,E} + N^{lr}_{[k]E,E} + N^{lr}_{[k]I,E} \tag{32}$$

With:
$N^{loc}_{[k]E,E}$ the number of NMDA receptors on the excitatory neurons coming from local connections.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

$N^{loc}_{[k]I,E}$ the number of NMDA receptors on the inhibitory neurons coming from local connections.
$N^{lr}_{[k]}$ the total number of NMDA receptors coming from long-range connections.
$N^{lr}_{[k]E,E}$ the number of NMDA receptors on the excitatory neurons coming from long-range connections.
$N^{lr}_{[k]I,E}$ the number of NMDA receptors on the inhibitory neurons coming from long-range connections.

In the model, $G^{NMDA,loc}_{E,E}$ and $G^{AMPA,loc}_{E,E}$ are set as follow:

$$\frac{G^{NMDA,loc}_{E,E}}{g^{NMDA}} = \frac{G^{AMPA,loc}_{E,E}}{g^{AMPA}} = G^{loc}_{E,E} \tag{33}$$

With g^{NMDA} the conductance of one NMDA channel. We also define:

$$\frac{G^{N,loc}_{I,E}}{g^{NMDA}} = G^{loc}_{I,E} \tag{34}$$

This equations can be expanded based on Equations (7), (10), (14) and (17)

$$\begin{aligned} NMDA_k =& z_{E,[k]}\kappa^{loc}G^{loc}_{E,E} \\ &+ z_{I[k]}G^{loc}_{I,E} \\ &+ z_{E,[k]}G^{lr}_{E,E}\sum_{l=1}^{n^{sub}} w_{[k,l]}(SLN_{[k,l]}\kappa^{FF}\rho^{FF} + (1-SLN_{[k,l]})\kappa^{FB}\rho^{FB}) \\ &+ z_{I[k]}G^{lr}_{I,E}\sum_{l=1}^{n^{sub}} w_{[k,l]}(SLN_{[k,l]}\kappa^{FF}(1-\rho^{FF}) + (1-SLN_{[k,l]})\kappa^{FB}(1-\rho^{FB})) \end{aligned} \tag{35}$$

In the reference model, used throughout Figures 1-7, κ^{loc} is the same across all cortical areas.

Similarly:

$$AMPA_k = A^{loc}_{[k]E,E} + A^{lr}_{[k]} \tag{36}$$

$$AMPA_k = A^{loc}_{[k]E,E} + A^{lr}_{[k]E,E} + A^{lr}_{[k]I,E} \tag{37}$$

It is worth noting that their are no local AMPA connections targeting the inhibitory pool.

$$\begin{aligned} AMPA_k =& z_{E,[k]}(1-\kappa^{loc})G^{loc}_{E,E} \\ &+ z_{E,[k]}G^{lr}_{E,E}\sum_{l=1}^{n^{sub}} w_{[k,l]}(SLN_{[k,l]}(1-\kappa^{FF})\rho^{FF} + (1-SLN_{[k,l]})(1-\kappa^{FB})\rho^{FB}) \\ &+ z_{I[k]}G^{lr}_{I,E}\sum_{l=1}^{n^{sub}} w_{[k,l]}(SLN_{[k,l]}(1-\kappa^{FF})(1-\rho^{FF}) + (1-SLN_{[k,l]})(1-\kappa^{FB})(1-\rho^{FB})) \end{aligned} \tag{38}$$

In practice, both $NMDA_k$ and $AMPA_k$ should be doubled (to represent the two excitatory populations), but as this affects all terms in the numerator and denominator, it will not affect the fraction $X_{[k]}$.

We show in Figure 7 how the proportion of feedforward and feedback connections mediated by AMPA and NMDA receptors should lie in a particular range in order to enable rapid 'ignition' of cortical activity. Therefore, for the receptor data-based model, we treat these long-range feedforward and feedback NMDA and AMPA fractions as fixed. We then set the overall NMDA fraction in each area to match the experimentally-measured value $X_{[k],data}$, shifted by a constant term c to account for the a mean shift between the raw receptor data and the reference model used in the rest of the paper. We can then calculate the local NMDA fraction $\kappa^{local}_{[k]}$ in each area required to match the observed NMDA fraction distribution across the cortex.

