Brainstem fMRI signaling of surprise across different types of deviant stimuli

Audrey Mazancieux[1]*, Franck Mauconduit[2], Alexis Amadon[2], Jan Willem de Gee[3], Tobias Donner[4], and Florent Meyniel[1]*

[1] Cognitive Neuroimaging Unit, NeuroSpin center, Institute for Life Sciences Frédéric Joliot, Fundamental Research Division, Commissariat à l'Energie Atomique et aux énergies alternatives, INSERM, Université Paris-Sud, Université, Paris-Saclay, Gif-sur-Yvette, France

[2] NeuroSpin, CEA, CNRS, BAOBAB, Université Paris-Saclay, Gif-Sur-Yvette, France

[3] Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA

[4] Section Computational Cognitive Neuroscience, Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

* Corresponding authors: Florent Meyniel (florent.meyniel@cea.fr), Audrey Mazancieux (audrey.mazancieux@cea.fr)

Summary

The ability to detect deviant stimuli is crucial to adapt our behavior. Previous work showed that infrequent (hence deviant) stimuli elicit phasic activation of the brainstem locus coeruleus (LC), which releases noradrenaline and controls central arousal. However, it is unclear whether deviance detection selectively recruits the LC or also other neuromodulatory systems related to dopamine, acetylcholine, and serotonin. It is also unclear whether deviant-related responses in those systems only track infrequent stimuli in a sequence (which can be computed with bottom-up processes), or also violations of the sequence structure (which requires higher-order processings). Here, we combined human fMRI recordings optimized for brainstem imaging with pupillometry (a peripheral marker of central arousal) to perform a mapping of deviant-related responses in subcortical structures. Participants were exposed to a "local-global" paradigm that distinguishes between deviance based on the stimulus probability and the sequence structure. fMRI responses to deviant stimuli were quite distributed, detected in the LC but also other subcortical nuclei and many cortical areas. Both types of deviance elicited responses in the pupil, LC and other neuromodulatory systems. Our results reveal that deviance detection in humans recruits several subcortical systems and generalizes across computationally different types of deviance.

Introduction

Detecting deviant stimuli (i.e. stimuli that violate some regularity) is crucial in a variety of processes, such as learning under uncertainty (Soltani & Izquierdo, 2019), interacting in a flexible manner with the environment (Devauges & Sara, 1990; Janitzky et al., 2015), and orienting behavior (Sara & Bouret, 2012). In terms of mechanisms, it seems clearly established that deviance detection triggers the phasic, brain-wide release of noradrenaline from the locus coeruleus (LC) located in the brainstem (Aston-Jones et al., 1994; Hervé-Minvielle & Sara, 1995; Poe et al., 2020; Sara et al., 1994; Sara & Bouret, 2012). This conclusion is supported by studies in non-human animals, using electrophysiological recordings of LC neurons during oddball tasks, in which frequent and rare stimuli are typically presented in a sequence. The LC responds specifically to the rare (hence deviant) stimulus (Aston-Jones et al., 1994, 1997; Foote et al., 1980; Rajkowski et al., 1994; Swick et al., 1994). Studies in humans provided converging evidence. Again in oddball tasks, the fMRI signal in the LC region increased after deviant stimuli (Krebs et al., 2018; Murphy et al., 2014) and propranolol (a blocker of the noradrenergic beta receptors) decreased fMRI signals in cortical regions that respond to deviant stimuli, supporting the idea that noradrenaline is released in cortex after deviance detection (Strange & Dolan, 2007).

This body of work leaves unclear at least two aspects of the implication of the LC in deviance detection. A first aspect is about anatomical specificity: is the deviance-related response specific to the LC or shared across multiple other subcortical structures, notably neuromodulatory centers? The latter seems likely because deviance detection overlaps with other notions like novelty (Debener et al., 2005; Schomaker & Meeter, 2015), salience (Harsay et al., 2012), and unexpectedness (Reichardt et al., 2020; Verleger & Śmigasiewicz, 2016) that are known to implicate other neuromodulators (Bassareo et al., 2002; Brown et al., 2015; Bunzeck & Düzel, 2006; Caldenhove et al., 2017; Glimcher, 2011; Pessiglione et al., 2006). For instance, dopamine encodes unexpected stimuli in the form of reward prediction error (Glimcher, 2011; Pessiglione et al., 2006) as well as salient stimuli related to novelty (Bassareo et al., 2002; Bunzeck & Düzel, 2006; Kutlu et al., 2021). Establishing a mapping between the different neuromodulators (notably, noradrenaline, dopamine, serotonin, acetylcholine) and their functions remains a challenging goal (Dayan, 2012) with major clinical implications since many neurological and psychiatric disorders are related to dysfunctional neuromodulation, e.g., in major depression (Dunlop & Nemeroff, 2007), Parkinson's disease (Goldstein et al., 2008), attention disorders (Frank et al., 2007).

There are already several indications from previous studies that the deviance-related LC response is embedded in a larger set of neuromodulatory responses. For instance, the P300 event-related potential (ERP) was used as an indicator of LC activity supported by

photoactivation studies in rats (Vazey et al., 2018), and it was found to be larger for deviant stimuli (Nieuwenhuis et al., 2005). However, this deviant-related P300 response was also found to be reduced following the administration of either scopolamine (a cholinergic antagonist) in rats (Ahnaou et al., 2018) and humans (Caldenhove et al., 2017) or clonidine (a noradrenaline alpha receptor agonist) in humans (Brown et al., 2015), suggesting that other neuromodulators such as acetylcholine are recruited by deviance detection. Pupillometry has also often been used as an indirect marker of the LC activity. The existence of pupil responses to deviant stimuli is clearly established (Gilzenrat et al., 2010; Murphy et al., 2011; Nieuwenhuis et al., 2011; Preuschoff et al., 2011). However, a change in pupil size is not necessarily due to a change in LC activity (Megemont et al., 2022) because other subcortical nuclei like the inferior colliculi (de Gee et al., 2017; Joshi & Gold, 2020) and neuromodulators are at play such as acetylcholine from the basal forebrain (Reimer et al., 2016) and more indirectly serotonin from the raphe nucleus (Cazettes et al., 2021).

Here, we propose to measure deviant-related responses not only in the LC but also in other structures of the brainstem, notably in neuromodulatory centers. Direct, concurrent electrophysiological recording of multiple neuromodulatory centers is very difficult in non-human animals (Varazzani et al., 2015) and is not an option in humans. In contrast, fMRI can provide a complete coverage of the brainstem (and beyond). However, brainstem fMRI is particularly challenging due to the presence of larger physiological noise compared to cortex and to the very small size of the structures of interest, such as the LC (Liu et al., 2017). Most fMRI studies that measured LC activity used anatomical atlases but this method is imprecise (Astafiev et al., 2010). We thus used fMRI methods optimized for the brainstem (de Gee et al., 2017; Krebs et al., 2018) and delineated the LC (noradrenaline), the substantia nigra / ventral tegmental (SN/VTA; dopamine) and the superior and inferior colliculi (involved in pupil size and auditory processing (Ayala & Malmierca, 2012; Joshi & Gold, 2020; Malmierca et al., 2009)) based on the participant's anatomy. Using atlases, we also included for comparison the activity of other neuromodulatory regions that are more difficult to delineate individually (the basal forebrain (BF) for acetylcholine, and the Raphe nucleus (RN) for serotonin) as well as other subcortical and cortical areas.

