Alertness modulates perceptual decision-making

Sridhar R. Jagannathan[1*], Corinne A. Bareham[1,2,3], Tristan A. Bekinschtein[1*]

[1]Department of Psychology, University of Cambridge, Cambridge, United Kingdom

[2]Department of Clinical neurosciences, University of Cambridge, Cambridge, United Kingdom

[3]School of Psychology, Victoria University of Wellington, Wellington, New Zealand

*Corresponding authors: j.sridharrajan@gmail.com, tb419@cam.ac.uk

ABSTRACT

The ability to make decisions based on external information, prior knowledge and context is a crucial aspect of cognition and it may determine the success and survival of an organism. Despite extensive and detailed work done on the decision making mechanisms, the understanding of the effects of arousal remain limited. Here we characterise behavioural and neural dynamics of decision making in awake and low alertness periods to characterise the compensatory signatures of the cognitive system when arousal decreases. We used an auditory tone-localisation task in human participants under conditions of fully awake and low arousal. Behavioural dynamics analyses using psychophysics, signal detection theory and drift-diffusion modelling showed slower responses, decreased performance and a lower rate of evidence accumulation due to alertness fluctuations. To understand the modulation in neural dynamics we used multivariate pattern analysis (decoding), identifying a shift in the temporal and spatial signatures involved. Finally, we connected the computational parameters identified in the drift diffusion modelling with neural signatures, capturing the effective lag exerted by alertness in the neurocognitive system underlying decision making. These results define the reconfiguration of the brain networks, regions and dynamics needed for the implementation of perceptual decision making, revealing mechanisms of resilience of cognition when challenged by decreases in arousal.

INTRODUCTION

The question of how decisions are made has shaped the world's systems of government, justice and social order (Buchanan and O'Connell, 2006). From a cognitive perspective, the modelling of psychological and neural factors allows for the search of underlying brain mechanisms, which are under specific constraints of the information provided from the external world and

the internal state of the corresponding system. Studies on how the brain implements simple decisions have revealed several neurocognitive processes at the perceptual, central integration and motor implementation levels (Sigman and Dehaene, 2005, 2008), but the modulatory effects of the internal milieu, homeostasis, arousal, alertness, and circadian influences on such processes have received less attention (Hull et al., 2003; Knowles, 1993). Specifically, alertness fluctuations on decision processes have been tackled by sleep deprivation and brain injury studies but hardly by normal variations in the stability of wakefulness (Goupil and Bekinschtein, 2012). The question arising is; how does the brain compensate for the detrimental effects of decreasing arousal while making an optimal decision?

Perceptual decision making is referred to as a set of processes which involves processing sensory information and making a single behavioural choice from a set of available options (Heekeren et al., 2008). The various elements of this process include: a) a sensory system that transforms physical stimulus intensities into decision variables in the brain; b) a decision system that integrates and accumulates such decision variables and makes an optimal choice of based on relative evidence; and c) a motor system that implements the appropriate motor plan for the choice by pressing a response button etc. The brain regions involved in such processes have been revealed by neuroimaging studies in monkeys and human observers. Signals from domain general regions like right insula, posterior left SFS/DLPFC, left TPJ, posterior parietal cortex (LIP) were identified as processes responsible for evidence accumulation (Heekeren et al., 2004, 2006; Ho et al., 2009; Roitman and Shadlen, 2002; Shadlen and Newsome, 1996, 2001). However, these studies have not directly measured the effect of alertness modulations on evidence accumulation or decision making parameters. Specifically alertness related modulations can be classed into 'tonic' fluctuations that span multiple trials and time-periods, and 'phasic' that are moment to moment modulations in arousal produced in response to an ongoing task. A few recent studies have shown in humans and mice that arousal measured by brain-stem based systems (using pupil response) modulate individual decision making, specifically relating task-evoked, moment to moment fluctuations of phasic arousal (or alertness) (de Gee et al., 2017; McGinley et al., 2015). Further to this, van Kempen and collaborators (van Kempen et al., 2019) showed that lower tonic and higher phasic arousal, as defined by the pupil measurements, predicted shorter reaction times and was associated with a centroparietal positivity in EEG space. All of the above mentioned studies though have only used proxies in the form of pupil response to measure fluctuations in arousal, hence several other physiological parameters are needed concurrently to accurately measure arousal (Wang et al., 2018). Furthermore they have not measured influences of tonic fluctuations alone. Hence, so far no study has directly investigated the effect of intrinsic alertness (tonic) fluctuations on decision-making systems as the arousal decreases.

While experiments in spontaneous fluctuations in tonic alertness are difficult to perform in animals since they stop responding when arousal drops, some have circumvented this problem by using sleep deprivation and measuring brain activity after a prolonged wakefulness period (Vyazovskiy et al., 2011). In the case of humans, patient studies have revealed the effect of

alertness on cognitive processes like attention. Patients that sustain a stroke and then have a lesion on one side of the brain have deficits in general levels of alertness, sometimes they also have difficulty in localising and paying attention to information on the side opposite to the lesion. This condition is referred to as unilateral spatial neglect and is usually, more persistent (Bowen et al., 1999; Karnath and Fetter, 1995) after a stroke in the right hemisphere (compared to left hemisphere) and hence also called *left neglect* (due to lack of attention to left side of space). This asymmetry is thought to arise due to an interaction between processes responsible for spatial attention: a bilateral dorsal fronto-parietal network and an alertness-related right lateralized ventral attention network (Corbetta and Shulman, 2002, 2011). Other studies have reasoned this behavioural asymmetry due to competition between hemispheres (Kinsbourne, 1970, 1977). However, it is important to note that studies in patients pose a problem to dissect the systematic effects of alertness on decision making due to the heterogeneous nature of the location and extent of lesions, and due to confounds in computational methods that attempt to directly measure tonic alertness levels. Hence an ideal scenario to study the effect of alertness, would be a study exposing such behavioural asymmetries (attentional bias) in attention in healthy participants while their alertness fluctuates (Bareham et al., 2014).

Following this line of thought, studies with healthy participants (Bareham et al., 2014; Goupil and Bekinschtein, 2012; Noreika et al.) have captured the effect of alertness on attention, semantic-audiomotor mapping and conscious access. However, these studies use spectral power (alpha-theta ratio) as a continuous measure or visually defined grapho-elements in the transition to sleep (Hori et al., 1994) to define the level of drowsiness. While measures using alpha-theta assume participants being equally alert and drowsy (not always true) and manual Hori scoring suffers from subjectivity dependent on the experience of the human scorer. Recently we published a novel computational method to track alertness levels using electroencephalography (EEG) that shows robust outcomes (Jagannathan et al., 2018), and has been successfully applied in other studies to explore the effect of alertness on executive control systems (Canales-Johnson et al.). Here we decided to take advantage of this technique and use an experiment (auditory spatial localisation task) with a known behavioural asymmetry (Bareham et al., 2014) to investigate how decision-making is modulated by fluctuations in tonic alertness.

In this study we aim to map the neurobehavioural changes in the decision making processes in fully alert and in moments of decreased alertness. First, we use a combination of multi-level modelling, psychophysics and signal detection theory to show how behaviour is modulated by variations in alertness. Second, we employ a hierarchical drift diffusion model to tease apart and parameterise the different elements of the decision-making process and examine individually how they are modulated by alertness. Third, we use decoding at the temporo-spatial levels to understand how the neural markers of such decision-making processes may be implemented in the brain, and how such neural processes are modulated by fluctuations in alertness. Fourth, we connect the distinct behavioural parameters of the decision making process to the neural parameters, thereby capturing how the brain reconfigures, spatial and temporally, to maintain behavioural performance when challenged by decreased alertness.

RESULTS

Behavioural evidence of decision making modulated by alertness

Error-proportion modulated by alertness

We first used multilevel modelling to understand how the errors made by each subject in an auditory tone-localization task (Figure 1A) was influenced by the stimulus presented ('left' or 'right' auditory tone) and the state of the subject ('alert' or 'drowsy'). We defined 4 different multilevel (linear mixed) models where errors were modulated by various combinations of stimulus and alertness states (see methods section). The analysis of variance table of the winning model shows that both alertness ($F(1, 95.07) = 14.04$, $p<0.0001$) and stimulus type ($F(1,95.07) = 18.80$, $p<0.0001$) have an effect on error-proportion. Further, there was a reliable interaction between alertness levels and stimulus type ($F(1,95.07) = 6.88$, $p<0.05$). Next, post-hoc analysis revealed a reliable difference between alert and drowsy conditions for left stimuli ($p<0.001$), and not for right stimuli ($p = 0.859$). These behavioural results (Figure 1B) replicate the findings of (Bareham et al., 2014), reporting an increase in location assignment errors on tones from the left side of the midline when people became drowsy.

Reaction times are modulated by alertness

Second, we aimed to quantify the modulation in the response profiles (reaction times) of individual subjects by alertness levels. A paired samples t-test with $t(31) = -7.78$, $p<0.05$ revealed a reliable effect of alertness on reaction times. These behavioural results (Figure 1C) replicate the findings of (Hori et al., 1994) indicating slower reaction times under lower levels of alertness.

Subjective midline modulated by alertness

Third, we used psychophysics to quantify the modulation of subjective midline per subject by alertness levels. For this purpose we fit a cumulative normal function (see methods section) to the proportion of rightward responses (per subject) under each stimulus condition from -60° to 60° under both alert and drowsy periods. The fit was weighted by the number of trials (responses) in each stimulus condition. The mean of the function referred to as 'bias' is the subjective midline (where subjects have 0.5 chance of pressing left or right responses). Most subjects (Figure 1D) had their bias point shifted to the left (as they became drowiser), indicating more left errors (as they overestimate the right side of space). Some subjects had bias points shifted to the right. Overall, a paired samples t-test with $t(20) = 3.33$, $p<0.005$ revealed a reliable difference in bias points between alert and drowsy periods.

Next, we also plotted (Figure 1E) the mean of the psychometric fits of individual subjects to show that overall the subjective midline has shifted to the left at the group level.