By reorganising equations 30 - 38, we can compute the local fraction $\kappa^{local}_{[k]}$ as a function of network parameters and real receptor data $X_{[k],data}$. We forced $\kappa^{local}_{[k]}$ to lie between 0 and 1.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 2. Parameters for Numerical Simulations

Parameter	Description	Value
τ^{NMDA}, τ^{AMPA}	E synaptic time constants	60ms,2ms
g^{NMDA}, g^{AMPA}	Channel Conductances	1000pA, 10000pA
τ^{GABA}	I synaptic time constant	5ms
τ^{R}	Firing rate time constant	2ms
γ_{NMDA}, γ_{AMPA}, γ_I	Synaptic rise constants	1.282, 2, 2
κ^{FF}, κ^{FB}, κ^{local}	NMDA fraction	0., 0.8, 0.91
ρ^{FF}, ρ^{FB}	Long-range E/I targets	1., 0.015
z^{min}	Min excitation value	0.6
z_I^{min}	Min excitation value	0.218
σ_{noise}	std. dev. of noise	5pA
I_E^{bg}, I_I^{bg}	Background inputs	329.4pA, 260pA
a, b, d	f-I curve (E cells)	0.135 Hz/pA, 54Hz, 0.308s
β_i, I_{th}	f-I curve (I cells)	153,75Hz/nA, 252Hz
b_1	Rescale FLN	0.3
$G_{E,E}^{N,loc}$	Excitatory NMDA strengths	480pA
$G_{E,E}^{A,loc}$	Excitatory AMPA strengths	4800pA
$G_{I,E}^{N,loc}$	Excitatory NMDA strengths to the Inhib Unit	10pA
$g_{E,I}$, $g_{I,I}$	Inhibitory strengths	-8800pA, -120pA
G_E^{NDMDA}, G_I^{NMDA}	Long-range NMDA strength	1500pA, 10.5pA
G_E^{AMPA}, G_I^{AMPA}	Long-range AMPA strength	15000pA, 105pA
G_0	Local balanced coupling	215pA
I^{stim}	Stimulus strength	250pA

Please note that this is a current-based model, so all synaptic strengths area given in units of pA.

References

Abbott, Larry F. and Frances S. Chance (2005). "Drivers and modulators from push-pull and balanced synaptic input". *Progress in Brain Research*. Vol. 149. Cortical Function: a View from the Thalamus. Elsevier, pp. 147–155.

Ardid, Salva, Xiao-Jing Wang, and Albert Compte (2007). "An Integrated Microcircuit Model of Attentional Processing in the Neocortex". *Journal of Neuroscience* 27.32, pp. 8486–8495.

Arnsten, Amy F. T., Dibyadeep Datta, and Min Wang (2020). "The genie in the bottle-magnified calcium signaling in dorsolateral prefrontal cortex". *Molecular Psychiatry*, pp. 1–17.

Arnsten, Amy F. T., Lu E. Jin, Nao J. Gamo, Brian Ramos, Constantinos D. Paspalas, Yury M. Morozov, Anna Kata, Nigel S. Bamford, Mark F. Yeckel, Leonard K. Kaczmarek, and Lynda El-Hassar (2019). "Role of KCNQ potassium channels in stress-induced deficit of working memory". *Neurobiology of Stress* 11, p. 100187.

Aru, Jaan, Mototaka Suzuki, and Matthew E. Larkum (2020). "Cellular Mechanisms of Conscious Processing". *Trends in Cognitive Sciences* 24.10, pp. 814–825.

Baars, Bernard J. (2005). "Global workspace theory of consciousness: toward a cognitive neuroscience of human experience". *Progress in Brain Research*. Ed. by Steven Laureys. Vol. 150. The Boundaries of Consciousness: Neurobiology and Neuropathology. Elsevier, pp. 45–53.

Baria, Alexis T., Brian Maniscalco, and Biyu J. He (2017). "Initial-state-dependent, robust, transient neural dynamics encode conscious visual perception". *PLOS Computational Biology* 13.11, e1005806.

Barone, Pascal, Alexandre Batardiere, Kenneth Knoblauch, and Henry Kennedy (2000). "Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule". *Journal of Neuroscience* 20.9, pp. 3263–3281.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Bastos, Andre M., W. Martin Usrey, Rick A. Adams, George R. Mangun, Pascal Fries, and Karl J. Friston (2012). "Canonical Microcircuits for Predictive Coding". *Neuron* 76.4, pp. 695–711.

Bellet, Joachim, Marion Gay, Abhilash Dwarakanath, Bechir Jarraya, Timo van Kerkoerle, Stanislas Dehaene, and Theofanis I Panagiotaropoulos (2022). "Decoding rapidly presented visual stimuli from prefrontal ensembles without report nor post-perceptual processing". *Neuroscience of Consciousness* 2022.1, niac005.

Block, Ned (2019). "What Is Wrong with the No-Report Paradigm and How to Fix It". *Trends in Cognitive Sciences* 23.12, pp. 1003–1013.

Boly, M., E. Balteau, C. Schnakers, C. Degueldre, G. Moonen, A. Luxen, C. Phillips, P. Peigneux, P. Maquet, and S. Laureys (2007). "Baseline brain activity fluctuations predict somatosensory perception in humans". *Proceedings of the National Academy of Sciences* 104.29, pp. 12187–12192.

Brette, Romain et al. (2007). "Simulation of networks of spiking neurons: A review of tools and strategies". *Journal of Computational Neuroscience* 23.3, pp. 349–398.

Busch, Niko A., Julien Dubois, and Rufin VanRullen (2009). "The Phase of Ongoing EEG Oscillations Predicts Visual Perception". *Journal of Neuroscience* 29.24, pp. 7869–7876.