A second aspect of the LC deviant-related response remains poorly characterized: its computational specificity. The studies mentioned so far used oddball (or similar) tasks (Aston-Jones et al., 1994, 1997; Foote et al., 1980; Krebs et al., 2018; Murphy et al., 2014; Rajkowski et al., 1994; Strange & Dolan, 2007), in which the deviance is characterized by the probability of items in a sequence. Item probability is only one and the most basic aspect of sequence knowledge in humans and other animals (Dehaene et al., 2015). Sequence knowledge also includes elements of structure, such as the existence of algebraic patterns in the sequence.

The neural mechanisms for detecting deviance based on item probability or sequence structure are likely to be very different. Bottom-up mechanisms such as stimulus-specific adaptation (i.e., reduction in neuronal firing rates for repeated stimuli) suffice to detect probability-deviant stimuli (Joshi & Gold, 2020); this mechanism is for instance at play in the inferior colliculi during auditory oddball tasks (Ayala & Malmierca, 2012; Malmierca et al., 2009). The detection of structure-deviant stimuli could require different mechanisms such as predictive coding (Heilbron & Chait, 2018) and even awareness (King et al., 2013; Strauss et al., 2015).

Here, we propose to use the local-global paradigm (Bekinschtein et al., 2009) to dissociates two types of deviance (based on item probability and sequence structure) and test whether deviant-related responses in the LC (and in comparison, other subcortical regions) are specific to the item probability or also reflect the sequence structure.

Results

Distinguishing between different types of mechanism for deviance detection

We investigated the effect of deviant stimuli using the local-global paradigm. This paradigm presents sequences of stimuli that exhibit two nested levels of structure. At the local level, stimuli form patterns of 4 sounds with a fifth one that is either identical, forming a *locally standard* pattern xxxxx, or different, forming a *locally deviant* pattern xxxxY, see Figure 1A. The global level is characterized by the succession of patterns separated by short pauses: in each block of the task, one pattern is frequent (80% of patterns) while the other is rare (20%). The local and global properties, as well as sound identity were crossed in a full factorial, 2x2x2 design. In each block, participants were first familiarized with the frequent pattern only (which defines the block identity: xxxxx block or xxxxY block) and then presented with a few rare patterns (called global deviants) interleaved among the frequent ones. Twenty participants performed the task in the scanner where they had to count the number of rare patterns for a total of 4 sessions.

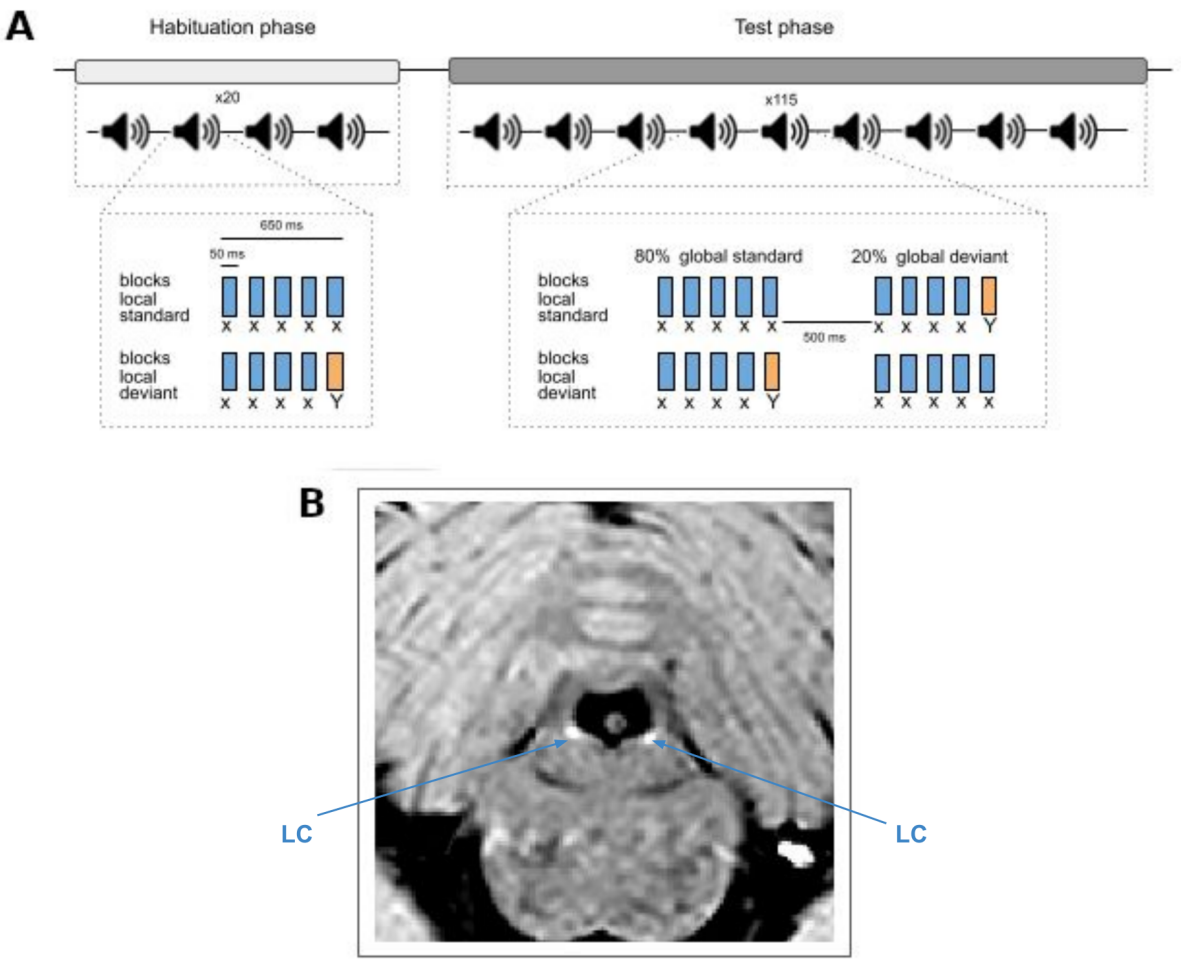

Figure 1. Task and example of anatomical images. A) The local-global paradigm. Patterns presented are either 5 identical tones (locally standard, xxxxx) or 4 identical tones and a different one (locally deviant, xxxxY). During the habituation phase, only the frequent pattern (either locally deviant or locally standard) is presented. During the test phase, this frequent pattern is presented with the rare pattern in a sequence. The sound x could be the low-pitch tone (as illustrated here) or the high-pitch tone (not shown, counterbalanced across blocks). B) Example slices of the anatomical Turbo Spin Echo (TSE) image used to delineate the LC in each participant. As the sequence is neuromelanin-sensitive, both structures are brighter.

Note that there is a key difference between the two types of global deviants. A rare xxxxY can be detected among frequent xxxxx based on low-level processes that operate on the probability of the sounds themselves in a sequence, like stimulus-specific adaptation, because the Y sound is extremely rare (occuring only with probability 0.2*⅕=0.04). In contrast, such low-level processes do not suffice to detect a rare xxxxx among frequent xxxxY because the x being the dominant sound (occurring with probability 0.8*1+0.2*⅘=0.96), xxxxY stands out more than xxxxx based on sound probability alone. A mechanism for the detection of a rare

xxxxx must operate at a higher level, namely the sequence of sound patterns. The local-global paradigm thus allows a distinction between different computations for deviance detection, operating on the stimulus probability and sequence structure respectively, and possibly different mechanisms, such as bottom-up and top-down processes, see Discussion.