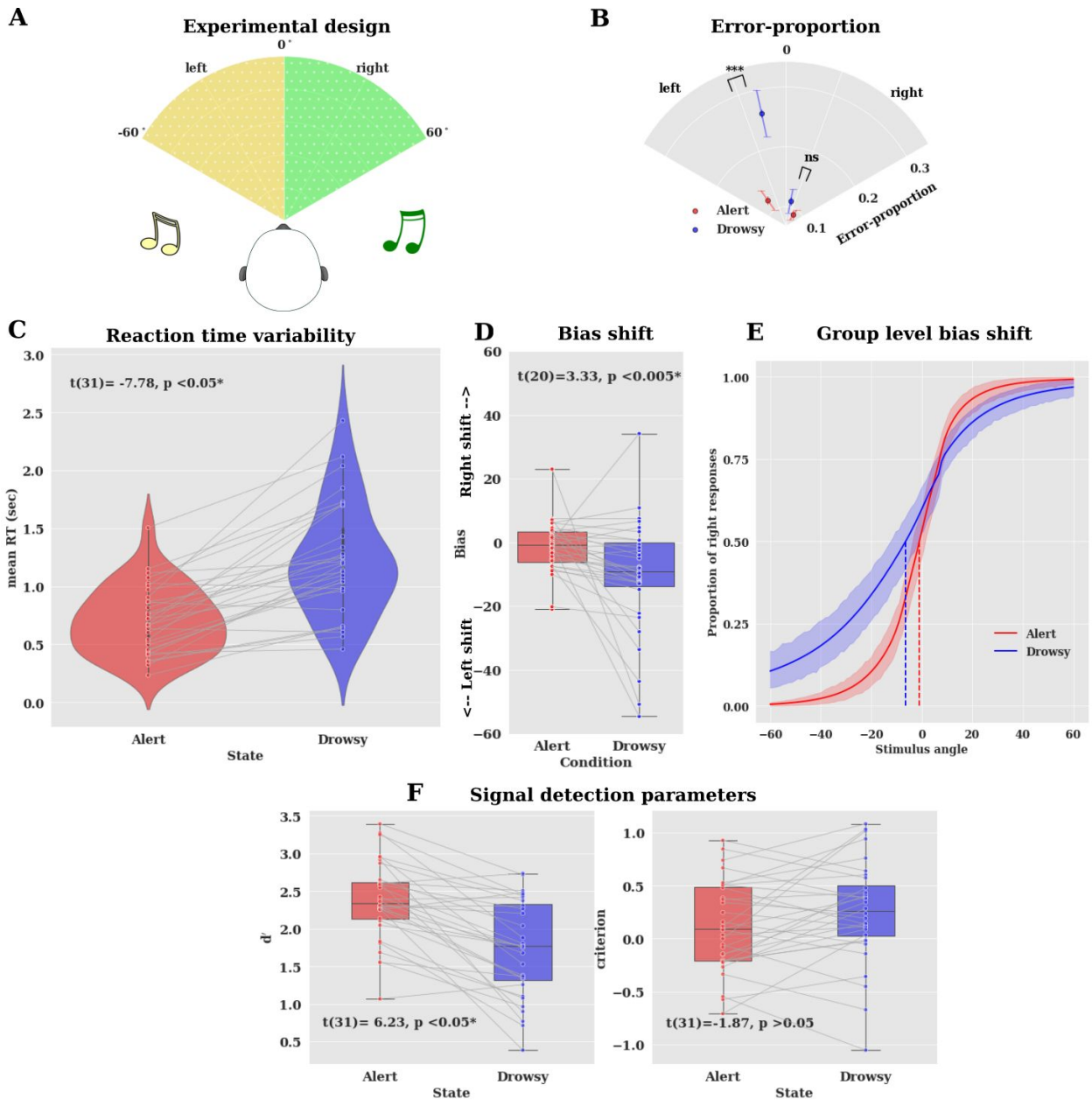

Figure 1: Auditory spatial attention task: A) subjects had to localize the direction of auditory tones coming from left and right side of the midline. B) proportion of errors committed in alert and drowsy periods across left and right stimuli. Multilevel modelling reveals that error rates depend on stimulus type (left,right) and state of subject (alert,drowsy). Post-hoc tests based on estimated marginal means indicate that error proportion is reliably modulated by alertness but only for left stimuli. *** indicates $p<0.001$, ns indicates not significant (or not reliable), error bars indicate standard error of the mean. C) mean reaction times for individual subjects. Reaction times are variable and slower under drowsy conditions. D) bias value per subject (using psychometric fits), bias level shifts towards the left side (indicating more left errors) for most subjects. Negative bias values indicate shifts in the subjective midline towards the left and positive values indicate shift towards the right. E) group level psychometric fits indicate the shift in subjective midline (dotted lines), shaded regions are confidence interval bounds.

F) signal detection analysis shows that only d' (sensitivity) is modulated by alertness and criterion remains unaffected.

Signal detection parameters modulated by alertness

Fourth, we used signal detection theory to understand what factors modulate decision making under varying levels of alertness. d' (sensitivity) was modulated by alertness with $t(31) = 6.23$, $p<0.05$ however criterion (response bias) was not modulated by alertness levels with $t(31) = -1.87$, $p>0.05$. This clearly indicates that internal representations in the brain in terms of sensory and noise distributions are modulated by alertness levels. Further, the insignificance of the response bias also indicates that subjects were not arbitrarily pressing right responses for uncertain stimuli.

Computational modelling of decision making modulated by alertness

Next, we aimed to quantify the different elements of the decision-making process using the drift-diffusion model. The drift-diffusion model captures the optimal procedure involved in performing a 2-alternative forced choice (2AFC) task. It assumes that the observer accumulates evidence for one or other alternative in every other time step (Ratcliff et al., 2016), until that integrated evidence reaches a threshold to make a decision (Figure 2A). The localization of tones to the left and right side of space is essentially a 2-choice task with the subject always forced to make a decision on the direction of tone. The model was implemented with a hierarchical Bayesian procedure using hierarchical drift diffusion model (HDDM) (See methods section). For the HDDMs, we fit the response of each subject instead of accuracy. This procedure is referred to as Stimulus-coding and is critical to uncover response bias (de Gee et al., 2017).

We examined 8 different variants of the model (see methods section), allowing parameters like: drift-rate (v), bias-point (z) to vary based on state (alert or drowsy) or stimulus (left or right) (Figure 2B,C). The winning model was chosen as the one with lowest deviance information criterion (DIC) which provides a balance between model fit and complexity. The winning model (#8) was composed of: drift-rate (v) varied according to state (alert or drowsy) and stimulus, whilst bias-point (z) varied according to stimulus (left or right) only. The winning model was then analysed for differences in posterior densities of parameters. For this purpose, we used a bayesian estimate which is fundamentally more informative and avoids the arbitrary choices like significance level and specific statistical tests used by the frequentist based methods (Kruschke, 2012).

We found in the winning model, the proportion of posterior overlap (between left and right stimuli) in the bias point was 26.1% (Figure 2E). This indicates that the bias-point was not reliably different between the different stimuli. Further, the proportion of posterior overlap (between left and right stimuli) in the drift rate for alert trials was 14.8% this reduces to 3.6% for drowsy trials (Figure 2H). This indicates that the drift-rate (evidence accumulation rate) was reliably different between left and right stimuli on drowsy trials.

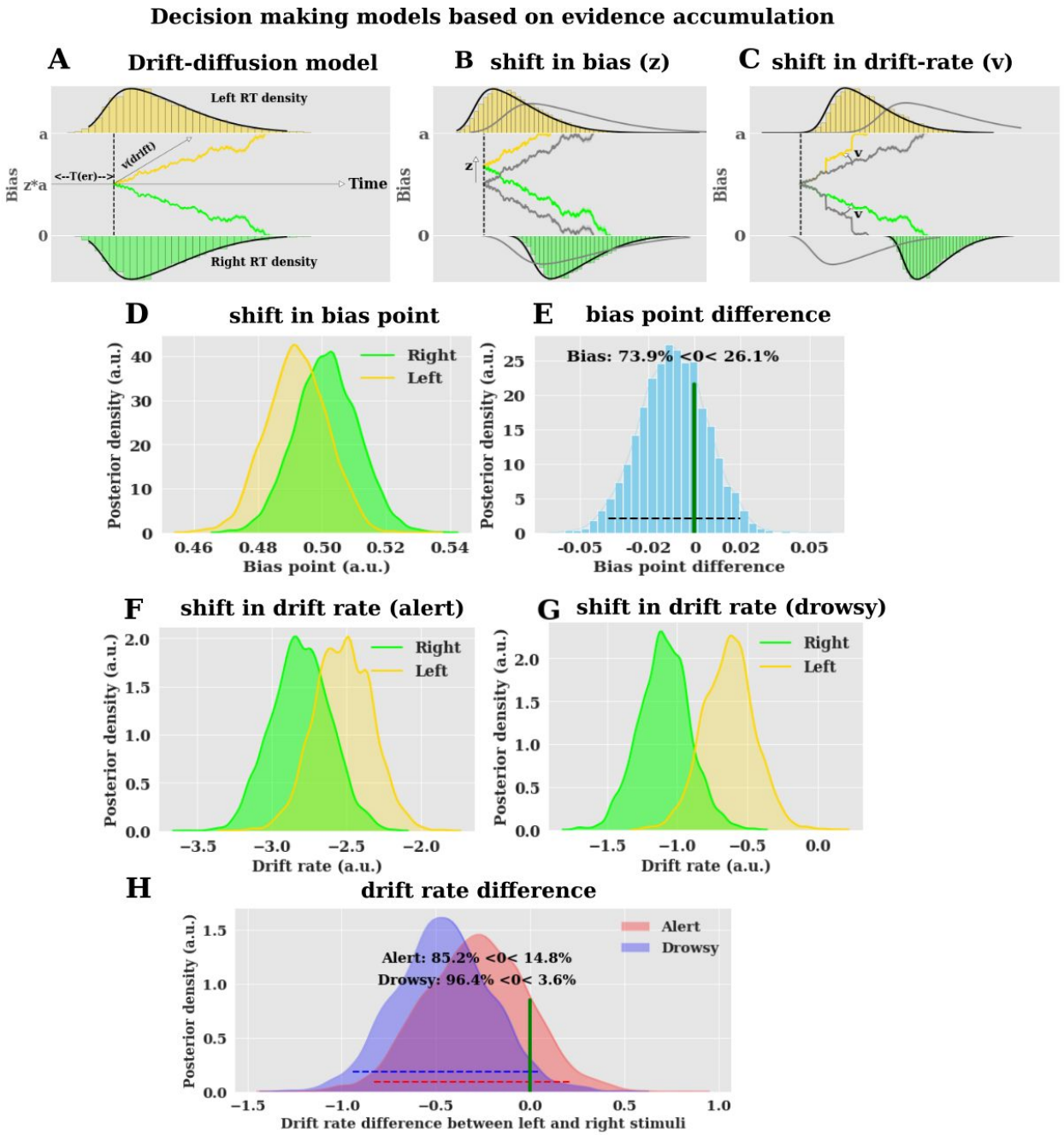

Figure 2: Evidence accumulation models: A) Drift-diffusion model accounts for the reaction time distributions of responses across left and right stimuli ('Stimulus Coding'). 'v(drift)' indicates evidence accumulation rate, 'a' indicates the boundary separation across left and right responses, 'z' indicates the bias point, usually z = 0.5 for unbiased responses. B,C) changes in response distributions explained by shift in 'z', shift in 'v' respectively from alert to drowsy trials. grey lines indicate unbiased condition. D) posterior densities of 'z' for left and right stimuli in the winning model #8. a.u. indicates arbitrary units. E) differences in the posterior densities indicate no evidence of bias point difference across stimuli. F,G) posterior densities of 'v' for left and right stimuli across alert and drowsy periods in the winning model #8. H) differences in the posterior densities of difference in drift rate indicate strong evidence in favour of drift rate difference across stimuli in the drowsy period compared to alert periods.

Neural evidence of decision making modulated by alertness

Next, we aimed to uncover how the neural patterns in the EEG data are modulated by alertness levels. Conventional ERP (Event related potentials) analysis for this purpose, relies on a-priori identified spatial locations or temporal segments to measure the differences across conditions. However decoding techniques do not rely on such a-priori definitions and perform much better in detecting differences across experimental conditions (Fahrenfort et al., 2018). We applied temporal decoding (see methods section) to understand the differences in the patterns of EEG data across conditions.

Spatial and temporal signatures involved in spatial localization across alert and drowsy periods:

First, we were interested in identifying the neural signatures involved in the performance of this task during alert condition. For this purpose decoding involved in identifying the stimulus (X - left or right tone) presented from the EEG data (Y). This process involves the identification of the W (classifier weights) that can produce the transformation, $Y_t = W_tX_t$ where 't' represents time (Figure 3A,B). The performance of the classifier (W) is evaluated by training and testing the data at each time-point (t) using a measure such as the area under the curve (AUC). In Figure 3C, the shaded region represents statistically significant periods. For more details refer to the methods section.

We found that when the subject was 'alert' the decoding of stimuli (Figure 3D) started from 160 ms after the stimulus was presented and lasted until 720 ms (reliable with cluster permutation, $p<0.05$) with mean AUC of 0.57. The peak discriminatory power was at 280 ms (AUC = 0.61). The average AUC between 200-300 ms was 0.58 ± 0.007, $p<0.05$, between 300-400 ms was 0.56 ± 0.002, $p<0.05$. However when the subjects became 'drowsy' the decoding of stimuli (Figure 3E) started from 420 ms after the stimulus was presented and lasted until 730 ms (reliable with cluster permutation, $p<0.05$) with mean AUC of 0.54. The peak discriminatory power was at 590 ms (AUC = 0.56). The average AUC between 420-500 ms was M = 0.53 ± 0.002, $p<0.05$, between 510-590 ms was M = 0.54 ± 0.004, $p<0.05$. This indicates that subjects are slower to discriminate between left and right stimuli under drowsy conditions and also the discriminatory power is lower.