Cavanagh, Sean E., John P. Towers, Joni D. Wallis, Laurence T. Hunt, and Steven W. Kennerley (2018). "Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex". *Nature Communications* 9.1, p. 3498.

Chaudhuri, Rishidev, Kenneth Knoblauch, Marie-Alice Gariel, Henry Kennedy, and Xiao-Jing Wang (2015). "A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex". *Neuron* 88.2, pp. 419–431.

Cohen, Michael A., Kevin Ortego, Andrew Kyroudis, and Michael Pitts (2020). "Distinguishing the Neural Correlates of Perceptual Awareness and Postperceptual Processing". *Journal of Neuroscience* 40.25, pp. 4925–4935.

Collins, Christine E., David C. Airey, Nicole A. Young, Duncan B. Leitch, and Jon H. Kaas (2010). "Neuron densities vary across and within cortical areas in primates". *Proceedings of the National Academy of Sciences* 107.36, pp. 15927–15932.

Deco, Gustavo, Kevin Aquino, Aurina Arnatkevičiūtė, Stuart Oldham, Kristina Sabaroedin, Nigel C. Rogasch, Morten L. Kringelbach, and Alex Fornito (2020). "Dynamical consequences of regional heterogeneity in the brain's transcriptional landscape". *bioRxiv*, p. 2020.10.28.359943.

Dehaene, S., M. Kerszberg, and J.-P. Changeux (1998). "A neuronal model of a global workspace in effortful cognitive tasks." *Proceedings of the National Academy of Sciences*.

Dehaene, S., C. Sergent, and J.-P. Changeux (2003). "A neuronal network model linking subjective reports and objective physiological data during conscious perception." *Proceedings of the National Academy of Sciences*.

Dehaene, Stanislas and Jean-Pierre Changeux (2005). "Ongoing Spontaneous Activity Controls Access to Consciousness: A Neuronal Model for Inattentional Blindness". *PLOS Biology* 3.5, e141.

Dehaene, Stanislas, Jean-Pierre Changeux, Lionel Naccache, Jérôme Sackur, and Claire Sergent (2006). "Conscious, preconscious, and subliminal processing: a testable taxonomy". *Trends in Cognitive Sciences* 10.5, pp. 204–211.

Dehaene, Stanislas and Jean-Rémi King (2016). "Decoding the Dynamics of Conscious Perception: The Temporal Generalization Method". *Micro-, meso-and macro-dynamics of the brain*, pp. 85–97.

Dehaene, Stanislas, Lionel Naccache, Laurent Cohen, Denis Le Bihan, Jean-François Mangin, Jean-Baptiste Poline, and Denis Rivière (2001). "Cerebral mechanisms of word masking and unconscious repetition priming". *Nature Neuroscience* 4.7, pp. 752–758.

Del Cul, Antoine, Sylvain Baillet, and Stanislas Dehaene (2007). "Brain dynamics underlying the nonlinear threshold for access to consciousness". *PLOS Biology* 5.10, e260.

Dellert, Torge, Miriam Müller-Bardorff, Insa Schlossmacher, Michael Pitts, David Hofmann, Maximilian Bruchmann, and Thomas Straube (2021). "Dissociating the Neural Correlates of Consciousness and Task Relevance in Face Perception Using Simultaneous EEG-fMRI". *Journal of Neuroscience* 41.37, pp. 7864–7875.

Demertzi, Athena, Enzo Tagliazucchi, Stanislas Dehaene, Gustavo Deco, Pablo Barttfeld, Federico Raimondo, Charlotte Martial, Davinia Fernández-Espejo, Benjamin Rohaut, and H. U. Voss (2019). "Human consciousness is supported by dynamic complex patterns of brain signal coordination". *Science advances* 5.2, eaat7603.

Doerig, Adrien, Aaron Schurger, and Michael H. Herzog (2021). "Hard criteria for empirical theories of consciousness". *Cognitive Neuroscience* 12.2, pp. 41–62.

Donahue, C. J., S. N. Sotiropoulos, S. Jbabdi, M. Hernandez-Fernandez, T. E. Behrens, T. B. Dyrby, T. Coalson, H. Kennedy, K. Knoblauch, D. C. Van Essen, and M. F. Glasser (2016). "Using diffusion tractography to predict cortical connection strength and distance: a quantitative comparison with tracers in the monkey". *J. Neurosci.* 36, pp. 6758–6770.

Elston, Guy N. (2007). "Specialization of the neocortical pyramidal cell during primate evolution". *Evolution of Nervous Systems*. Elsevier, pp. 191–242.

Felleman, Daniel J. and David C. Van Essen (1991). "Distributed hierarchical processing in the primate cerebral cortex." *Cerebral cortex* 1.1, pp. 1–47.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Fleming, Stephen M. and Hakwan C. Lau (2014). "How to measure metacognition". *Frontiers in Human Neuroscience* 8.