Robust responses to global deviants in the pupil-linked arousal system

Pupil size is controlled by the autonomic nervous system. It provides a marker of arousal that is known to transiently increase when deviant stimuli are detected (Quirins et al., 2018). To characterize its response to global deviance, we performed a baseline-corrected, epoch-based analysis to isolate the phasic evoked response. This analysis included only a subset of 12 participants who have a large-enough number of trials after artifact rejection (see Method). Pupil size exhibited a clear response to global deviants, with larger pupil size for rare patterns compared to frequent patterns (t_{max} = 6.56, p_{max} < 0.001, d_{max} = 1.89, cluster p_{FWE} < 0.001), see Figure 2A. Note that response was remarkably similar between the two types of global deviants (there was no effect of local deviance, t_{max} = 1.55, p_{max} = 0.149; and or interaction between local and global deviances, t_{max} = 1.49, p_{max} = 0.164).

Global deviance transiently increases LC activity

Having established that both types of global deviants in the local-global paradigm elicit strong transient activation of the pupil-linked arousal system, we next investigated their effects in our primary region of interest (ROI), namely the LC. As for the pupil, we used epoch-based analyses. We extracted time series from the LC, removed several confounds (see methods), epoched the signal (from -2 to 12 s around the onset of patterns), and corrected for the baseline (-2 to 0 s). This analysis allowed us to track the fMRI activity of the LC in response to deviant stimuli with no assumption about the shape of the hemodynamic response, which may be different in subcortical structures compared to the cortex where the canonical hemodynamic responses has been defined (Friston et al., 2007). In addition, as for the pupil size analyses, the baseline correction removes the autocorrelation that may exist in the signal before and after the pattern onset, and captures the change in activity evoked by the pattern. Thus, this analysis captures the phasic activity of the LC and removes the tonic activity.

The LC activity showed a main effect of global deviance (t_{max} = 2.64, p_{max} = 0.016, d_{max} = 0.59, cluster p_{FWE} = 0.037) with a greater increase in fMRI activity for rare patterns compared to frequent patterns, without effect of the local deviance (t_{max} = 1.31, p_{max} = 0.206), see Figure 2B. The interaction between the local and global deviance effects formed two small clusters that did not reach significance after correction (p_{FWE} = 0.460 and p_{FWE} = 0.433) suggesting that responses were similar between the two types of global deviants.

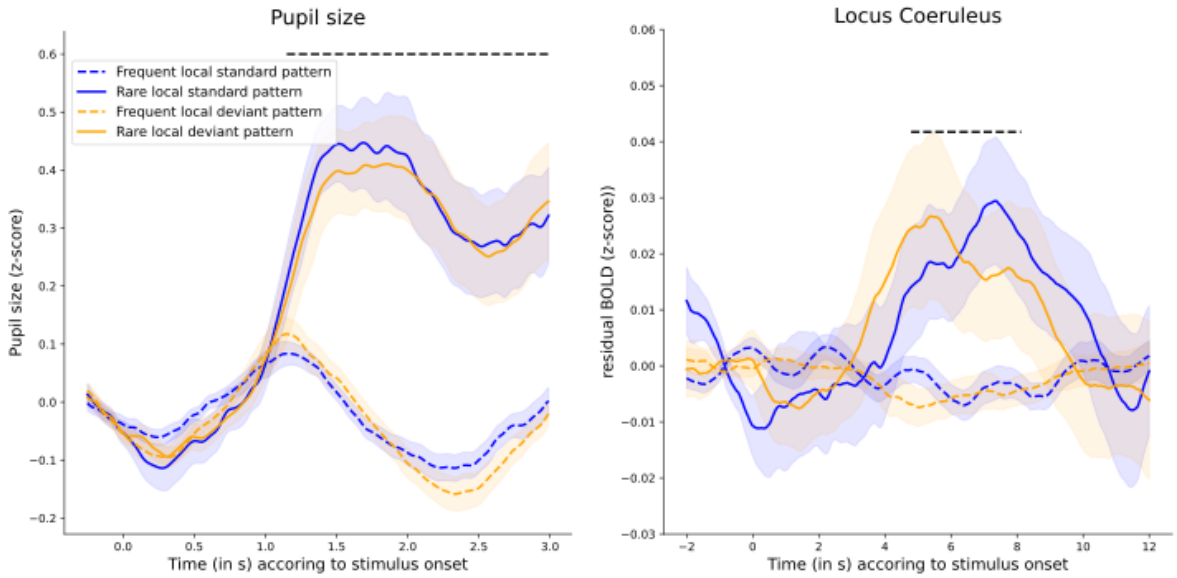

Figure 2. Epoch-based analyses of the 4 stimuli for pupil and LC data. A) Pupil size (z-score) evoked by the 4 types of patterns. Error shading is standard error. Black dashed lines indicate significant clusters for the global effect and the interaction between global and local effects, respectively (p_{FWE}<0.05). B) Time course of fMRI activity (z-score) in the LC evoked by the 4 types of pattern. Error shading is standard error. Black dashed lines indicate significant clusters for the effect of rare patterns (global effect, p_{FWE}<0.05).

Mapping of deviance-related responses across regions of interest

We repeated the epoch-based analysis to quantify deviant-related responses in several other brain structures sorted into three categories. 1) The other neuromodulator nuclei included the SN/VTA (individually delineated using Turbo-Spin-Echo anatomical images similarly to the LC, see Figure S5), the BF and the RN (based on anatomical atlases). 2) the other subcortical structures included the superior colliculi (involved in orienting responses), the inferior colliculi (involved in auditory processing), and the hippocampus (involved in sequence processing). 3) Cortical structures, where the fMRI signal-to-noise ratio is higher than in subcortical structures, included the anterior and posterior superior temporal gyri (previously identified to respond to rare patterns with the same task, see (Bekinschtein et al., 2009)), the primary auditory and visual cortices, and the ventral medial prefrontal cortex (which is part of the default mode network and thus not expected to respond to rare patterns).

Figure 3 shows the 4 types of stimuli for each ROI and Table 1 summarizes the corresponding statistics. All neuromodulator nuclei showed an increase in fMRI activity for rare patterns compared to frequent patterns (which remained significant when corrected for multiple comparisons across time, except in the BF). The dopaminergic SN/VTA is the structure

that elicits the largest response. For the LC, the maximum effect size was smaller than in the SN/VTA and the RN, probably due to the smaller size of the LC ROI. The ventral medial prefrontal cortex and the hippocampus showed the reverse pattern later in the time window: fMRI activity in these regions was higher for frequent stimuli compared to rare stimuli. The other structures (cortical and subcortical) all exhibited a larger response for rare patterns compared to frequent patterns. Larger responses were found in cortical ROIs than in sub-cortical areas, potentially due to their different sizes and signal-to-noise ratios. The shape of the response also differed between cortical and subcortical regions: cortical regions had a more canonical response with a clear peak while several subcortical regions exhibit a kind of plateau (see Figure 3). Note that a strict comparison between all these structures cannot be entirely made as they slightly differ conceptually and in terms of methodology. First cortical and subcortical structure differ in terms of SNR. Second, within subcortical structures, 4 have been defined in native space and others using atlas render the latest less neuroanatomically accurate.