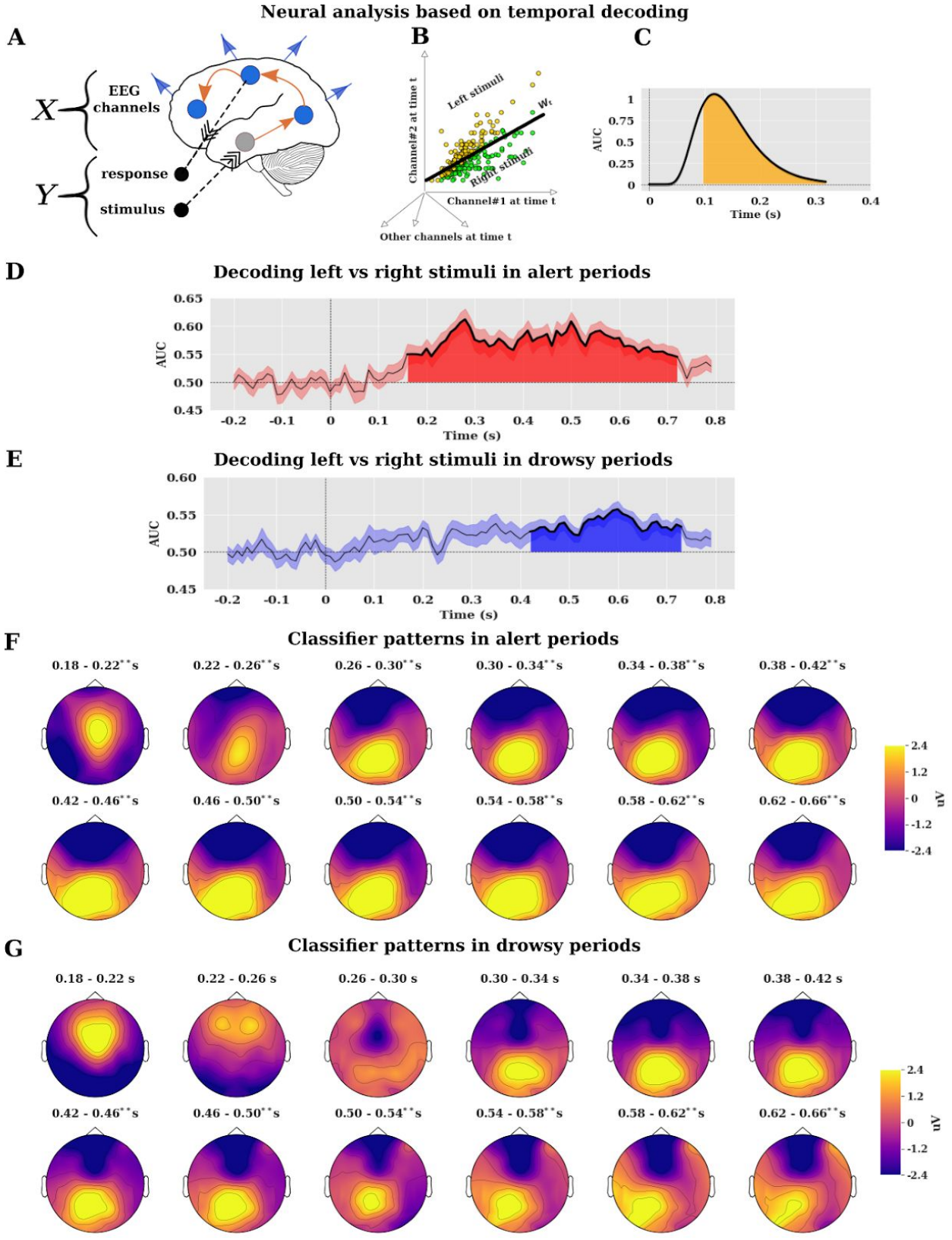

Figure 3: Temporal decoding: A) decoding consists of identifying Y (responses or stimuli) from X (EEG data). The model thus consists of $Y_t = W_t X_t$, where 't' represents times and W_t represents transformation

(classifier weights). B) classifier weights W_t are determined by the optimal separation of different classes (here left, right stimuli). C) decoding performance is assessed using the area under curve (AUC) where shaded regions represent reliably different time periods. D) AUC under alert periods, where the classifier was trained to discriminate between targets of left and right stimuli. Shaded regions show time periods with reliable discriminatory power. E) AUC under drowsy periods, where the classifier was trained to discriminate between targets of left and right stimuli. Shaded regions show time periods with reliable discriminatory power, note that the scale is different when compared to alert periods. F,G) Coefficients of classifier patterns (derived from classifier weights) under alert and drowsy periods. These patterns explain how EEG data are generated from discriminatory neural sources and the strength is directly proportional to discriminability of those channels. ** indicates $p<0.05$ in the corresponding period in temporal decoding.

Next, we plotted the coefficients of the classifier patterns (derived from classifier weights W) which are neurophysiologically interpretable (Haufe et al., 2014). The classifier patterns for the 'alert' time periods for every 40 ms between 180 ms to 660 ms is presented in Figure 3F. The pattern between 180-220 ms indicates strong response in the fronto-central electrodes. The pattern shifts to more posterior regions (centro-parietal electrodes in the right side of the scalp) between 220 ms to 260 ms. Further, the topography of the patterns shifts to more parietal and occipital electrodes from 260 ms to 300 ms. This indicates that the decoding pattern moves from the frontal electrodes to more posterior (central, parietal) indicating that the decision making areas would possibly converge on parietal regions in the brain.

When the subjects become drowsier, the classifier patterns (Figure 3G) between 180 - 220 ms indicates the strong response in the frontal electrodes (though not reliably decodable). Interestingly the topography from 220-260 ms stays in frontal electrodes (not reliable). The pattern shifts to more posterior regions (centro-parietal electrodes, reliably decodable) from 380 ms to 420 ms. Further, the pattern shifts to more parietal and occipital electrodes between 420 ms to 460 ms. This suggests that evidence accumulation in drowsy periods could possibly be delayed when compared to alert periods.

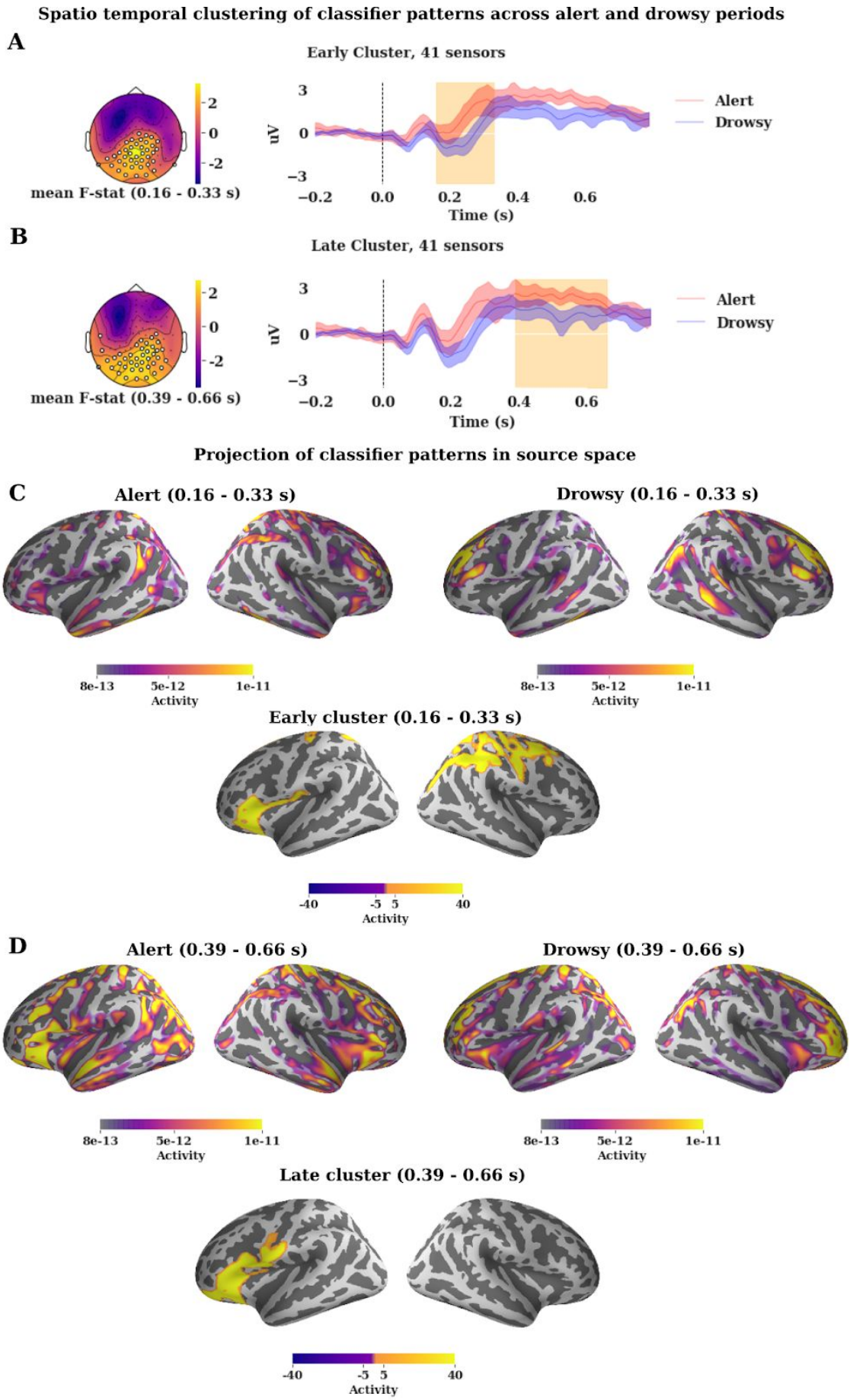

Figure 4: Classifier patterns: A,B) spatio-temporal clustering of differences in classifier patterns shown in Figure 3 (F,G) between alert and drowsy periods. The differences indicate a cluster in early (160 - 330 ms after stimulus) and later (390 - 660 ms after stimulus) time periods. C) classifier patterns (early) in electrode space are projected to source space and spatio-temporal clustering identifies differences in source patterns (mostly in right parietal regions) D) classifier patterns (later) in sensor space are projected to source space and spatio-temporal clustering identifies differences in source patterns (mostly in left frontal regions).

The descriptive analysis of the classifier patterns indicate differences between alert and drowsy periods. To establish the spatial and temporal signatures of such differences we performed a cluster permutation test and identified regions where activity patterns in alert periods is higher than drowsy periods. The other contrast (drowsy periods > alert periods) is not meaningful as the AUC (discriminatory power) in alert periods is overall always higher than drowsy periods (Figure 3 D,E). The permutation test revealed an early cluster (160 ms to 330 ms) and a late cluster (390 ms to 660 ms). The electrodes that are reliably different in both the early and late cluster along with the mean temporal signature of the classifier pattern in those electrodes is plotted in Figure 4A,B.

To identify the neural sources of these differences in electrodes, we used a source reconstruction procedure to project the classifier patterns of both the early and later cluster back to their neural sources separately (see methods section). In the source space again we performed cluster permutation tests to identify neural regions where activity in alert periods is higher than drowsy periods. We used the actual value of the source activity (instead of absolute values), as we are interested in the distance between the two different patterns and not in overall activity levels. In the early cluster, the permutation tests revealed regions in the right hemisphere, mainly superior, inferior parietal and pre, post central gyrus etc and left prefrontal cortex. In the later cluster, the permutation tests revealed a region in the left frontal side.

The above analysis reveals regions of the brain wherein elements of the decision making process (by using classifier patterns) may be implemented in the alert periods, however, the critical piece still missing in the puzzle is how these implementations are modified in the drowsy periods.

Neurobehavioral evidence of decision making modulated by alertness

Next, we aimed to uncover the neural signatures of how the elements of the decision making process might be implemented in the drowsy periods. To establish this we decided to develop a novel method based on the drift-diffusion analysis developed earlier. First, we used the winning model (#8) and computed trial-by-trial drift parameter (dv). Second, we used dv as a dependent variable and regressed the same against ERP data (see methods section).