Froudist-Walsh, Sean, Daniel P. Bliss, Xingyu Ding, Lucija Rapan, Meiqi Niu, Kenneth Knoblauch, Karl Zilles, Henry Kennedy, Nicola Palomero-Gallagher, and Xiao-Jing Wang (2021a). "A dopamine gradient controls access to distributed working memory in the large-scale monkey cortex". *Neuron* 109.21, 3500–3520.e13.

Froudist-Walsh, Sean, Ting Xu, Meiqi Niu, Lucija Rapan, Karl Zilles, Daniel S. Margulies, Xiao-Jing Wang, and Nicola Palomero-Gallagher (2021b). "Gradients of receptor expression in the macaque cortex". *bioRxiv*, p. 2021.02.22.432173.

Galvin, Veronica C., Sheng Tao Yang, Constantinos D. Paspalas, Yang Yang, Lu E. Jin, Dibyadeep Datta, Yury M. Morozov, Taber C. Lightbourne, Adam S. Lowet, Pasko Rakic, Amy F. T. Arnsten, and Min Wang (2020). "Muscarinic M1 Receptors Modulate Working Memory Performance and Activity via KCNQ Potassium Channels in the Primate Prefrontal Cortex". *Neuron* 106.4, 649–661.e4.

Goldman, Jennifer S., Núria Tort-Colet, Matteo di Volo, Eduarda Susin, Jules Bouté, Melissa Dali, Mallory Carlu, Trang-Anh Nghiem, Tomasz Górski, and Alain Destexhe (2019). "Bridging Single Neuron Dynamics to Global Brain States". *Frontiers in Systems Neuroscience* 13.

Golos, Mathieu, Viktor Jirsa, and Emmanuel Daucé (2015). "Multistability in Large Scale Models of Brain Activity". *PLOS Computational Biology* 11.12, e1004644.

González-Burgos, Guillermo, Takeaki Miyamae, Yosef Krimer, Yelena Gulchina, Diego E. Pafundo, Olga Krimer, Holly Bazmi, Dominique Arion, John F. Enwright, Kenneth N. Fish, and David A. Lewis (2019). "Distinct Properties of Layer 3 Pyramidal Neurons from Prefrontal and Parietal Areas of the Monkey Neocortex". *Journal of Neuroscience* 39.37, pp. 7277–7290.

Goulas, Alexandros, Jean-Pierre Changeux, Konrad Wagstyl, Katrin Amunts, Nicola Palomero-Gallagher, and Claus C. Hilgetag (2021). "The natural axis of transmitter receptor distribution in the human cerebral cortex". *Proceedings of the National Academy of Sciences* 118.3.

Graziano, Michael SA and Taylor W. Webb (2015). "The attention schema theory: a mechanistic account of subjective awareness". *Frontiers in psychology* 6, p. 500.

Green, David Marvin and John A. Swets (1966). *Signal detection theory and psychophysics*. Vol. 1. Wiley New York.

Harrison, Stephenie A. and Frank Tong (2009). "Decoding reveals the contents of visual working memory in early visual areas". *Nature* 458.7238, pp. 632–635.

He, Biyu J. (2018). "Robust, Transient Neural Dynamics during Conscious Perception". *Trends in Cognitive Sciences* 22.7, pp. 563–565.

Huang, Chengcheng, Douglas A. Ruff, Ryan Pyle, Robert Rosenbaum, Marlene R. Cohen, and Brent Doiron (2019). "Circuit Models of Low-Dimensional Shared Variability in Cortical Networks". *Neuron* 101.2, 337–348.e4.

Impieri, Daniele, Karl Zilles, Meiqi Niu, Lucija Rapan, Nicole Schubert, Claudio Galletti, and Nicola Palomero-Gallagher (2019). "Receptor density pattern confirms and enhances the anatomic-functional features of the macaque superior parietal lobule areas". *Brain Structure and Function* 224.8, pp. 2733–2756.

James, W. (1890). *Principles of Psychology*.

Joglekar, Madhura R., Jorge F. Mejias, Guangyu Robert Yang, and Xiao-Jing Wang (2018). "Inter-areal Balanced Amplification Enhances Signal Propagation in a Large-Scale Circuit Model of the Primate Cortex". *Neuron* 98.1, 222–234.e8.

Kalisman, Nir, Gilad Silberberg, and Henry Markram (2005). "The neocortical microcircuit as a tabula rasa". *Proceedings of the National Academy of Sciences* 102.3, pp. 880–885.

Kerkoerle, Timo van, Matthew W. Self, and Pieter R. Roelfsema (2017). "Layer-specificity in the effects of attention and working memory on activity in primary visual cortex". *Nature Communications* 8.1, pp. 1–14.

King, J.-R. and S. Dehaene (2014). "Characterizing the dynamics of mental representations: the temporal generalization method". *Cell Press*.