No ROI showed a main effect for local deviance. Interaction between the local and global effect was significant only for cortical areas involved in auditory processing, namely the primary auditory cortex (t_{max} = 7.19, p_{max} < 0.001, d_{max} = 1.61, cluster p_{FWE} < 0.001) and the anterior (t_{max} = 7.17, p_{max} < 0.001, d_{max} = 1.60, cluster p_{FWE} = 0.001) and posterior superior giry (t_{max} = 5.21, p_{max} < 0.001, d_{max} = 1.16, cluster p_{FWE} = 0.012). In these regions, the rare xxxxY pattern (the stimulus-probability deviant) elicited a higher and earlier response than the rare xxxxx pattern (the structure deviant). We estimated this delay on the group-average responses. The three 3 cortical auditory processing regions showed a significant difference in the peak of the response between the two types of rare patterns: 0.887 s (95% confidence interval: 0.347, 1.415) in the primary auditory cortex, 1.139 (95% confidence interval: 0.734, 1.734) in the anterior superior temporal gyrus and 0.998 (95% confidence interval: 0.483, 1.902) in the posterior superior temporal gyrus. In contrast, no difference in the timing of the peaks was observed in the primary visual cortex (mean difference: 0.001: 95% confidence interval: -0.912, 1.080).

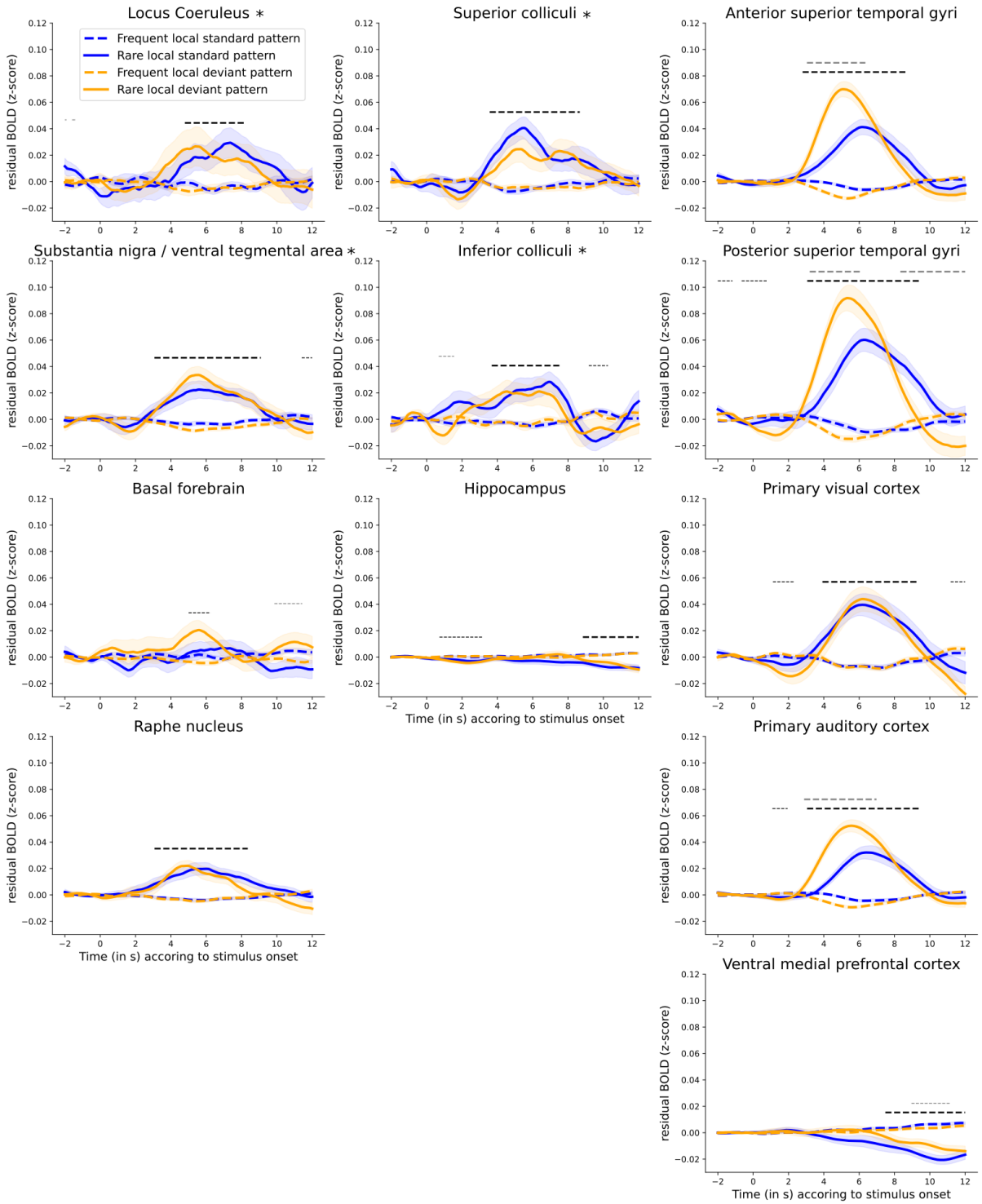

Figure 3. Time course of fMRI activity (z-score) evoked by the 4 types of patterns. The first column shows neuromodulator nuclei, the second column other subcortical ROIs, and the third column cortical ROIs. Stars indicate ROI defined in native space by manual delineation. Error shading is standard error. Bold black dashed lines indicate significant clusters for the effect of rare patterns (global effect) and bold gray dashed lines indicate clusters for the interaction between the global effect and the local effect ($p_{FWE}<0.05$).

Neuromodulation-related nuclei	Other subcortical nuclei	Cortical areas
LC t_{max} = 2.64 p_{max} = 0.016 d_{max} = 0.59 cluster p_{FWE} = 0.037	**Superior colliculi** t_{max} = 4.34 p_{max} < 0.001 d_{max} = 0.97 cluster p_{FWE} = 0.005	**Anterior superior temporal gyri** t_{max} = 10.36 p_{max} = < 0.001 d_{max} = 2.31 cluster p_{FWE} = < 0.001
SN/VTA t_{max} = 5.53 p_{max} < 0.001 dz-$_{max}$ = 1.24 cluster p_{FWE} < 0.001	**Inferior colliculi** t_{max} = 3.14 p_{max} = 0.005 d_{max} = 0.70 cluster p_{FWE} =0 .008	**Posterior superior temporal gyri** t_{max} = 9.00 p_{max} = < 0.001 d_{max} = 2.01 cluster p_{FWE} = <0 .001
BF t_{max} = 2.34 p_{max} = 0.030 d_{max} = 0.52 cluster p_{FWE} = 0.153	**Hippocampus** t_{max} = -2.10 p_{max} = 0.049 d_{max} = -0.47 cluster p_{FWE} = 0.033	**Primary visual cortex** t_{max} = 5.97 p_{max} = < 0.001 d_{max} = 1.34 cluster p_{FWE} = < 0.001
RN t_{max} = 5.14 p_{max} < 0.001 d_{max} = 1.15 cluster p_{FWE} = 0.001		**Primary auditory cortex** t_{max} = 10.04 p_{max} = < 0.001 d_{max} = 2.24 cluster p_{FWE} = < 0.001
		Ventral medial prefrontal cortex t_{max} = -2.10 p_{max} = 0.049 d_{max} = -0.47 cluster p_{FWE} = 0.004

Table 1. Statistics for the global effect (max t-values, max p-values, max Cohen's d, and p-values for clusters with FWE correction) for all ROIs.