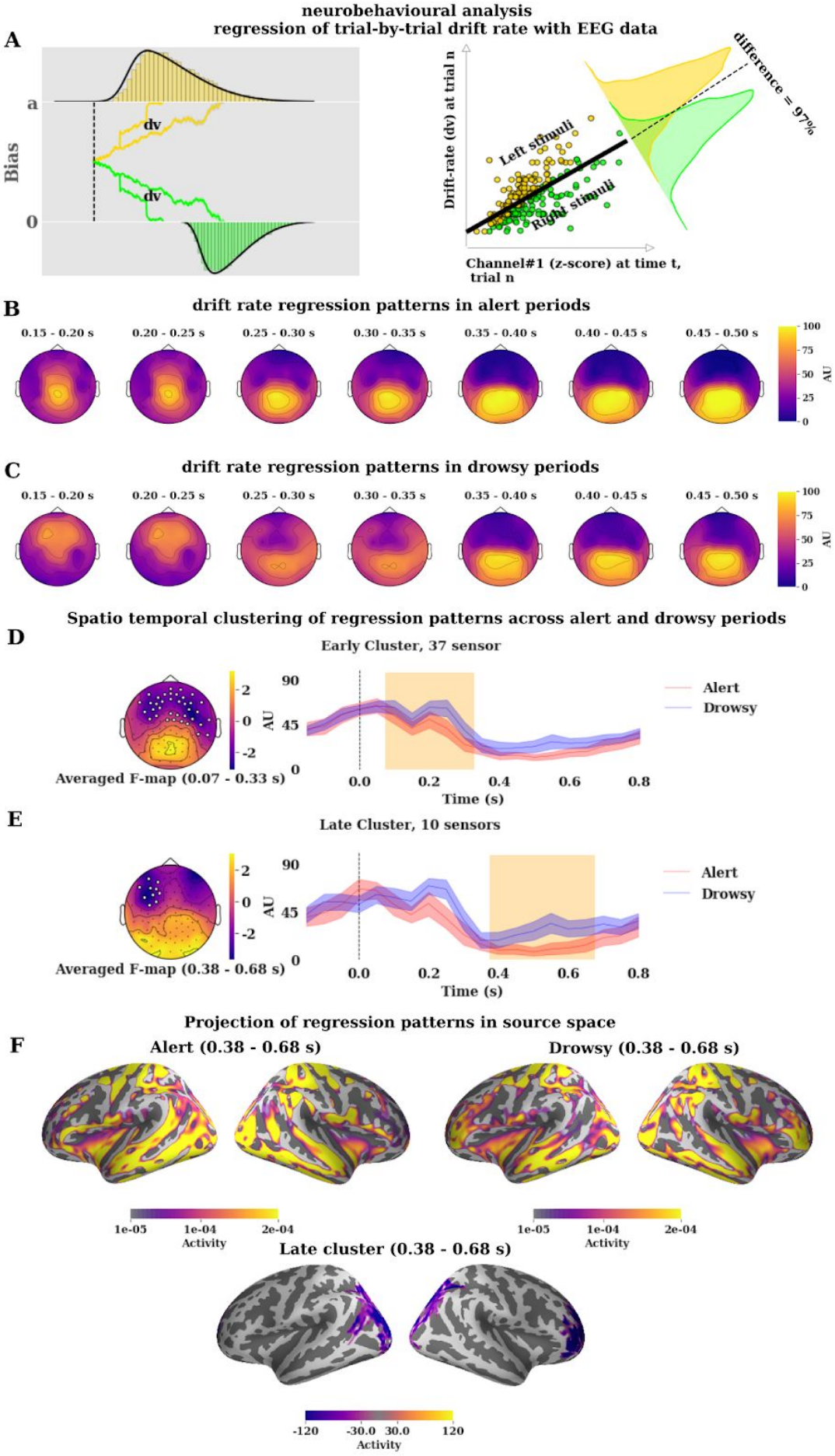

Figure 5: Regression patterns: A) trial by trial drift rate (dv) was regressed against EEG data (z-scored and averaged per 50 ms), the difference in the posterior of regression rate was quantified per electrode per time-point per subject. B,C) mean regression patterns in alert and drowsy periods. Regression patterns could be interpreted like activation patterns in Figure 3 (F,G), higher values indicate greater discriminatory power to identify between left and right stimuli. D,E) spatio-temporal clustering identifies differences in regression patterns between alert and drowsy periods in early (70 - 330 ms after stimulus) and later (380 - 680 ms after stimulus) time periods. F) regression patterns (later time period) in sensor space are projected to source space and spatio-temporal clustering in source space identifies differences in source patterns (right and left parietal regions).

It can be expressed as

$$dv \sim \beta_0 + \beta_1(ERP)$$

The above equation can be written in patsy form as below.

$$dv \sim ERP : C(state, Treatment(`Alert`)) : C(stim, Treatment(`Right`))$$

Here, dv represents trial-by-trial drift-rate, ERP represents z-scored ERP data per trial per time point per electrode, state represents alertness levels ('alert' or 'drowsy'), stim represents stimulus types ('left' or 'right'). Further we compute the proportion of the overlap of the posterior distributions of the trace obtained for both left and right stimuli separately under alert and drowsy condition (Figure 5A). This analysis is repeated for each subject and we obtain regression patterns (similar to the classifier patterns) per subject (see methods section).

Next, we plot the regression patterns in alert and drowsy periods. The patterns in the alert periods (Figure 5B) indicate that the topography focuses initially in the fronto-central electrodes (150 ms to 250 ms), which then shifts to centro-parietal electrodes (250 ms to 350 ms). The patterns in the drowsy periods (Figure 5C) indicate that the topography initially focuses on the frontal electrodes (though weaker compared to alert periods) in the interval from 150 ms to 250 ms. Further the patterns shift to more central electrodes (again weaker when compared to alert periods) in the interval from 250 ms to 350 ms.

To establish the spatial and temporal signatures of such differences we performed a cluster permutation test and identified regions where "activity" in drowsy periods is higher than alert periods (opposite of the classifier analysis). The permutation test revealed an early cluster (70 ms to 330 ms) and a late cluster (380 ms to 680 ms). The electrodes that are reliably different in both the early and late cluster along with the mean temporal signature of the regression pattern in those electrodes is plotted in Figure 5D,E.

Further, we projected the patterns in source space and performed cluster permutation analysis as before. We used the absolute value of the source activity (instead of the actual value) as the proportion of posterior overlaps has always a positive sign in the sensor space, furthermore using absolute values helps us understand which condition generates a larger response at

specific time points. There were no reliable differences in the source space for the early cluster, For the later cluster, the permutation tests revealed posterior parietal regions in both the left and right hemisphere and right frontal pole region. This analysis reveals the regions of the brain wherein elements of the decision making process (by using regression patterns) are possibly implemented in the drowsy periods.

DISCUSSION

In this project, we were interested in, first, how changes of arousal modulate specific aspects of spatial attention, and second, how general mechanisms of perceptual decision making are affected by tonic alertness fluctuations. By using our previous data and results (Bareham et al., 2014; Noreika et al.) we were able to formulate specific hypotheses to test particular aspects of theories in cognition, informative to general frameworks of decision making.

We investigated how alertness modulates different neurocognitive components of perceptual decision-making using a tone localisation task. We replicated the spatial bias induced by alertness changes that has been shown in healthy participants and in patients with brain lesions (spatial neglect). We characterised behavioural changes in decision making as impacted by the fluctuations in arousal, revealing a decrease in performance due to sensitivity to the stimuli but not response bias. We also showed a specific effect of alertness delaying the rate of evidence accumulation in a drift diffusion model. At the neural level we show that the formation of brain patterns (relating to stimuli presented) is delayed and decreased by lower levels of alertness, suggesting a reconfiguration of the neural networks of the decision. We confirmed this effect by mapping the behavioural model parameters of evidence accumulation into the neural evoked response, showing delays in the neural processes and further recruitment of parietal and frontal areas. For spatial attention theories we were able to add support to the role of distributed parietal networks when attention systems are challenged, and to tease apart the change in the predisposition of the system (bias-point, prior) from the modifications to the decision process itself (drift-rate). For decision making theories, we have added a missing level of description by characterising possible mechanisms of resilience of the neurocognitive processes when tonic alertness decreases. We were able to show the behavioural flexibility of the system, its behavioural dynamics, and capture the temporal and spatial neural progression and reconfiguration caused by the change in the internal context - describing its cognitive homeostasis.

Behavioural changes in decision making show specific modulations of spatial attention with decreased arousal

We established, using multi-level modelling, that the errors for left-sided stimuli increased as participants became drowsy. However the same did not occur for stimuli coming from the right side. This directly replicates (Bareham et al., 2014), and provides evidence for inattention or neglect of stimuli originating from the left side when healthy people become drowsy.

Furthering the previous behavioural results, we fitted the proportion of rightward responses of individual subjects and identified the subjective midline (tone angle where a subject is equally likely to press left or right response), finding that the subjective midline shifted to the left side (indicating more right responses) as individuals became drowsy. This is in agreement with line-bisection studies (Jewell and McCourt, 2000) showing healthy individuals bisect horizontally presented lines to the left of the veridical midline, overestimating the right side of space. Further studies (Benwell et al., 2014) have also identified possible neural origins of such bias ('pseudoneglect') and identify task independent activity in the ventral attention network as one of the main modulating factors behind this shift. However these studies did not probe whether trial to trial alertness of individuals was modulated. Several studies (Corbetta and Shulman, 2011; Robertson et al., 1998) have shown that activity in the ventral attention network interacts with the arousal system (tonic alertness). Hence arousal instability, as shown directly in this study, could be a cause for these effects.

Evidence accumulation is delayed by decreases in alertness

To characterise more general aspects of decision making modulated by alertness, we decided to dissect the different elements of the decision-making process. Using signal detection theory to identify the changes in sensory signal and noise distributions due to alertness, we found that sensitivity (d') was reliably modulated by alertness whereas criterion (response bias) was not. This suggests that the sensory representations have indeed been modulated by alertness, and that participants were not arbitrarily pressing more rightward responses (in the face of uncertain stimuli) as their alertness decreases. To complement signal detection theory and further characterise how arousal modulates decision making, we used the rich variability in reaction times produced by awake and drowsy trials in a drift diffusion model, under the assumption that in each trial noisy evidence is slowly accumulated over time until criterion to make a decision is reached (Ratcliff et al., 2016). We constructed several models that allowed independent variation of parameters, such as starting point (bias) and drift rate (evidence accumulation), based on stimulus type and alertness levels of the participant. The statistical model with the higher evidence showed that drift rate was modulated by both alertness levels and stimulus type, whereas the starting point was only affected by stimulus type (left or right tone). Further analysis of the winning model also revealed that the evidence for modulation of the starting point was weak while that for drift rate was reliably different for stimuli from the left side compared to the right side. This shows that alertness indeed differentially modulates evidence coming from the left side of space compared to the right side. Decision making studies in the visual modality (Smith and Ratcliff, 2009) have shown that drift rates are closely related to the strength of stimulus encoding (attentional strength) and evidence accumulation to a decision threshold. Further studies have also shown neural correlates of such parameters in the EEG (Nunez et al., 2017; O'Connell et al., 2012). Hence we consider the effect of alertness modulation on drift rate as acting specifically on spatial attention itself, as well as general mechanisms by which partially independent systems, arousal and decision making, interact.

Alertness delays the neural signatures of perceptual decision making and alters the dynamics of the neural networks temporally and spatially

We probed the neural signatures of decision making processes using decoding analysis. The primary aim was to understand how the brain decodes information to "decide" if the stimuli presented was 'left' or 'right' under alert and drowsy conditions. Under alert conditions the decoding starts early, at 160 ms and lasts until 720 ms, however, under drowsy conditions decoding starts later, at 420 ms and lasts till 730 ms. This ~260 ms delay may indicate that, under lower arousal, the brain requires a longer time to decipher the direction of stimuli as suggested by slower reaction times and a change in the drift (evidence accumulation) of the diffusion model. Furthermore, the overall strength of decoding of the alert condition was systematically higher than that of drowsy, which may suggest that the brain processes responsible for decision making are less variable under the alert condition compared to the drowsy condition, or it may indicate less efficient decision processing during low alertness.

Furthermore, we aimed to understand how such neural signatures are depicted at the sensor level. For this purpose, we identified classifier patterns, which are neurophysiologically interpretable (Haufe et al., 2014) from classifier weights. In the alert condition, the early pattern (180-220 ms) indicates activity over frontal, central sites, further shifting to more central and parietal sites later on. In the drowsy condition, the pattern initially starts at frontal sites (though not enough to be decoded) and shifts to the more central and parietal sites later on from 380 ms onwards. This indicates that the formation of central, parietal patterns takes longer in drowsy periods compared to alert. Thus the activity over central and parietal electrodes could be considered vital for deployment of spatial attention and evidence accumulation. The spatial and temporal distribution of the decoding pattern resembles both a classic P300, which has been dubbed as the build-to-threshold signal (Twomey et al., 2015), and the centro-parietal positivity (CPP), a more specific signal of evidence accumulation (Loughnane et al., 2016; O'Connell et al., 2012). O'Connell et al showed that the CPP (an EEG evoked response) tracked evidence accumulation in perceptual decision making and was sensitive to perturbation, opening the window for the tracking of signals with sufficient temporal resolution to map the dynamics of decision making in humans (O'Connell et al., 2012) beyond what we have learnt in monkeys (Shadlen and Newsome, 2001). Following this line of thought, our results highlight the neural dynamics of decision making when the external world remains unaltered, physical evidence held constant, but the internal milieu fluctuates exerting a modulatory influence on the cognitive decision making systems that caused them to reconfigure to solve the task at hand.

As we identified the differences in the classifier patterns across alert and drowsy periods at both the sensor and source space level, the sensor level showed early and late clusters with only the late cluster showing reliable decoding in drowsy metastable states. To identify the spatial location of these patterns we projected the data at the sensor level to the source space in the two clusters. The early cluster mainly revealed a dominance in the right hemisphere, superior and inferior parietal, temporo-parietal junctions along with regions of the central gyrus. This adds

evidence to a body of attention studies across modalities (Dietz et al., 2014; Shulman et al., 2010) where right-hemisphere dominance is shown to occur in frontal, parietal and temporal sites (specifically in the temporo-parietal junction). The later cluster reveals mainly frontal regions of the left hemisphere, sometimes associated with fronto-parietal interaction that has been shown as early correlates of decision making. Here we have identified the neural signatures of decision making during alert periods and how it is distinctly different from the drowsy periods. It must also be noted that alertness itself, in the classic attentional aspect (Posner, 2008), has been associated to right hemisphere networks (Petersen and Posner, 2012; Sturm and Willmes, 2001) and the current results could be interpreted in that framework as a reconfiguration of those attentional network. Alternatively, we could interpret our findings as an indication for the alerting network impacting the evidence accumulation network, where the neural reconfiguration is a reflection of neural resilience and compensatory mechanisms that the human brain develops to maintain performance when challenged (Canales-Johnson et al.). We think that aspects of wakefulness (Bekinschtein et al., 2009; Goupil and Bekinschtein, 2012) and cognitive flexibility should be called upon to gain explanatory power in light of other frameworks.