King, Jean-Rémi, Niccolo Pescetelli, and Stanislas Dehaene (2016). "Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information". *Neuron* 92.5, pp. 1122–1134.

Koch, Christof, Marcello Massimini, Melanie Boly, and Giulio Tononi (2016). "Neural correlates of consciousness: progress and problems". *Nature Reviews Neuroscience* 17.5, pp. 307–321.

Kouider, Sid, Carsten Stahlhut, Sofie V. Gelskov, Leonardo S. Barbosa, Michel Dutat, Vincent de Gardelle, Anne Christophe, Stanislas Dehaene, and Ghislaine Dehaene-Lambertz (2013). "A Neural Marker of Perceptual Consciousness in Infants". *Science* 340.6130, pp. 376–380.

Lafuente, Victor de and Ranulfo Romo (2005). "Neuronal correlates of subjective sensory experience". *Nature Neuroscience* 8.12, pp. 1698–1703.

– (2006). "Neural correlate of subjective sensory experience gradually builds up across cortical areas". *Proceedings of the National Academy of Sciences* 103.39, pp. 14266–14271.

Lamme, Victor A. F. and Pieter R. Roelfsema (2000). "The distinct modes of vision offered by feedforward and recurrent processing". *Trends in Neurosciences* 23.11, pp. 571–579.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Lau, Hakwan and David Rosenthal (2011). "Empirical support for higher-order theories of conscious awareness". *Trends in cognitive sciences* 15.8, pp. 365–373.

Leavitt, Matthew L., Diego Mendoza-Halliday, and Julio C. Martinez-Trujillo (2017). "Sustained Activity Encoding Working Memories: Not Fully Distributed". *Trends in Neurosciences* 40.6, pp. 328–346.

Lebarbier, Emilie and Tristan Mary-Huard (2004). "Le critère BIC: fondements théoriques et interprétation". PhD Thesis. INRIA.

Levinson, Max, Ella Podvalny, Steven H. Baete, and Biyu J. He (2021). "Cortical and subcortical signatures of conscious object recognition". *Nature Communications* 12.1, p. 2930.

Lo, Chung-Chuan, Cheng-Te Wang, and Xiao-Jing Wang (2015). "Speed-accuracy tradeoff by a control signal with balanced excitation and inhibition". *Journal of Neurophysiology* 114.1, pp. 650–661.

Markov, N.T., J. Vezoli, P. Chameau, A. Falchier, and R. Quilodran (2014). "Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex." *The Journal of comparative neurology*.

Markov, N.T. et al. (2012). "A weighted and directed interareal connectivity matrix for macaque cerebral cortex." *Cerebral cortex*.

Markov, Nikola T., Mária Ercsey-Ravasz, David C. Van Essen, Kenneth Knoblauch, Zoltán Toroczkai, and Henry Kennedy (2013). "Cortical high-density counterstream architectures". *Science* 342.6158, p. 1238406.

Markram, Henry, Joachim Lübke, Michael Frotscher, Arnd Roth, and Bert Sakmann (1997). "Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex." *The Journal of Physiology* 500.2, pp. 409–440.

Mashour, George A., Pieter Roelfsema, Jean-Pierre Changeux, and Stanislas Dehaene (2020). "Conscious Processing and the Global Neuronal Workspace Hypothesis". *Neuron* 105.5, pp. 776–798.

Mejias, Jorge F., J. Murray, H. Kennedy, and X.-J. Wang (2016). "Feedforward and feedback frequency-dependent interactions in a large-scale laminar network of the primate cortex." *ScienceAdvances*.

Mejias, Jorge F. and Xiao-Jing Wang (2022). "Mechanisms of distributed working memory in a large-scale network of macaque neocortex". *eLife* 11, e72136.

Melloni, Lucia, Carlos Molina, Marcela Pena, David Torres, Wolf Singer, and Eugenio Rodriguez (2007). "Synchronization of Neural Activity across Cortical Areas Correlates with Conscious Perception". *Journal of Neuroscience* 27.11, pp. 2858–2865.

Melloni, Lucia, Liad Mudrik, Michael Pitts, and Christof Koch (2021). "Making the hard problem of consciousness easier". *Science* 372.6545, pp. 911–912.

Melloni, Lucia, Caspar M. Schwiedrzik, Notger Müller, Eugenio Rodriguez, and Wolf Singer (2011). "Expectations Change the Signatures and Timing of Electrophysiological Correlates of Perceptual Awareness". *Journal of Neuroscience* 31.4, pp. 1386–1396.

Merker, Björn (1983). "Silver staining of cell bodies by means of physical development". *Journal of Neuroscience Methods* 9.3, pp. 235–241.

Meyers, Ethan M., David J. Freedman, Gabriel Kreiman, Earl K. Miller, and Tomaso Poggio (2008). "Dynamic Population Coding of Category Information in Inferior Temporal and Prefrontal Cortex". *Journal of Neurophysiology* 100.3, pp. 1407–1419.