To rule out that the detection of global deviant-related responses depends on the specifics of our analysis approach, we compared frequent and rare patterns across brain structures using Finite Impulse Response (FIR) analyses and General Linear Model (GLM) analyses (see Supplementary results). Epoch-based analyses do not model the potential superposition of effects of the current and previous patterns in the time window of interest. In contrast, FIR and GLM analyses are designed to model this superposition, and differ in their assumptions about the hemodynamic response (which is unconstrained or assumed to be

canonical, respectively). The epoch-based analysis also contains a baseline-correction that aims to suppress endogenous fluctuations in the signal - therefore focusing on phasic activity - which are ignored by the FIR and GLM analysis. Those three analyses are thus complementary.

Overall, results were qualitatively consistent across the three types of analyses suggesting that the global effect does not depend on the type of analyses we performed. The only notable qualitative differences concern the hippocampus and the ventral medial prefrontal cortex where the late negative effect of global deviance found in epoch-based analyses (and FIR analyses) differed from the positive (non-significant) effect was found with GLM analyses, probably because the late difference is not well captured by the canonical response function.

Anatomical specificity of the response to global deviants around the in the LC region

The global deviant-related response is very much distributed in cortical and subcortical structures, raising the concern that the effect found in the LC may not be specific to this region but instead widespread within the pons. To test for the anatomical specificity of global deviance within the pons, we repeated the same epoch-based analyses but after shifting the ROI corresponding to the LC in space. We shifted this ROI toward the anterior part of the pons (from 1 to 5 voxels, i.e. +2 to +10 mm) and toward the posterior part, which falls in the fourth ventricle (from 1 to 3 voxels, i.e. -2 to -6 mm). This axis is more relevant for the shift of the LC ROI than shifting toward left or right because the LC is a bilateral structure, and more relevant than toward the spinal cord or the midbrain because the LC is elongated along this axis.

Figure 4A-B shows the time course of the difference in fMRI signals between rare patterns and frequent patterns, for different shifts of the LC ROI. The effect of global deviance overall decreased as larger shifts were applied: the maximum difference between rare and frequent pattern changed from 0.055 (sd = 0.023) originally to 0.058 (sd = 0.023), 0.047 (sd = 0.025), 0.045 (sd = 0.024), 0.024 (sd = 0.022), and 0.034 (sd = 0.021) in the anterior direction (+2 mm to +10 mm shifts) and to 0.049 (sd = 0.024), 0.043 (sd = 0.020), 0.027 (sd = 0.024) in the posterior direction (-2 mm to -6 mm shift). Only the signal for the actual LC (no shift) and the signal for the +2 mm shift showed significant clusters for the effect of global deviance (no shift: $p_{FWE} = .038$; +2 mm shift: $p_{FWE} = .008$). Direct comparisons of the global deviance in shifted and unshifted data showed time points with significant differences ($p < 0.05$ one-sided test, for shifts of +8, -4 and -6 mm, but the corresponding cluster p_{FWE} remained > 0.05).

We also performed GLM analysis for the effect of global deviant stimuli. Significant voxels in the brainstem showed a small recovery with the LC atlas (see Figure 4B) which is likely due to fact whole brain analyses are more sensitive to cortical effects due to a higher SNR in the cortex. However, other voxels in the LC do not exhibit strong effects either. Those results support an anatomical specificity of the effect of global deviance in the LC region compared to its vicinity, but without sharp boundaries.

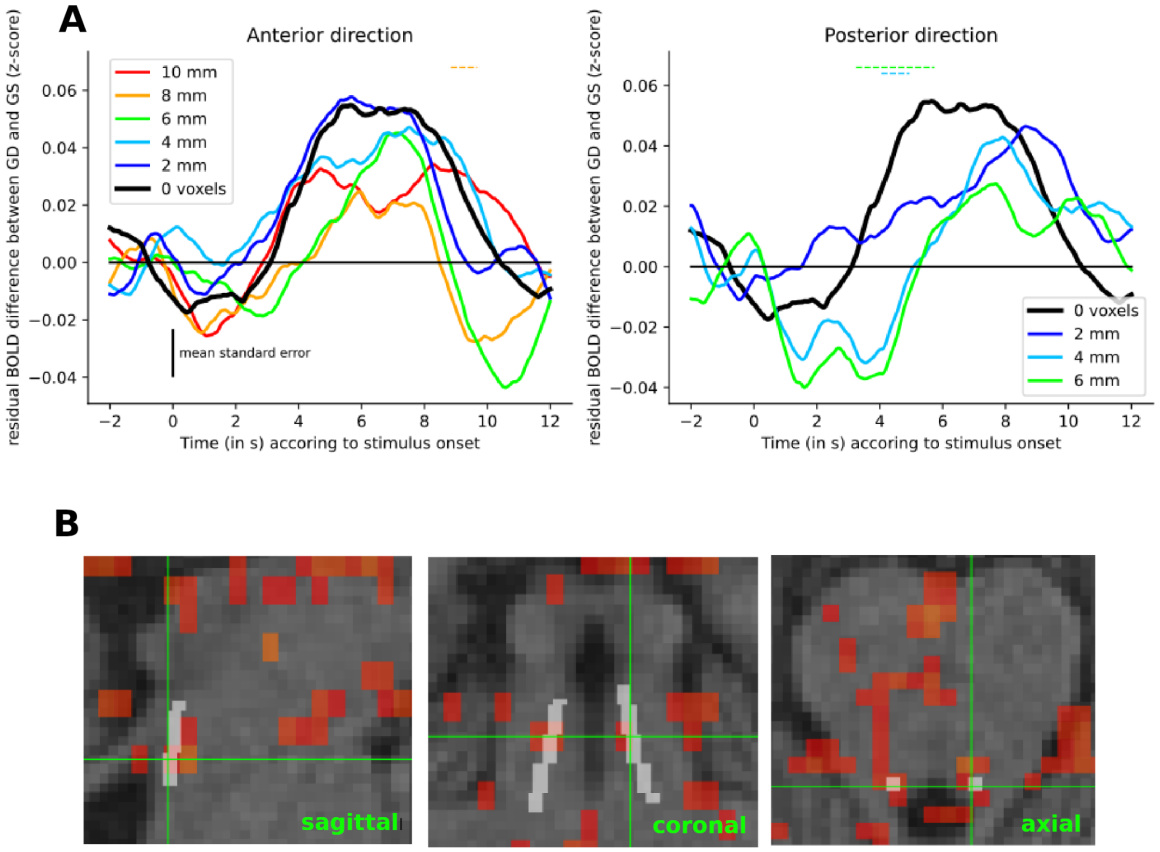

Figure 4. Anatomical specificity around the LC in the pons. A) Time course of the effect of global deviance (difference in z-scored fMRI activity between rare and frequent patterns) for different shifts of the anatomically-defined LC ROI in millimeters (2 mm corresponds to 1 voxel). Left panel refers to shifts toward the anterior direction of the pons. Right panel refers to shifts toward the posterior direction of the pons. Black line refers to the original, non-shifted LC ROI. Colored horizontal dashed lines refer to identified clusters for the difference between the corresponding color and the black line. None of them remains significant after correction for multiple comparisons (one-sided tests). B) Statistical z-map for the effect of global deviant stimuli, thresholded at z = 2.8 (corresponding to p = 0.005). White voxels correspond to the LC atlas from keren et al. (2015).