Evidence accumulation reveals a late and extended frontoparietal neural network to resolve the decision when challenged by arousal

The decoding analysis is not guided by parameters that directly test mechanistic hypotheses, and the decoding score of the alert periods are always higher than drowsy periods, shielding any contrast to only reveal signatures of the primacy of the alert patterns. To overcome this issue and guide neural exploration with a direct readout of evidence accumulation of the decision, we developed a novel method wherein we use regression patterns that identified differences in drift rate across alert and drowsy periods into the neural dynamics. We performed a cluster permutation test to identify regions in the brain where regression patterns of drowsy periods dominate over the alert periods. This revealed clusters in the early time period (70 to 330 ms) and a later (380 - 680 ms) one, matching the direct decoding results. Unsurprisingly the early cluster was not reliable in source space as evidence accumulation in the drowsy periods started later (around 420 ms from decoding analysis). However, the later cluster showed frontal electrodes in sensor space, and source space parietal regions (superior, inferior) in both the left and right hemispheres and a frontal region in the right hemisphere are associated exclusively with low alertness. This evidence suggests the processing has shifted from right hemisphere dominant (in alert periods) to both the right and left hemisphere. Such a signature of processing agrees closely with the inter-hemispheric model of competition (Kinsbourne, 1970, 1977), wherein under conditions of low alertness, there is competition between left and right hemisphere for attention. The brain thus reconfigures from being right hemisphere dominant to one involving inter-hemispheric communication as it becomes drowsy. From a decision making perspective the neural reconfiguration seems to expand to bilateral parietal sources with arousal fluctuations, and from inferior and dorsolateral prefrontal areas into the frontal pole, consistent with an increase in cognitive demand proposed by the Multiple Demand system (Duncan, 2010). Another line of evidence suggesting neural fluctuations

influencing evidence accumulation comes from Summerfield and Wyart (Wyart et al., 2012) where the phase of ongoing parietal delta oscillations impacted choice. We interpret that the neural reconfiguration occurring in the transition to lower states of consciousness (inattentive, mildly drowsy) brings slower rhythms and suppresses fast oscillatory events (disappearing alpha and beta), and that this changes the nature of the noise in the brain, forcing the cognitive system to exchange information differently.

As we lose consciousness, the neural system responsible for decision making adapts to the internal challenge of decreasing arousal and shows its resilience, exerting homeostatic regulations at the cognitive level to maintain performance. The attention and wakefulness fluctuations experienced in humans (and animals) are common during the day and they are not only dependent on the circadian and sleep-wake regulation pressure (Borbély et al., 2016), but also on genetic, epigenetic, environmental and life history factors that shape the alertness aspects of attention as well as the alertness aspect of arousal (Bekinschtein et al., 2009; Mitchell, 2020).

Concluding remarks

In this study we investigated the effects of arousal fluctuation on decision making and proposed that despite the internally generated pressure due to the conscious state change, the decision making system is flexible enough to maintain functional performance by reallocating neural resources. We defined measures of tonic alertness in pre-trial neural EEG signatures by separating each trial between awake and early drowsy arousal metastable levels, and evaluated the changes in behavioural model parameters and its relationship with neural decoding features and neural dynamics of perceptual decision-making. We found that the decision becomes more liberal (in terms of performance) as alertness decreases, showing lower sensitivity with a less steep psychophysics curve. We further show that an evidence accumulation parameter (drift rate) explains the change in the patterns of responses with decreased alertness. At the neural level we demonstrate a delay in the processing of the decision both by direct decoding of the neural dynamics, and when guided by the evidence accumulation behaviour in each trial.

Transition of consciousness in the near-awake to light-decrease of alertness is emerging as a model for internally caused interactions (Canales-Johnson et al.; Comsa et al., 2019; Noreika et al.; Song and Tagliazucchi, 2020; Tagliazucchi and Laufs, 2014). With lesion and pharmacological challenges, neuropsychology and cognitive neuroscience have tried to define the necessary and sufficient brain networks, areas and dynamics of the brain to implement cognition, revealing compensation, reconfiguration and plasticity (Adolphs, 2016; Valero-Cabré et al., 2017; Yeung et al., 2018). Semi-causal effect of the internal interference exerted by the arousal system in the cognitive process paves the way for the use of wakefulness and arousal challenges in a principled manner, adding new tools for cognitive brain research. Cognitive neuroscience uses models, correlational and causal methods to reach consensus about the underlying mechanism of thought (Krakauer et al., 2017). Here, we have followed a theoretically motivated question about perceptual decision making systems using behavioural

modelling to understand the system, but at the same time causally modulated the neural networks with arousal to uncover the mechanisms of brain function.

EXPERIMENTAL PROCEDURES

Stimuli and Protocol

Forty-one healthy subjects (no auditory, neurological or psychiatric abnormalities) participated in this study. Data from nine subjects was discarded due to a) technical problems with headphone amplifier (8) b) Not following task instructions (1). Thus only data from 32 subjects (24.46 ± 3.72 years old, 14 males) was considered for further analysis. All participants were self-reported to be right-handed and the same was also established by using Edinburgh Handedness Scale (Oldfield, 1971). Each subject had a handedness score of above 0 (right-handed) with mean 80.26 ± 23.59. Only easy sleepers (as per self report) were recruited and further they were administered with the Epworth Sleepiness scale (Johns, 1991) on the day of the experiment. 29 subjects had a sleepiness score >= 7 (classified as easy sleepers) and 3 of them had a sleepiness score >=4. All subjects were asked not to consume any stimulant like Coffee/Tea before the experiment that day. The study and the experimental protocol was approved by the cambridge psychology research ethics committee and written informed consent was provided by all subjects. A monetary compensation of £30 was provided for participation in the study.

Each subject underwent two experimental sessions a) Alert b) Drowsy.

Alert session: Subjects were presented with 124 complex harmonic tones (guitar chord) that fell on the left or right side of their veridical midline (0°) ranging from -59.31° to +59.31°. These tones were recorded using in-ear microphones in free-field (Bareham et al., 2014). Six tones from -59.31° to -39.26° were presented two times each; twelve tones from -35.24° to - 1.86° were presented four times each. A similar pattern was repeated on the right side with twelve tones from 1.86° to 35.24° presented four times each, six tones from 39.26° to 59.31° presented two times each. The tones in the midline (0°) were presented four times, resulting in a total of 124 tones. The order of tones presented was randomized per subject. Further, subjects were instructed to keep their eyes closed and respond (as quickly and as accurately as possible) with a button press (using left/right thumb) indicating the location of the tone (left or right). Each trial began after a random interval of 2-3 seconds and if the subject did not respond for 5 seconds, the next trial was started. The subjects were also instructed to stay awake throughout this session.

Drowsy session: Subjects were presented with 740 complex harmonic tones (as above) that fell on the left or right of their veridical midline (0°) again ranging from -59.31° to +59.31°. Six tones from -59.31° to -39.26° were presented twenty times each; twelve tones from -35.24° to -1.86°

were presented twenty times each. Similar pattern was repeated on the right side with twelve tones from 1.86° to 35.24° being presented twenty times each, six tones from 39.26° to 59.31° presented twenty times each. The tone in the midline (0°) was presented twenty times, resulting in a total of 740 tones. The order of tones was again randomized per subject as in the alert session. Subjects were again instructed to keep their eyes closed and respond (as quickly and as accurately as possible) with a button press (by left/right thumb) indicating the direction of the tone (left or right). Each trial began after a random interval of 4-5 seconds and if the subject did not respond for 5 seconds, the next trial was started. In this session, the subjects were allowed to fall asleep (and become drowsy) and were gently awoken if they didn't respond to more than 3 trials consecutively.

Before the start of the experiment, the subjects were allowed a practise session to familiarise with the task.

Preprocessing

EEG data was acquired with 129 Ag/AgCl electrodes (Electrical Geodesics Inc) using Cz as the reference electrode. The impedances of all electrodes were kept below 100 KΩ (to ensure higher signal to noise ratio) and data was acquired at a sampling rate of 500 Hz. EEG data was pre-processed with custom made scripts in MATLAB (MathWorks Inc. Natick, MA, USA) using EEGLAB toolbox (Delorme and Makeig, 2004). The preprocessing steps are as follows: First, the peripheral electrodes that covered the regions of forehead, cheeks and neck were removed to reduce artifacts related to eye and muscle movements, thus retaining only 92 channels that covered the scalp. Second, the data was bandpass filtered with zero phase shift between 1 and 40 Hz using hamming windowed-sinc FIR filter and further resampled to 250 Hz. Third, pre-trial and post-trial epochs per trial were created. For the pre-trial epochs, the data was epoched from -4000ms to 0ms prior to the onset of the stimuli. The pre-trial epochs were created only in the drowsy session and not in the alert session (details below). For the post-trial epochs, the data was epoched from -200ms to 800ms to the onset of the stimuli for both the alert and drowsy sessions. Fourth, the trials that exceeded the amplitude threshold of ±250 μV were removed in a semi-automatic fashion. Fifth, the bad channels were detected in a two-step fashion: a) channels are considered bad (zero activity) if channel variance is below 0.5. b) The normalized power spectrum of the remaining channels was computed and any channel that exceeded the mean power spectrum by ±3 standard deviations was marked bad. Sixth, to remove further artifacts related to eye-blink and muscle movement, independent component analysis (ICA) was performed on the channels not marked as bad in the previous step. ICA components that correspond to artifacts were rejected by manual inspection. Seventh, the bad channels were now interpolated using spherical interpolation. Eighth, the bad trials were detected again using an amplitude threshold of ±250 μV and bad electrodes (those exceeding the threshold) in such trials were interpolated in a trial-by-trial fashion. Ninth, the post-trial

epochs were re-referenced to the average of all channels (whereas the pre-trial epochs were maintained with the recorded reference - Cz).

Alertness levels

The preliminary step in both behavioural and neural analysis is to classify periods of data into 'alert' and 'drowsy'. The data from the pre-trial epochs were used to classify each trial into alert or drowsy. For the alert session, all pre-trial periods were considered to be awake as subjects were explicitly instructed to stay awake (and none of the subjects failed to respond to any of the trials in the alert session). In the drowsy session, the subjects were allowed to fall asleep and hence some trials would be considered as alert and some would be drowsy. For each trial in the drowsy session, pre-trial epochs were analysed using the micro-measures algorithm (Jagannathan et al., 2018). Each trial was classed as 'alert', 'drowsy(mild)' , 'drowsy(severe)'. Further only trials that were classified as 'drowsy(mild)' were used as drowsy trials. The other trials were excluded because usually subjects don't respond when under 'drowsy(severe)' trials. We similarly excluded the 'alert' trials in the drowsy session to instead compare the drowsy trials of the drowsy session to all of the trials in the alert session.

Behavioral analysis

Error proportion

In order to understand how the rate of errors differs across different stimuli (tones to the left or right side of midline) and how it is modulated by change in alertness levels we performed the following analysis. First, we computed the proportion of errors made by each subject under each alertness level ('awake', 'drowsy') and under each stimulus type ('left' or 'right' tone). If the total number of trials for any participant under any condition is less than 5, then the corresponding error proportion (for that condition) is ignored in the further analysis. We defined 4 different multilevel models to understand the modulation of error proportion by state of participant ('alert', 'drowsy') and stimulus ('left', 'right'). We used multilevel models, as different subjects had different levels of alertness (differing number of alert, drowsy trials). In the null model, the error proportion depends only on the mean per subject (fixed effect) and the subject id (random effect). In the second model (state model), the error proportion depends only on the state of the subject ('alert' or 'drowsy' as fixed effect) and individual subject (subject id is used as random effect). In the third model (stimulus model), the error proportion depends only on the stimulus ('left' or 'right' as fixed effect) and individual subject (subject id is used as random effect). In the fourth model (state-stimulus model), the error proportion depends on a combination of state of subject ('alert' or 'drowsy') and the stimulus ('left' or 'right'), both used as fixed effects and individual subjects (subject id is used as random effect). These 4 models were fitted using the 'lmer' function ('lmerTest' package) in R (Kuznetsova et al., 2017) and the winning model is identified as the one with the highest log-likelihood by comparing it with the

null model and performing a likelihood ratio chi-square test ($\chi 2$). Finally the top two winning models were compared against each other using 'anova' function in R (Fox and Weisberg, 2018), to validate whether the winning model (if it is more complex) is actually better than the losing model (if it is simpler). The state-stimulus model emerged as the winning model.