Murray, John D., Alberto Bernacchia, Nicholas A. Roy, Christos Constantinidis, Ranulfo Romo, and Xiao-Jing Wang (2017). "Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex". *Proceedings of the National Academy of Sciences* 114.2, pp. 394–399.

Nachmias, J. (1981). "On the psychometric function for contrast detection." *Vision Research*.

Nimchinsky, Esther A., Bernardo L. Sabatini, and Karel Svoboda (2002). "Structure and function of dendritic spines". *Annual Review of Physiology* 64.1, pp. 313–353.

Niu, Meiqi, Daniele Impieri, Lucija Rapan, Thomas Funck, Nicola Palomero-Gallagher, and Karl Zilles (2020). "Receptor-driven, multimodal mapping of cortical areas in the macaque monkey intraparietal sulcus". *eLife* 9. Ed. by Timothy E Behrens, e55979.

Niu, Meiqi, Lucija Rapan, Thomas Funck, Seán Froudist-Walsh, Ling Zhao, Karl Zilles, and Nicola Palomero-Gallagher (2021). "Organization of the macaque monkey inferior parietal lobule based on multimodal receptor architectonics". *NeuroImage* 231, p. 117843.

Oizumi, Masafumi, Larissa Albantakis, and Giulio Tononi (2014). "From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0". *PLoS computational biology* 10.5, e1003588.

Palomero-Gallagher, Nicola and Karl Zilles (2018). "Chapter 24 - Cyto- and receptor architectonic mapping of the human brain". *Handbook of Clinical Neurology*. Ed. by Ingeborg Huitinga and Maree J. Webster. Vol. 150. Brain Banking. Elsevier, pp. 355–387.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg (2011). "Scikit-learn: Machine learning in Python". *the Journal of machine Learning research* 12, pp. 2825–2830.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Petreanu, Leopoldo, Tianyi Mao, Scott M. Sternson, and Karel Svoboda (2009). "The subcellular organization of neocortical excitatory connections". *Nature* 457.7233, pp. 1142–1145.

Pitts, Michael A., Lydia A. Lutsyshyna, and Steven A. Hillyard (2018). "The relationship between attention and consciousness: an expanded taxonomy and implications for 'no-report' paradigms". *Philosophical Transactions of the Royal Society B: Biological Sciences* 373.1755, p. 20170348.

Pitts, Michael A., Antígona Martínez, and Steven A. Hillyard (2012). "Visual Processing of Contour Patterns under Conditions of Inattentional Blindness". *Journal of Cognitive Neuroscience* 24.2, pp. 287–303.

Pitts, Michael A., Jennifer Padwal, Daniel Fennelly, Antígona Martínez, and Steven A. Hillyard (2014). "Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness". *NeuroImage* 101, pp. 337–350.

Podvalny, Ella, Matthew W. Flounders, Leana E. King, Tom Holroyd, and Biyu J. He (2019). "A dual role of prestimulus spontaneous neural activity in visual object recognition". *Nature Communications* 10.1, p. 3910.

Powers, A. R., C. Mathys, and P. R. Corlett (2017). "Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors". *Science* 357.6351, pp. 596–600.

Rapan, Lucija, Sean Froudist-Walsh, Meiqi Niu, Ting Xu, Thomas Funck, Karl Zilles, and Nicola Palomero-Gallagher (2021). "Multimodal 3D atlas of the macaque monkey motor and premotor cortex". *NeuroImage*, p. 117574.

Rees, Geraint, Gabriel Kreiman, and Christof Koch (2002). "Neural correlates of consciousness in humans". *Nature Reviews Neuroscience* 3.4, pp. 261–270.

Romo, Ranulfo and Román Rossi-Pool (2020). "Turning Touch into Perception". *Neuron* 105.1, pp. 16–33.

Sadaghiani, Sepideh, Guido Hesselmann, and Andreas Kleinschmidt (2009). "Distributed and Antagonistic Contributions of Ongoing Activity Fluctuations to Auditory Stimulus Detection". *Journal of Neuroscience* 29.42, pp. 13410–13417.

Salti, Moti, Simo Monto, Lucie Charles, Jean-Remi King, Lauri Parkkonen, and Stanislas Dehaene (2015). "Distinct cortical codes and temporal dynamics for conscious and unconscious percepts". *eLife* 4. Ed. by Heidi Johansen-Berg, e05652.

Sanz Leon, Paula, Stuart A. Knock, M. Marmaduke Woodman, Lia Domide, Jochen Mersmann, Anthony R. McIntosh, and Viktor Jirsa (2013). "The Virtual Brain: a simulator of primate brain network dynamics". *Frontiers in Neuroinformatics* 0.

Schiff, N. D., J. T. Giacino, K. Kalmar, J. D. Victor, K. Baker, M. Gerber, B. Fritz, B. Eisenberg, J. O'Connor, E. J. Kobylarz, S. Farris, A. Machado, C. McCagg, F. Plum, J. J. Fins, and A. R. Rezai (2007). "Behavioural improvements with thalamic stimulation after severe traumatic brain injury". *Nature* 448.7153, pp. 600–603.