Comparison of LC activity in native space and atlas

The anatomical specificity of the effect of global deviance in the LC region can also be assessed by comparing the results obtained with anatomical delineation of the LC in each subject (in native space) to the expectedly less accurate ones obtained from a probabilistic atlas of the LC (in standardized space). We extracted fMRI time series based on an atlas of the LC (see Method section) that identified 10 voxels in standardized space. Note that our delineation in native space identified only a mean of 5.6 voxels for the LC (min = 4 voxels, max = 9 voxels,

across participants). A comparison of the two approaches revealed that the voxels identified in native space were systematically in the more anterior part of the atlas of the LC, thus, closer to the midbrain, which is consistent with previous studies (Keren et al., 2009). Therefore, to allow a fair comparison with the individual delineations (Figure 2), we also performed a second analysis matched in voxel number (using 6 voxels in the altas that were the most anterior). The two atlas-based analyses showed significant effects of global deviance (full atlas: $t_{max} = 4.34$, $p_{max} < .001$, $d_{max} = 0.97$, cluster $p_{FWE} = 0.001$; atlas with 6 voxels: $t_{max} = 4.00$, $p_{max} < 0.001$, $d_{max} = 0.90$, cluster $p_{FWE} < 0.001$). No significant cluster was identified for the effect of local deviance but the interaction between local deviance and global deviance was significant in both analyses (full atlas: $t_{max} = 4.17$, $p_{max} < .001$, $d_{max} = 0.93$, cluster $p_{FWE} = .026$; atlas with 6 voxels: $t_{max} = 2.93$, $p_{max} = .009$, $d_{max} = 0.66$, cluster $p_{FWE} = 0.022$). In contrast to the analysis performed after individual delineation of the LC, a region corresponding to probabilistic atlas exhibits a response to global deviance that is mostly driven by stimulus probability (rare xxxxY patterns), for which low-level deviance detection mechanisms suffice.

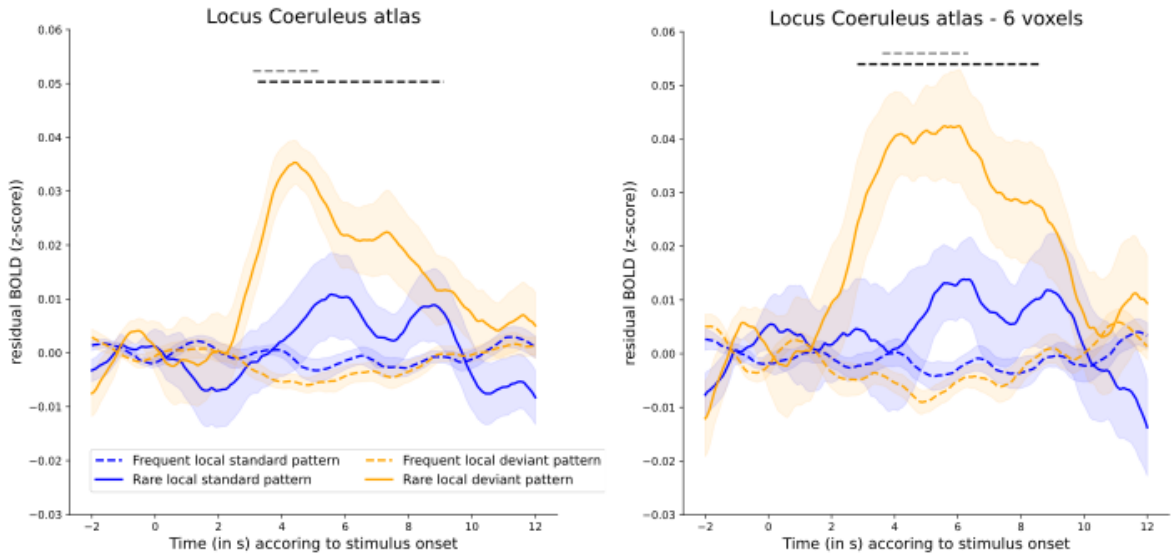

Figure 5. Anatomical specificity assessed with a normalized atlas of the LC. Time course of fMRI activity (z-score) in the LC evoked by the 4 types of patterns based on an anatomical atlas (left) and a selection of 6 voxels in the most anterior part of this atlas (right). Error shading is standard error. Black and gray dashed lines indicate significant clusters for the global effect and the interaction between global and local effects, respectively ($p_{FWE}<0.05$).

Comparison of responses to global deviants in pupil response and brain activity

Pupil response to auditory stimuli is controlled directly and indirectly by various neuromodulators (Joshi & Gold, 2020). We tested for stimuli-evoked fMRI responses in our

various ROIs and stimuli-evoked pupil responses to global deviants. Epochs containing a rare pattern were sorted in 2 conditions: large or small pupil response (median of values averaged baseline-corrected pupil size within 1.15-3 s, which is when a significant effect of the global deviant was detected). Figure 6 shows the time course of the effect of global deviance (difference between frequent and rare patterns) in all ROIs conditioned on pupil response. Note that this analysis is based on only 12 participants, with a small number of trials per condition due to artifact rejections (mean=63.4, min=29, max=88), statistical power is thus low.

Among all ROIs, only the SN/VTA showed a strong difference in responses to global deviance between epochs with large and small pupil response ($t_{max} = 4.34$, $p_{max} < 0.001$, $d_{max} = 0.97$, cluster $p_{FWE} = 0.016$). Other clusters of difference were detected in the anterior superior temporal gyrus, the primary auditory cortex, the inferior colliculi and the RN, but they did not reach significance after correction for multiple comparisons across time