The different models along with their log-likelihood values are shown below:

Model comparison:

Model	**Parameters**	**Log-likelihood**	**Pr(>$\chi 2$)**
Null	Fixed: mean, Random: subject id	61.24	-
State	Fixed: state, Random: subject id	66.62	<0.001
Stimulus	Fixed: stimulus, Random: subject id	68.45	<0.001
State-Stimulus	Fixed: state*stimulus, Random: subject id	**77.98**	<0.001

Type III Analysis of Variance Table with Satterthwaite's method of the winning model (State-Stimulus):

Model elements	**Sum Sq**	**Mean Sq**	**NumDF**	**DenDF**	**F value**	**Pr(>F)**
State	0.21023	0.21023	1	95.07	14.0465	0.0003061
Stimulus	0.28140	0.28140	1	95.07	18.8018	3.607e-05
State:Stimulus	0.10312	0.10312	1	95.07	6.8897	0.0101019

The state-stimulus (winning model) was further analysed with the 'anova' function and it was found that there was a reliable main effect of both state ($p<0.001$) and stimulus ($p<0.0001$) on error proportion and also the interaction between state and stimulus also had a reliable effect ($p<0.05$) on error proportion. Estimated marginal means were computed using the 'emmeans' package in R to perform post-hoc tests using tukey adjustment for multiple comparisons.

Subjective midline shifts

The change in subjective midline was quantified by fitting psychometric functions to the responses produced by each subject under alert and drowsy conditions.

The proportion of rightward responses for each subject under each stimulus condition from -60° to +60° was fitted with a cumulative normal function using the generalized linear model 'glm' function in R. The link function used in the model was 'probit' which quickly asymptotes to the axes compared to the 'logit' function. The fit also constrained in such a way the number of trials under each stimulus angle weighed the shape of the function. The mean of such a cumulative

normal function (the point where the curve crosses 0.5 in the y-axis) is referred to as the subjective midline ('bias'). The subjective midline is the stimulus (angle) at which the subject performs at chance (0.5), which in an ideal world would be closer to the veridical midline (0°). This fitting procedure was performed for each subject separately under each condition (alert, drowsy). Examples of fits for two subjects are shown in the supplemental information (Figure s1). The standard deviation of the cumulative function (slope or steepness) represents the 'sensitivity' of the system. In general, large variations in the bias point tend to reduce the sensitivity of the system. Further we also ignored subjects (leaving 11 subjects) in the drowsy condition that had a bias point of more than 60° (as the overall stimulus angle can vary only between -60° to +60°) and standard deviation of more than 30°.

Drift-diffusion modelling

The different elements of the decision-making process can be teased apart by using the drift-diffusion model. The primary parameters of this model include
i) drift-rate - 'v' -- rate of evidence accumulation
ii) bias point - 'z' -- starting point of the decision making process
iii) boundary separation distance – 'a' -- distance between the two decision boundaries
iv) non-decision time –'Ter' -- for accounting other processes like stimulus encoding (before the start of evidence accumulation process), response execution (after the end of evidence accumulation).

We implemented the drift diffusion process using hierarchical drift diffusion model (HDDM) (Wiecki et al., 2013) (version 0.7.1) that allows for a hierarchical Bayesian procedure to estimate the model parameters. The principal reason for using such hierarchical procedures is because different subjects fall asleep in different ways (differing number of alert and drowsy trials). Using a hierarchical bayesian procedure allows for robust estimation of model parameters under such limited conditions of trials available (Zhang et al., 2016). We used the response of each subject ('left' or 'right' button press) instead of accuracy ('correct' or 'incorrect') to fit the HDDM. Such a procedure is referred to as Stimulus-coding and allows for robust estimation of 'z' without being biased by accuracy. To identify the best parameters, we identified 8 different models where in each model allows the parameters ('z','v') to vary depending on state ('alert' or 'drowsy') or stimulus ('left' or 'right'). In each model, 15000 samples from the posterior distribution were estimated using Markov chain Monte Carlo methods. 5000 samples were further discarded as burn-in to discard the effect of initial values on the posterior distribution estimation. The winning model was chosen as the one with the lowest deviance information criterion (DIC). DIC allows for computing model accuracy while penalising for model complexity (Spiegelhalter et al., 2002). Further the winning model was also checked for model convergence using the Gelman-Rubin statistic. The convergence statistic was computed for 5 different model runs and was found to be closer to 0.99 (values closer to 1 but less than 1.2 indicate convergence) (Gelman et al., 2013; Spiegelhalter et al., 2002). Posterior predictive fits for a few sample subjects are shown in the supplemental information (Figure s2). The DIC scores for different models are given below

Model	Drift-rate (v)	Bias-point (z)	DIC score
1	'state'	'state'	22516.09
2	'state'	'stim'	22474.67
3	'state'	'state','stim'	22273.48
4	'stim'	'state'	24057.42
5	'stim'	'stim'	23901.97
6	'stim'	'state','stim'	22110.07
7	'state','stim'	'state'	21713.90
8	'state','stim'	'stim'	**21700.95**

Neural analysis

Decoding

We used multivariate pattern analysis (MVPA) techniques to analyse the divergent pattern in EEG data. Decoding is a form of MVPA in which we analyse patterns of brain activity in order to predict the experimental condition (or stimuli) under which that pattern was generated. Traditional ERP (Event related potentials) analyses rely on using a-priori identified spatial locations or temporal segments in the data to measure the differences across conditions. However decoding techniques do not rely on a-priori definitions and perform much better in detecting differences across experimental conditions (Fahrenfort et al., 2018).

Temporal decoding:

Temporal decoding involves using EEG data (X) composed of size: [Electrodes x Time points x Trials] to predict the experimental condition (Y). The experimental condition predicted here was the direction of the stimuli presented e.g. left or right stimuli. The first step in the decoding analysis consists of fitting an estimator (w) to a subset of the data (X) called ($Xtrain$) to predict a subset of the experimental condition (Y) called ($Ytrain$). The second step involves using this trained estimator on another subset of the data (X) called ($Xtest$) to predict subset of the experimental condition (Y) called ($Ytest$). The third step involves evaluating the performance of this estimator using a measure (e.g. accuracy) by comparing the prediction ($\hat{Y}test$) with the actual label ($Ytest$).

Estimator construction:

First, the EEG data is subjected to a standard scaler (using StandardScaler() from scikit-learn version 0.22) that removes the mean of the data and scales it by its variance. Second, we used logistic regression to estimate the model parameters for maximally separating the different

experimental conditions. Third, we implemented temporal decoding by using the sliding estimator (using SlidingEstimator() from scikit-learn) to fit the logistic regression model per time-point.

Cross-validation:
The EEG data was first downsampled to 100 Hz and further binary classification was performed between two conditions ('left' and 'right' stimuli) separately across alert and drowsy conditions per individual subject. As the target of the classification was stimuli being presented we only considered the trials wherein the subject made the correct decision. The subject was considered for classification only if they had at least 25 trials under each condition. Further 5-fold cross validation was performed such that 4 folds were used for training and the last fold was used as a testing set.

Validation measure:
The classifier performance was evaluated using Area Under the Curve (AUC) of the receiver-operating characteristic (ROC). It is implemented using 'roc_auc' in the sliding estimator function in scikit-learn. When AUC is about 0.5 the classifier performs at chance, while the AUC score of 1 has a very good separability across classes. We computed the AUC-ROC score per subject as the average of the score across all the cross-validation folds.

Group statistics:
In order to identify the reliability of the AUC score at the group level, we performed a cluster permutation test (subjects x timepoints) using MNE version 0.19.2 (spatio_temporal_ cluster_1samp_test) (Gramfort et al., 2013). Thus producing p-values per time point at the group level, from which time points wherein we can infer those regions wherein AUC is reliably different from chance (0.5) at the group level.

Coefficients of patterns:
The parameters of the decoding model are not neurophysiologically interpretable in a straightforward way (Haufe et al., 2014). Hence the parameters of the backward model (decoding) need to be transformed to produce the forward model. This is done by obtaining the coefficients of the estimator model per subject using the 'get_coef' function from MNE('_patterns'). For performing group statistics in electrode space, we used the same cluster permutation based approach as described earlier.

Source reconstruction of patterns:
The coefficients created above can be projected into the source space to infer the brain regions involved in the pattern activity. Source reconstruction was achieved using a combination of Freesurfer (Fischl, 2012) and MNE. First, the surface was reconstructed using 'recon-all' (Freesurfer) using the default ICBM152 template for the structural magnetic resonance image (MRIs). Next, the Boundary element model (BEM) was created using 'make_watershed_bem' from MNE. Further, scalp surfaces for the different element boundaries were created using

'make_scalp_surface' from MNE. Second, we performed the registration of the scalp surface with the default EEG channel locations (with fiducials as markers) manually using 'coregistration' from MNE. Third, forward solution was computed using 'make_bem_model' from MNE with conductivity = [0.3, 0.006, 0.3]. Fourth, to test if the source reconstruction of the electrode data is accurate we projected the ERP data of a sample subject into source space and analysed data from different regions of interest to confirm its validity as shown in the supplemental information (Figure s4). The classifier patterns of each subject were then projected into source space in the following manner. First, we computed the noise covariance using the baseline data from -0.2 to 0 ms. Second, we used the forward solution (8196 vertices) and the noise covariance to create an inverse operator using 'minimum_norm.make_inverse_operator' from MNE (loose=0.2). Third, used the individual classifier pattern per subject and applied the inverse operator on it (method = MNE, SNR = 3, lambda2 = $1/(SNR)^2$) to produce the source reconstruction of the classifier patterns per subject. For performing group statistics in the source space we used the same cluster permutation based approach as described earlier.

Neuro-behavioural analysis

Regression with drift diffusion model

To identify the correlates of the evidence accumulation process, we used the model parameters generated by the drift diffusion model (winning model #8). First, the ERP data (post trial epochs) were z-scored per electrode per trial. Second, the ERP data was baseline corrected with pre-trial data from -200ms to 0ms. Third, the ERP data was averaged every 50ms per electrode per trial to create 20 time points (-200ms to 800ms) per electrode per trial. Fourth, the ERP data were entered into a regression with trial by trial estimation of the drift rate using the HDDMRegressor from the HDDM toolbox (Wiecki et al., 2013). The drift rate was allowed to vary per trial based on the ERP data per state ('alert' or 'drowsy') per stimuli ('left' or 'right'). Fifth, the traces were computed per condition (state and stimuli combination). Sixth, the differences in drift rate (between left and right stimuli) per time point per electrode were computed in both alert and drowsy condition. This difference represents the discriminability of the electrode in identifying the left and right stimuli at that time point. Thus this analysis yielded differences in drift rate per electrode, per time point, per condition, per subject. The differences can then be considered similar to the classifier patterns and can be analysed both in electrode and source spaces. Further we also computed group level differences using the cluster permutation techniques described earlier.

AUTHOR CONTRIBUTIONS

Conceptualization: S.R.J and T.A.B.
Methodology: S.R.J.
Software: S.R.J.
Investigation: S.R.J.
Formal Analysis: S.R.J.
Resources: S.R.J, C.A.B. and T.A.B.
Data Curation: S.R.J and C.A.B.
Writing – Original Draft: S.R.J.
Writing – Review & Editing: S.R.J, C.A.B, and T.A.B.
Visualization: S.R.J.
Supervision: C.A.B. and T.A.B.
Project Administration: S.R.J.
Funding Acquisition: S.R.J and T.A.B.

ACKNOWLEDGMENTS

This research was funded by the Gates Cambridge Scholarship awarded to S.R.J and the Wellcome Trust Biomedical Research Fellowship WT093811MA awarded to T.A.B. as well as small departmental funding to S.R.J. and T.A.B. We thank Annamaria Laudini and Dritan Nikolla for assistance with data collection, Andrés Canales-Johnson, Will Harrison, Valdas Noreika and other members of the Consciousness and Cognition Lab in Cambridge for their valuable comments and support.