Schlossmacher, Insa, Torge Dellert, Michael Pitts, Maximilian Bruchmann, and Thomas Straube (2020). "Differential Effects of Awareness and Task Relevance on Early and Late ERPs in a No-Report Visual Oddball Paradigm". *Journal of Neuroscience* 40.14, pp. 2906–2913.

Schmack, K., M. Bosc, T. Ott, J. F. Sturgill, and A. Kepecs (2021). "Striatal dopamine mediates hallucination-like perception in mice". *Science* 372.6537, eabf4740.

Schmidt, Maximilian, Rembrandt Bakker, Kelly Shen, Gleb Bezgin, Markus Diesmann, and Sacha Jennifer van Albada (2018). "A multi-scale layer-resolved spiking network model of resting-state dynamics in macaque visual cortical areas". *PLOS Computational Biology* 14.10, e1006359.

Schurger, Aaron, Ioannis Sarigiannidis, Lionel Naccache, Jacobo D. Sitt, and Stanislas Dehaene (2015). "Cortical activity is more stable when sensory stimuli are consciously perceived". *Proceedings of the National Academy of Sciences* 112.16, E2083–E2092.

Self, Matthew W., Roxana N. Kooijmans, Hans Supèr, Victor A. Lamme, and Pieter R. Roelfsema (2012). "Different glutamate receptors convey feedforward and recurrent processing in macaque V1". *Proceedings of the National Academy of Sciences* 109.27, pp. 11031–11036.

Sergent, C., S. Baillet, and S. Dehaene (2005). "Timing of the brain events underlying access to consciousness during the attentional blink." *Nature Neuroscience*.

Sergent, C. and S Dehaene (2003). "Is consciousness a gradual phenomenon?" *Psychological Science*.

Sergent, Claire, Martina Corazzol, Ghislaine Labouret, François Stockart, Mark Wexler, Jean-Rémi King, Florent Meyniel, and Daniel Pressnitzer (2021). "Bifurcation in brain dynamics reveals a signature of conscious processing independent of report". *Nature Communications* 12.1, p. 1149.

Seth, Anil (2007). "Models of consciousness". *Scholarpedia* 2.1, p. 1328.

Seth, Anil K and Jakob Hohwy (2021). "Predictive processing as an empirical theory for consciousness science". *Cognitive Neuroscience* 12.2, pp. 89–90.

Seth, Anil K. and Tim Bayne (2022). "Theories of consciousness". *Nature Reviews Neuroscience*.

Shea, Nicholas and Chris D. Frith (2019). "The Global Workspace Needs Metacognition". *Trends in Cognitive Sciences* 23.7, pp. 560–571.

Signorelli, Camilo Miguel, Joanna Szczotka, and Robert Prentner (2021). *Explanatory profiles of models of consciousness- towards a systematic classification*. Tech. rep. PsyArXiv.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Solt, Ken, Christa J. Van Dort, Jessica J. Chemali, Norman E. Taylor, Jonathan D. Kenny, and Emery N. Brown (2014). "Electrical Stimulation of the Ventral Tegmental Area Induces Reanimation from General Anesthesia". *Anesthesiology* 121.2, pp. 311–319.

Strogatz, Steven H. (2018). *Nonlinear dynamics and chaos*. CRC press.

Supèr, Hans, Henk Spekreijse, and Victor A. F. Lamme (2001). "A Neural Correlate of Working Memory in the Monkey Primary Visual Cortex". *Science* 293.5527, pp. 120–124.

Suzuki, Mototaka and Matthew E. Larkum (2020). "General Anesthesia Decouples Cortical Pyramidal Neurons". *Cell* 180.4, 666–676.e13.

Takahashi, Naoya, Christian Ebner, Johanna Sigl-Glöckner, Sara Moberg, Svenja Nierwetberg, and Matthew E. Larkum (2020). "Active dendritic currents gate descending cortical outputs in perception". *Nature Neuroscience* 23.10, pp. 1277–1285.

Thorpe, Simon J. and Michèle Fabre-Thorpe (2001). "Seeking Categories in the Brain". *Science* 291.5502, pp. 260–263.

Tononi, Giulio (2004). "An information integration theory of consciousness". *BMC neuroscience* 5.1, pp. 1–22.

Trübutschek, Darinka, Sébastien Marti, Andrés Ojeda, Jean-Rémi King, Yuanyuan Mi, Misha Tsodyks, and Stanislas Dehaene (2017). "A theory of working memory without consciousness or sustained activity". *eLife* 6. Ed. by Tatiana Pasternak, e23871.

Trübutschek, Darinka, Sébastien Marti, Henrik Ueberschär, and Stanislas Dehaene (2019). "Probing the limits of activity-silent non-conscious working memory". *Proceedings of the National Academy of Sciences* 116.28, pp. 14358–14367.