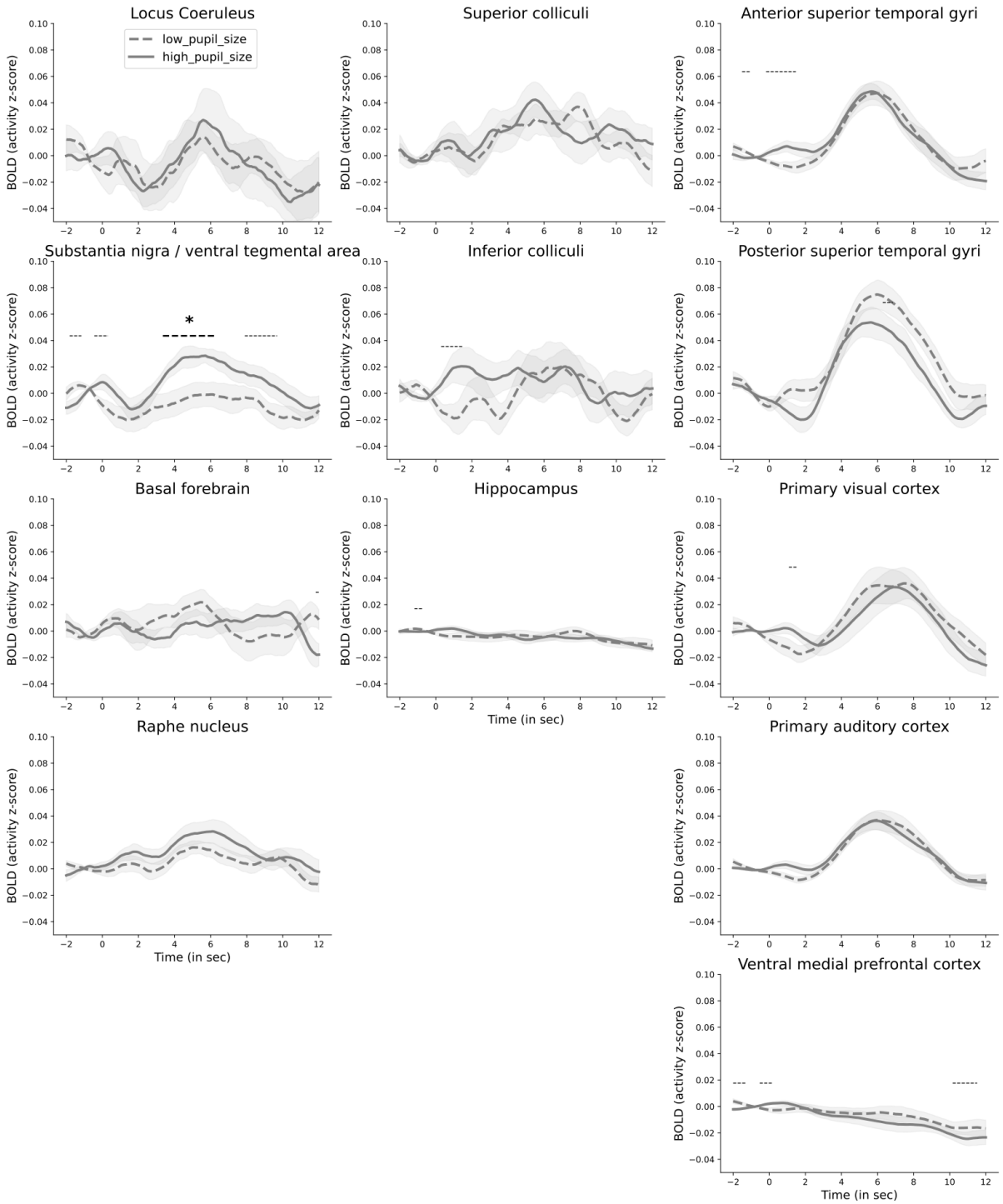

Figure 6. Time course of the effect of global deviance (z-scored fMRI activity elicited by rare), sorted by pupil response, in all ROIs. The first column refers to neuromodulator nuclei, the second column refers to other subcortical ROIs, and the third column refers to cortical ROIs. Horizontal dashed black lines correspond to clusters of significant difference ($p < 0.05$) and bold dashed black lines with a star to significant clusters after FWE correction ($p_{FWE}<0.05$).

Discussion

We performed a systematic response mapping in subcortical structures using fMRI coupled with pupillometry in a task that involves two types of deviants that requires computations based on the stimulus probability and sequence structure respectively. Global deviance evoked transient LC responses, which was our primary region of interest since it is well established that the central noradrenergic system vigorously responds to deviant stimuli (Aston-Jones et al., 1994, 1997; Rajkowski et al., 1994). Similar responses were found in other neuromodulatory centers (the SN/VTA and the RN), other subcortical nuclei (the superior and the inferior colliculi), and cortical regions (the anterior and posterior superior temporal gyri, the primary auditory and visual cortices). Local and global deviances interacted in cortical responses related to auditory processing where global deviants elicited stronger and earlier responses when corresponding to a stimulus-probability deviant than a structure deviant. In contrast, subcortical structures (and the visual cortex) exhibited similar responses to both types of global deviants. This response to both types of global deviants showed some anatomical specificity to the LC region within the pons, because it decreased when moving away from the subject-specific, anatomically defined LC region, and became driven by stimulus-probability deviants when using a probabilistic, normalized atlas of the the LC. Pupil size exhibited similar responses to both types of global deviants. When sorting global deviants by the magnitude of pupil response, only the SN/VTA showed a larger fMRI response when pupil response was larger

The local-global task allowed us to investigate a hierarchy of processes implicated in deviance detection. Previous studies showed with electrophysiology that this hierarchy of processes recruits an anatomically defined set of hierarchically organized cortical regions. Local deviance (xxxxY vs. xxxxx) elicits an early response in sensory cortices whereas global deviance (rare vs. frequent patterns) elicit a later response that is distributed across brain areas and reaches the frontal lobe both in humans (Bekinschtein et al., 2009; Dürschmid et al., 2016; Karoui et al., 2014; Wacongne et al., 2011) and macaques (Chao et al., 2018). The effects of local deviance and global deviance are propagated across cortical areas through different frequency bands, the gamma band and beta-alpha band respectively (Chao et al., 2018; Karoui et al., 2014) which are distinct functional markers of bottom-up and top-down processes (Bastos et al., 2015; Siegel et al., 2012; Wang, 2010). Our results focused on a comparison of the two types of global deviance and showed that rare patterns elicited stronger and earlier responses when they corresponded to the xxxxY patterns (stimulus probability deviant) than to the xxxxx pattern (structure deviant) in regions of the temporal lobe related to auditory processing, consistent with the idea that the detection of a rare xxxxx pattern recruits top-down processes (unfortunately, our partial brain coverage excluded most of the prefrontal cortex). We note that the distinction between global deviance based on stimulus probability and sequence structure is

not tested in several previous studies (Bekinschtein et al., 2009; Karoui et al., 2014; Quirins et al., 2018), leaving unclear whether the global effect analyzed in those studies is driven by both types of global deviants, or just one.

Other auditory paradigms can be used to discriminate different mechanisms for the detection of deviant stimuli and to map them to an anatomically defined hierarchy of cortical-subcortical regions. For instance, the contrast between standard and deviant sounds in an oddball task, and the same sound inserted in a sequence of ascending or descending sounds (which thus becomes entirely predictable) makes it possible to adjudicate between stimulus-specific adaptation (also referred to as repetition suppression) and prediction error (Lesicko et al., 2022; Parras et al., 2017). Using such a paradigm, stimulus specific adaptation and prediction error were found to dominate in the inferior colliculi and auditory cortex respectively, which are also respectively lower and higher in the anatomical hierarchy (Parras et al., 2017). Stimulus specific adaptation in the inferior colliculi of rodents was preserved when inactivating the auditory cortex, by cooling (Anderson & Malmierca, 2013) and to a lower extent by optogenetic stimulation (Lesicko et al., 2022), indicating that stimulus specific adaptation does not require cortical feedback in the inferior colliculi. In contrast, prediction error signals in the inferior colliculi were reduced when the auditory cortex was inactivated (Anderson & Malmierca, 2013; Lesicko et al., 2022), suggesting that there exists an anatomical and functional hierarchy of deviance detection mechanisms and that deviants detected by more complex mechanisms in the cortex can be fed back to subcortical structures.