SUPPLEMENTAL INFORMATION

Subjective midline by sigmoid fits

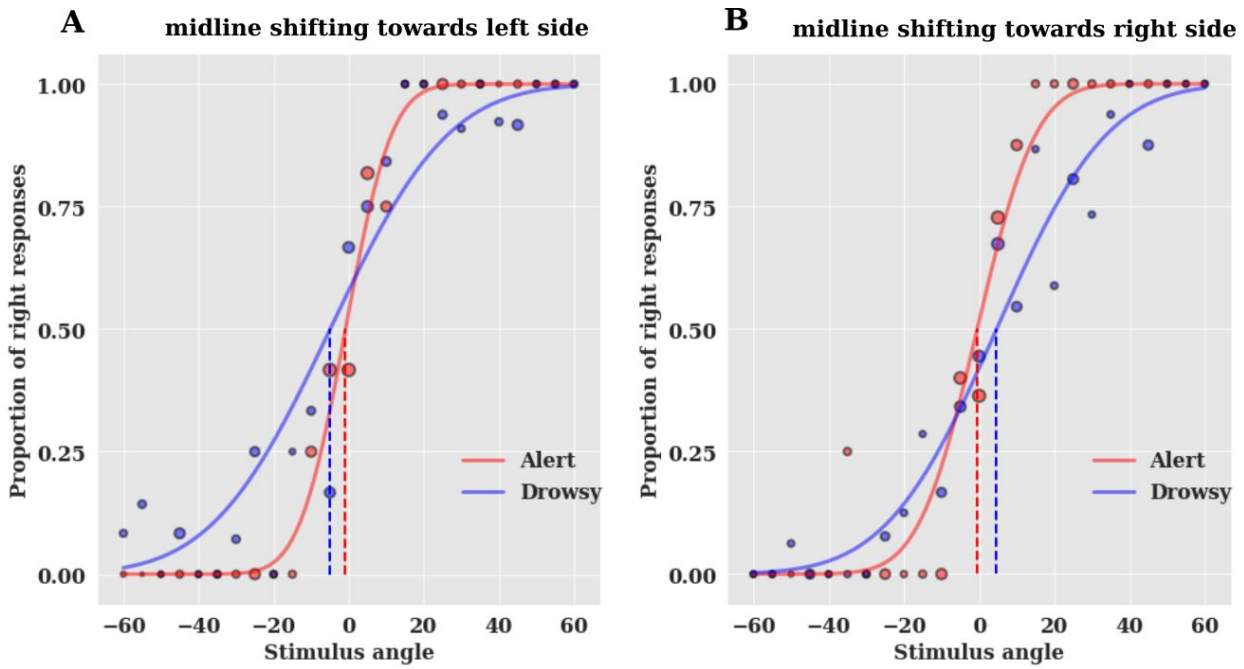

Figure s1: Example fits of two different subjects are shown. For each subject under each condition ('alert', 'drowsy'), the proportion of rightward responses were fitted to the stimulus angle varying from -60° to +60°. The 'glm' fit was constrained in such a way that the number of responses under each angle was used as weights. The size of the dots represent the normalised (per condition) number of trials under each angle. A) Here, we notice that the bias point (dotted line) shifts towards the left side as the subject becomes drowsy. In other words, as the subject becomes drowsy they overestimate the right side of space. B) Here, we notice that the bias point (dotted line) shifts towards the right side as the subject becomes drowsy. In other words, as the subject becomes drowsy they overestimate the left side of space

Posterior predictive fits from the best model

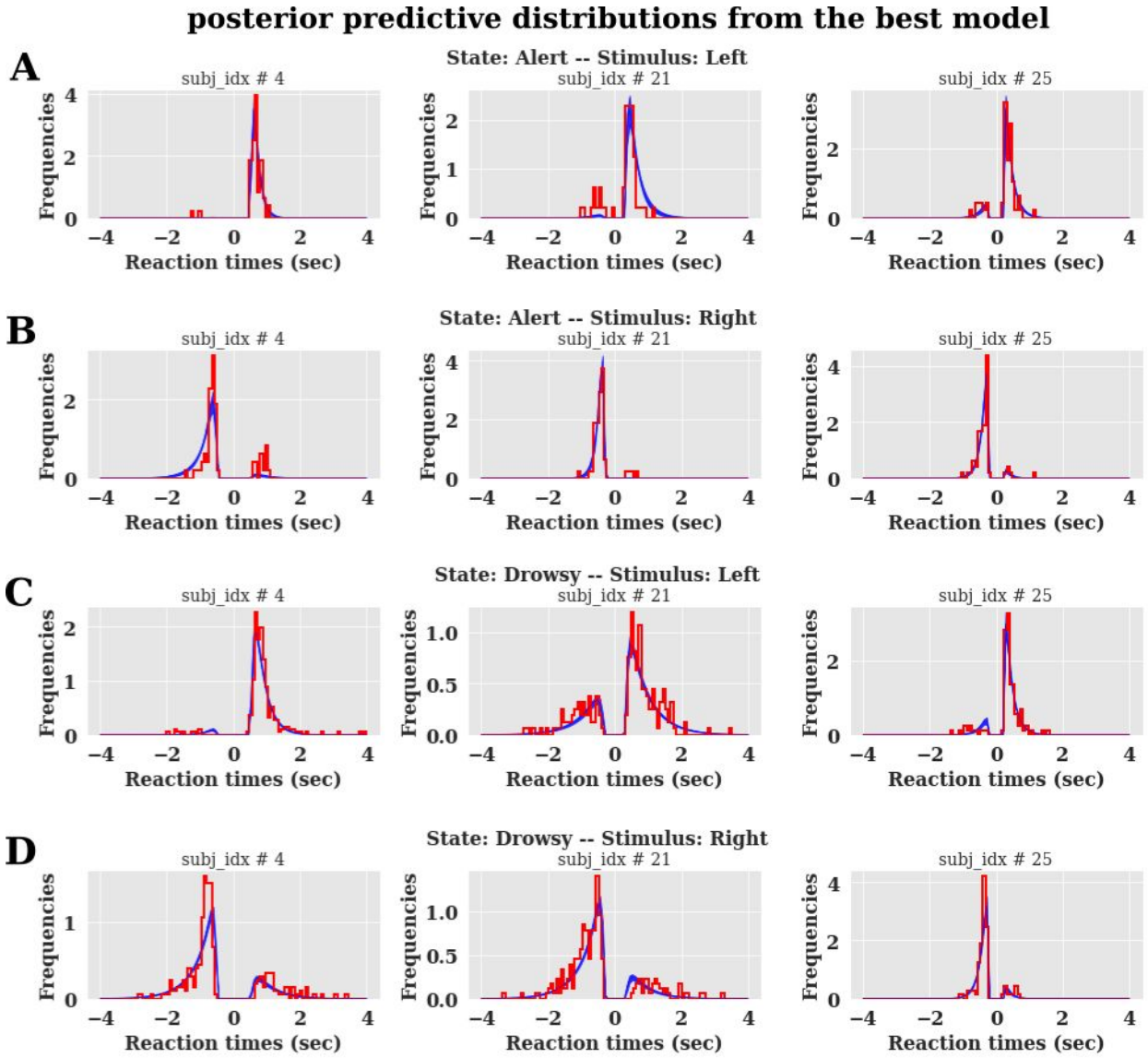

Figure s2: Posterior predictive fits for the best model (#8) in hddm for some sample subjects are shown. The x-axis in each plot represents the reaction times in seconds, the positive values are times for a 'left' button presses, whereas those on the negative x-axis are for 'right' button presses. The histograms (in red) represent distributions of the observed data, whereas the blue lines represent model prediction. A) Posterior predictive distributions when the subject is 'alert' and the stimulus presented was 'left' tone. B) Posterior predictive distributions when the subject is 'alert' and the stimulus presented was 'right' tone. C) Posterior predictive distributions when the subject is 'drowsy' and the stimulus presented was 'left' tone. D) Posterior predictive distributions when the subject is 'drowsy' and the stimulus presented was 'right' tone.

Event related potential (ERP) values at significant electrodes from classifier patterns

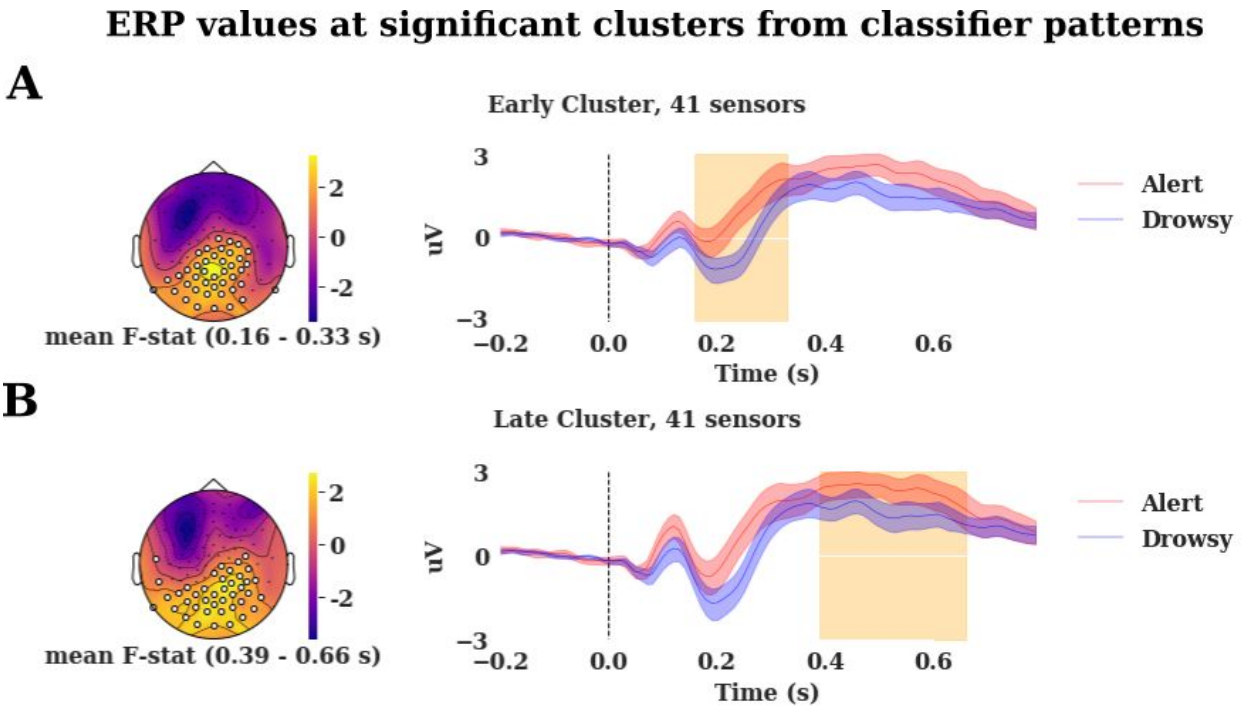

Figure s3: The spatio-temporal clustering of the classifier patterns across alert and drowsy periods reveals significant clusters at different time periods (From Figure 4 A,B). Here the significant electrodes are taken and their corresponding ERP values are plotted. The time course of the ERP values look similar to the time course of the classifier patterns itself.

Validation of source reconstruction

To verify the accuracy of the spatial localization procedure, we used the following methodology. First, we used EEG data from a sample subject (in the drowsy periods, as it has a larger number of trials) and computed noise covariance from 0.2 sec to 0 sec (pre stimulus period) using the 'shrunk' method in mne. Next the computed covariance was regularised. Second, the forward solution was computed using the transformation file, source space, beamformer solution as described in the methods section. Third, we computed the inverse operator using the noise covariance and the forward solution. Fourth, we applied the inverse operator on the raw data with parameters of snr = 12, lambda2 = 1.0 / snr ^{2}, method = 'MNE'. For validation we used the 'MNE' method instead of 'dSPM' used in the decoding analysis, as we wanted to have a positive and negative amplitude in the source signals. Further we only picked the source signals whose orientations were normal to the surface. Fifth, we used labels from the left hemisphere in the source space and extracted signals in the regions of auditory, occipital, middle temporal, and frontal pole using the mode 'mean_flip'. Sixth, we extracted

mean time courses inside each ROI and plotted the same in Figure s3. We identified N100 (peak negative signal 100 ms followed by stimulus presentation) only in the time course from the auditory region, whereas the other ROIs displayed weaker N100 amplitudes as shown below. This validates the accuracy of source localization as N100 is ERP amplitude is known to be localised in the auditory region.