Van Vugt, B., Bruno Dagnino, Devavrat Vartak, Houman Safaai, Stefano Panzeri, Stanislas Dehaene, and Pieter R. Roelfsema (2018). "The threshold for conscious report: Signal loss and response bias in visual and frontal cortex". *Science* 360.6388, pp. 537–542.

Van Vugt, Bram, Timo van Kerkoerle, Devavrat Vartak, and Pieter R. Roelfsema (2020). "The contribution of AMPA and NMDA receptors to persistent firing in the dorsolateral prefrontal cortex in working memory". *Journal of Neuroscience* 40.12, pp. 2458–2470.

Vijayraghavan, Susheel, Min Wang, Shari G. Birnbaum, Graham V. Williams, and Amy F. T. Arnsten (2007). "Inverted-U dopamine D1 receptor actions on prefrontal neurons engaged in working memory". *Nature Neuroscience* 10.3, pp. 376–384.

Vohryzek, Jakub, Gustavo Deco, Bruno Cessac, Morten L. Kringelbach, and Joana Cabral (2020). "Ghost Attractors in Spontaneous Brain Activity: Recurrent Excursions Into Functionally-Relevant BOLD Phase-Locking States". *Frontiers in Systems Neuroscience* 14, p. 20.

Wang, Min, Yang Yang, Ching-Jung Wang, Nao J. Gamo, Lu E. Jin, James A. Mazer, John H. Morrison, Xiao-Jing Wang, and Amy F. T. Arnsten (2013). "NMDA Receptors Subserve Persistent Neuronal Firing during Working Memory in Dorsolateral Prefrontal Cortex". *Neuron* 77.4, pp. 736–749.

Wang, Xiao-Jing (1999). "Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory". *Journal of Neuroscience* 19.21, pp. 9587–9603.

– (2002). "Probabilistic Decision Making by Slow Reverberation in Cortical Circuits". *Neuron* 36.5, pp. 955–968.

– (2021). "50 years of mnemonic persistent activity: quo vadis?" *Trends in Neurosciences* 44.11, pp. 888–902.

Wasmuht, D. F., E. Spaak, T. J. Buschman, E. K. Miller, and M. G. Stokes (2018). "Intrinsic neuronal dynamics predict distinct functional roles during working memory". *Nature Communications* 9.1, p. 3499.

Wong, Kong-Fatt and Xiao-Jing Wang (2006). "A recurrent network mechanism of time integration in perceptual decisions". *Journal of Neuroscience* 26.4, pp. 1314–1328.

Wyart, Valentin and Catherine Tallon-Baudry (2009). "How Ongoing Fluctuations in Human Visual Cortex Predict Perceptual Awareness: Baseline Shift versus Decision Bias". *Journal of Neuroscience* 29.27, pp. 8715–8725.

Yang, Guangyu Robert, John D. Murray, and Xiao-Jing Wang (2016). "A dendritic disinhibitory circuit mechanism for pathway-specific gating". *Nature Communications* 7.1, p. 12815.

Yang, Sheng-Tao, Min Wang, Constantinos D Paspalas, Johanna L Crimins, Marcus T Altman, James A Mazer, and Amy F T Arnsten (2018). "Core Differences in Synaptic Signaling Between Primary Visual and Dorsolateral Prefrontal Cortex". *Cerebral Cortex* 28.4, pp. 1458–1471.

Yang, Yang, Constantinos D. Paspalas, Lu E. Jin, Marina R. Picciotto, Amy F. T. Arnsten, and Min Wang (2013). "Nicotinic α7 receptors enhance NMDA cognitive circuits in dorsolateral prefrontal cortex". *Proceedings of the National Academy of Sciences* 110.29, pp. 12078–12083.

Yaron, Itay, Lucia Melloni, Michael Pitts, and Liad Mudrik (2021). "How are theories of consciousness empirically tested? The Consciousness Theories Studies (ConTraSt) database". *Journal of Vision* 21.9, p. 2195.

Yoo, Sang-Ah, Julio C. Martinez-Trujillo, Stefan Treue, John K. Tsotsos, and Mazyar Fallah (2021). "Feature-based attention induces surround suppression during the perception of visual motion". *bioRxiv*, p. 2021.02.17.431646.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.20.481230; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Yuste, Rafael, Michael J. Gutnick, Drorit Saar, Kerry R. Delaney, and David W. Tank (1994). "Ca2+ accumulations in dendrites of neocortical pyramidal neurons: An apical band and evidence for two functional compartments". *Neuron* 13.1, pp. 23–43.

Zandvakili, Amin and Adam Kohn (2015). "Coordinated neuronal activity enhances corticocortical communication". *Neuron* 87.4, pp. 827–839.

Zilles, Karl and Nicola Palomero-Gallagher (2017). "Multiple Transmitter Receptors in Regions and Layers of the Human Cerebral Cortex". *Frontiers in Neuroanatomy* 11.