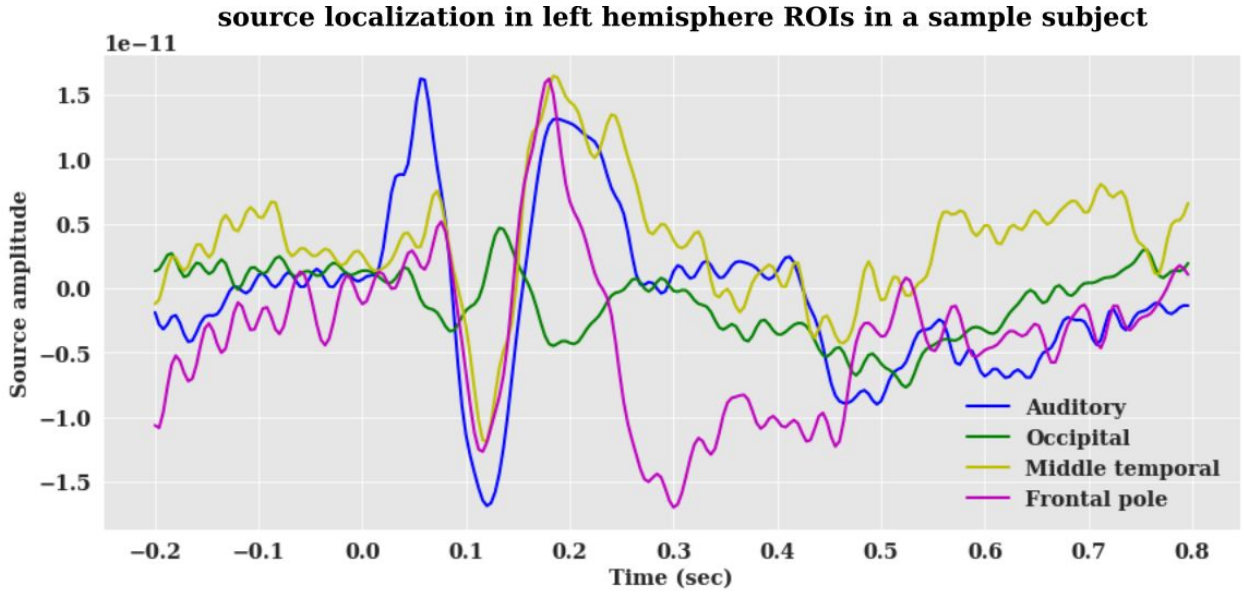

Figure s4: Source localization performed in a sample subject clearly reveals N100 presence in auditory regions at 100 ms after stimulus presentation. However occipital regions do not reveal the presence of N100 thereby validating the source localization methods used.

REFERENCES

Adolphs, R. (2016). Human Lesion Studies in the 21st Century. Neuron *90*, 1151–1153.

Bareham, C.A., Manly, T., Pustovaya, O.V., Scott, S.K., and Bekinschtein, T.A. (2014). Losing the left side of the world: rightward shift in human spatial attention with sleep onset. Sci. Rep. *4*, 5092.

Bekinschtein, T.A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. Proc. Natl. Acad. Sci. U. S. A. *106*, 1672–1677.

Benwell, C.S.Y., Harvey, M., and Thut, G. (2014). On the neural origin of pseudoneglect: EEG-correlates of shifts in line bisection performance with manipulation of line length. Neuroimage *86*, 370–380.

Borbély, A.A., Daan, S., Wirz-Justice, A., and Deboer, T. (2016). The two-process model of sleep

regulation: a reappraisal. J. Sleep Res. *25*, 131–143.

Bowen, A., McKenna, K., and Tallis, R.C. (1999). Reasons for variability in the reported rate of occurrence of unilateral spatial neglect after stroke. Stroke *30*, 1196–1202.

Buchanan, L., and O'Connell, A. (2006). A brief history of decision making. Harv. Bus. Rev. *84*, 32–41, 132.

Canales-Johnson, A., Beerendonk, L., Blain, S., Kitaoka, S., Ezquerro-Nassar, A., Nuiten, S., Fahrenfort, J., van Gaal, S., and Bekinschtein, T.A. Decreased alertness reconfigures cognitive control networks.

Comsa, I.M., Bekinschtein, T.A., and Chennu, S. (2019). Transient Topographical Dynamics of the Electroencephalogram Predict Brain Connectivity and Behavioural Responsiveness During Drowsiness. Brain Topogr. *32*, 315–331.

Corbetta, M., and Shulman, G.L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. *3*, 201–215.

Corbetta, M., and Shulman, G.L. (2011). Spatial neglect and attention networks. Annu. Rev. Neurosci. *34*, 569–599.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods *134*, 9–21.

Dietz, M.J., Friston, K.J., Mattingley, J.B., Roepstorff, A., and Garrido, M.I. (2014). Effective connectivity reveals right-hemisphere dominance in audiospatial perception: implications for models of spatial neglect. J. Neurosci. *34*, 5003–5011.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends Cogn. Sci. *14*, 172–179.

Fahrenfort, J.J., van Driel, J., van Gaal, S., and Olivers, C.N.L. (2018). From ERPs to MVPA Using the Amsterdam Decoding and Modeling Toolbox (ADAM). Front. Neurosci. *12*, 368.

Fischl, B. (2012). FreeSurfer. Neuroimage *62*, 774–781.

Fox, J., and Weisberg, S. (2018). An R Companion to Applied Regression (SAGE Publications).

de Gee, J.W., Colizoli, O., Kloosterman, N.A., Knapen, T., Nieuwenhuis, S., and Donner, T.H. (2017). Dynamic modulation of decision biases by brainstem arousal systems. Elife *6*.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2013). Bayesian Data Analysis, Third Edition (CRC Press).

Goupil, L., and Bekinschtein, T.A. (2012). Cognitive processing during the transition to sleep. Arch. Ital. Biol. *150*, 140–154.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., et al. (2013). MEG and EEG data analysis with MNE-Python.

Front. Neurosci. *7*, 267.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., and Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage *87*, 96–110.

Heekeren, H.R., Marrett, S., Bandettini, P.A., and Ungerleider, L.G. (2004). A general mechanism for perceptual decision-making in the human brain. Nature *431*, 859–862.

Heekeren, H.R., Marrett, S., Ruff, D.A., Bandettini, P.A., and Ungerleider, L.G. (2006). Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. Proc. Natl. Acad. Sci. U. S. A. *103*, 10023–10028.

Heekeren, H.R., Marrett, S., and Ungerleider, L.G. (2008). The neural systems that mediate human perceptual decision making. Nat. Rev. Neurosci. *9*, 467–479.

Ho, T.C., Brown, S., and Serences, J.T. (2009). Domain general mechanisms of perceptual decision making in human cortex. J. Neurosci. *29*, 8675–8687.

Hori, T., Hayashi, M., and Morikawa, T. (1994). Topographical EEG changes and the hypnagogic experience. Sleep Onset: Normal and Abnormal Processes. 237–253.

Hull, J.T., Wright, K.P., Jr, and Czeisler, C.A. (2003). The influence of subjective alertness and motivation on human performance independent of circadian and homeostatic regulation. J. Biol. Rhythms *18*, 329–338.

Jagannathan, S.R., Ezquerro-Nassar, A., Jachs, B., Pustovaya, O.V., Bareham, C.A., and Bekinschtein, T.A. (2018). Tracking wakefulness as it fades: Micro-measures of alertness. Neuroimage *176*, 138–151.

Jewell, G., and McCourt, M.E. (2000). Pseudoneglect: a review and meta-analysis of performance factors in line bisection tasks. Neuropsychologia *38*, 93–110.

Johns, M.W. (1991). A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep *14*, 540–545.

Karnath, H.O., and Fetter, M. (1995). Ocular space exploration in the dark and its relation to subjective and objective body orientation in neglect patients with parietal lesions. Neuropsychologia *33*, 371–377.

van Kempen, J., Loughnane, G.M., Newman, D.P., Kelly, S.P., Thiele, A., O'Connell, R.G., and Bellgrove, M.A. (2019). Behavioural and neural signatures of perceptual decision-making are modulated by pupil-linked arousal. Elife *8*.

Kinsbourne, M. (1970). The cerebral basis of lateral asymmetries in attention. Acta Psychol. *33*, 193–201.

Kinsbourne, M. (1977). Hemi-neglect and hemisphere rivalry. Adv. Neurol. *18*, 41–49.

Knowles, J.B. (1993). Sleep, Sleepiness and Performance Timothy H. Monk (Ed.) John Wiley &

Sons, 1991. Journal of Organizational Behavior *14*, 703–705.

Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., and Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. Neuron *93*, 480–490.

Kruschke, J.K. (2012). Bayesian Estimation Supersedes the t Test. PsycEXTRA Dataset.

Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software *82*.

Loughnane, G.M., Newman, D.P., Bellgrove, M.A., Lalor, E.C., Kelly, S.P., and O'Connell, R.G. (2016). Target Selection Signals Influence Perceptual Decisions by Modulating the Onset and Rate of Evidence Accumulation. Current Biology *26*, 496–502.

McGinley, M.J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C.R., Tolias, A.S., Cardin, J.A., and McCormick, D.A. (2015). Waking State: Rapid Variations Modulate Neural and Behavioral Responses. Neuron *87*, 1143–1161.

Mitchell, K.J. (2020). Innate: How the Wiring of Our Brains Shapes Who We Are (Princeton University Press).

Noreika, V., Canales-Johnson, A., Johnson, A., Arnatkevičiūtė, A., Koh, J., Chennu, S., and Bekinschtein, T.A. Wakefulness fluctuations elicit behavioural and neural reconfiguration of awareness.

Nunez, M.D., Vandekerckhove, J., and Srinivasan, R. (2017). How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters. J. Math. Psychol. *76*, 117–130.

O'Connell, R.G., Dockree, P.M., and Kelly, S.P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nat. Neurosci. *15*, 1729–1735.

Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia *9*, 97–113.

Petersen, S.E., and Posner, M.I. (2012). The attention system of the human brain: 20 years after. Annu. Rev. Neurosci. *35*, 73–89.

Posner, M.I. (2008). Measuring alertness. Ann. N. Y. Acad. Sci. *1129*, 193–199.

Ratcliff, R., Smith, P.L., Brown, S.D., and McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. Trends Cogn. Sci. *20*, 260–281.

Robertson, I.H., Mattingley, J.B., Rorden, C., and Driver, J. (1998). Phasic alerting of neglect patients overcomes their spatial deficit in visual awareness. Nature *395*, 169–172.

Roitman, J.D., and Shadlen, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J. Neurosci. *22*, 9475–9489.

Shadlen, M.N., and Newsome, W.T. (1996). Motion perception: seeing and deciding. Proc. Natl.

Acad. Sci. U. S. A. *93*, 628–633.

Shadlen, M.N., and Newsome, W.T. (2001). Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey. Journal of Neurophysiology *86*, 1916–1936.

Shulman, G.L., Pope, D.L.W., Astafiev, S.V., McAvoy, M.P., Snyder, A.Z., and Corbetta, M. (2010). Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network. J. Neurosci. *30*, 3640–3651.

Sigman, M., and Dehaene, S. (2005). Parsing a Cognitive Task: A Characterization of the Mind's Bottleneck. PLoS Biology *3*, e37.

Sigman, M., and Dehaene, S. (2008). Brain mechanisms of serial and parallel processing during dual-task performance. J. Neurosci. *28*, 7585–7598.

Smith, P.L., and Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. Psychol. Rev. *116*, 283–317.

Song, C., and Tagliazucchi, E. (2020). Linking the nature and functions of sleep: insights from multimodal imaging of the sleeping brain. Current Opinion in Physiology *15*, 29–36.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) *64*, 583–639.

Sturm, W., and Willmes, K. (2001). On the functional neuroanatomy of intrinsic and phasic alertness. Neuroimage *14*, S76–S84.

Tagliazucchi, E., and Laufs, H. (2014). Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. Neuron *82*, 695–708.

Twomey, D.M., Murphy, P.R., Kelly, S.P., and O'Connell, R.G. (2015). The classic P300 encodes a build-to-threshold decision variable. Eur. J. Neurosci. *42*, 1636–1643.

Valero-Cabré, A., Amengual, J.L., Stengel, C., Pascual-Leone, A., and Coubard, O.A. (2017). Transcranial magnetic stimulation in basic and clinical neuroscience: A comprehensive review of fundamental principles and novel insights. Neuroscience & Biobehavioral Reviews *83*, 381–404.

Vyazovskiy, V.V., Olcese, U., Hanlon, E.C., Nir, Y., Cirelli, C., and Tononi, G. (2011). Local sleep in awake rats. Nature *472*, 443–447.

Wang, C.-A., Baird, T., Huang, J., Coutinho, J.D., Brien, D.C., and Munoz, D.P. (2018). Arousal Effects on Pupil Size, Heart Rate, and Skin Conductance in an Emotional Face Task. Frontiers in Neurology *9*.

Wiecki, T.V., Sofer, I., and Frank, M.J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. Front. Neuroinform. *7*, 14.

Wyart, V., de Gardelle, V., Scholl, J., and Summerfield, C. (2012). Rhythmic fluctuations in

evidence accumulation during decision making in the human brain. Neuron *76*, 847–858.

Yeung, A.W.K., Tzvetkov, N.T., and Atanasov, A.G. (2018). When Neuroscience Meets Pharmacology: A Neuropharmacology Literature Analysis. Frontiers in Neuroscience *12*.

Zhang, J., Rittman, T., Nombela, C., Fois, A., Coyle-Gilchrist, I., Barker, R.A., Hughes, L.E., and Rowe, J.B. (2016). Different decision deficits impair response inhibition in progressive supranuclear palsy and Parkinson's disease. Brain *139*, 161–173.