In subcortical structures, in the pupil, and in the primary visual cortex, responses to both types of rare patterns were largely similar, in contrast to the temporal cortex. Determining the source and target of neural activity in those cortico-subcortical networks would be valuable but would require better time-resolved techniques than fMRI, like electrophysiology, which is technically difficult to obtain. A possible explanation for the widespread responses to rare sound patterns, which even includes the primary visual cortex here, is that the activation of the LC, which projects to the entire cortex and various subcortical structures (including the superior colliculi, RN and SN/VTA), arouses almost the entire brain (Sara & Bouret, 2012). Previous studies showed that spiking and fMRI activity in sensory cortices, including the visual cortex (Burlingham et al., 2022) tightly reflects arousal (Safaai et al., 2015; Stringer et al., 2019).

If the LC plays a central role in arousing the brain in response to deviant stimuli, then it would be valuable to determine the afferents that signal the deviance to the LC. Studies in rodents and in primates showed that afferent LC inputs mainly come from subcortical nuclei and that in cortex, only the prefrontal region directly projects to the LC (Arnsten & Goldman-Rakic, 1984; Luppi et al., 1995). Subcortical structures seem to be suitable candidates to signal the presence of rare xxxxY patterns, but as discussed above, they seem to lack the mechanisms to

detect rare xxxxx patterns. The detection of the latter seems to rely on higher-order regions like the prefrontal cortex, which could directly signal those types to deviants to the LC. This is consistent with our timing analyses in auditory processing regions where the rare xxxxY pattern elicits an earlier response than the rare xxxxx deviant.

In the local-global task, the increase in central arousal that follows rare patterns depends on the participants state of consciousness and their awareness of a sequence structure. Previous studies showed that the global deviance detection vanishes in patients with disorders of consciousness (Faugeras et al., 2012), when healthy subjects fall asleep (Strauss et al., 2015), and when they are not aware (or do not pay attention to) the task structure (Quirins et al., 2018). Interestingly, the effect of global deviance (notably rare xxxxx patterns) is more difficult to detect, and with a reduced extent, in brain recordings of macaque monkeys (Bellet et al., 2021; Chao et al., 2018; Jiang et al., 2022; Uhrig et al., 2014), for which global deviants are not behavioral relevant and thus potentially not attended, compared to healthy human participants who are told about the existence of global deviants and often asked to count them (Bekinschtein et al., 2009; Karoui et al., 2014; Quirins et al., 2018; Strauss et al., 2015; Wacongne et al., 2011). Here, we also asked participants to count the global deviants, which probably enhanced their detection and the associated brain responses.

LC responses are also known to depend on attentional effects. During active oddball tasks, LC neurons exhibited a higher response when monkeys correctly (vs. incorrectly) detected rare stimuli (Rajkowski et al., 1994). More generally, there are state-dependent changes in tonic LC activity: higher tonic activity coincides with periods when monkeys have more motor activity that is irrelevant to the task; in contrast, periods of drowsiness, immobility, and eye-closure reduce LC activity (Rajkowski et al., 1994). Thus, phasic LC responses to deviant stimuli may occur for a particular level of tonic LC activity, when the subject is sufficiently focused (and not too much) on the current task (Aston-Jones & Cohen, 2005; Dubois et al., 2021). In this study, we reported baseline-corrected analyses to focus on the phasic component and suppress the additive effect of tonic fluctuations (but ignoring potential non-linear effects previously reported in spiking activity of LC neurons (Aston-Jones & Cohen, 2005) and in pupil size (Knapen et al., 2016)). In contrast, the non-baseline corrected analyses (FIR and GLM analyses) also focus on the phasic component but ignore the tonic fluctuations (the ones above 1/128 Hz that remain after preprocessing). We note that our results are consistent across baseline and non-baseline corrected analyses, which suggests in retrospect that the detection of rare patterns that manifests itself in an increased arousal of various brain structures was not masked by fast (above 1/128 Hz) fluctuations of tonic arousal levels (which would have penalized the non-baseline corrected analyses).

The current work will in addition be of methodological interest to people interested in the measure of LC activity with fMRI and more indirectly with pupillometry. The possibility to estimate the LC activity with fMRI is a contentious issue; doing so requires dedicated methods such as the subject-specific anatomical delineation of the LC, e.g. see technical comment (Astafiev et al., 2010; Eckert et al., 2010). We report a comparison of results obtained with a subject-specific delineation and a probabilistic, normalized anatomical atlas of the LC. Although both analyses showed an effect of global deviant patterns, this effect interacted with the local pattern type and was actually driven by the rare xxxxY pattern when using the altas (there is no such interaction when using the subject-specific delineation, or in pupil size). Given what is observed in other brain regions, the pupil and previous work on the LC, we assume that the LC responds to both types of global deviants, and thus that the results obtained with subject-specific delineation are closer to the ground truth. In other words, using a subject-specific delineation (rather than an atlas) seems necessary in fMRI studies of the LC, despite being time and resource consuming. We also propose that the effect of the rare xxxxx patterns could be a quality check of a correct identification of the LC region, possibly in a trimmed down version of the local-global paradigm that only presents rare xxxxx among frequent xxxxY. Note that at our resolution (3x2x2 mm^3), the subject-specific delineation leads only to a few fMRI voxels putatively corresponding to the LC, with high variability across subjects (from 4 to 9 voxels), suggesting that a large enough number of participants is needed.

Our results are also informative concerning the use of peripheral arousal (measured as non luminance-based change in pupil size) as an approximation of central arousal (more precisely, LC activity). The sensitivity of pupil size to LC activity is demonstrated based on direct LC recording in non-human animals (Costa & Rudebeck, 2016; Reimer et al., 2016). However, those studies also demonstrated that this correlation is not specific to the LC activity, but also related to central acetylcholine (Reimer et al., 2016) and serotonin (Cazettes et al., 2021) levels. A consequence of this lack of specificity is that changes in pupil dilation may not reflect changes in LC activity (Megemont et al., 2022). Here, we found an effect of the global deviance, without interaction with the local deviance, in both pupil size and fMRI activity in the LC region, suggesting that peripheral and central arousals are similar. Note that those similar effects could arise from the LC influencing pupil size (e.g. via the intermediolateral cell column, the Edinger-Westphal nucleus or the superior colliculi notably, (Joshi & Gold, 2020)), or from a common input (e.g. the nucleus gigantocellularis that activates both the LC and the autonomic system (Sara & Bouret, 2012)). Those two hypotheses would have been supported by correlated responses to global deviance in the pupil and the LC region, but we did not find such a relationship, which was significant only between pupil size and fMRI activity in the SN/VTA region. This null result in the LC is not evidence for the absence of a relationship, notably

because our analyses were limited by the small number of included trials and participants. The result found for the SN/VTA region (which replicates the one from (de Gee et al., 2017) in a different task) could simply be due to better data quality, this region being much larger than the LC (here, all effects were stronger in the SN/VTA than in the LC). The SN/VTA has no direct connection to the systems controlling pupil size (Joshi & Gold, 2020) but it receives direct input from the LC (Sara & Bouret, 2012); the effect found in the SN/VTA could thus be due to an effect in the LC that our data failed to detect.

Overall, the current study showed an effect of deviance that generalizes across two types (stimulus probability and sequence structure) in many subcortical regions, including neuromodulatory centers, and several cortical regions. Our results are consistent with the idea that deviant sound patterns engage the arousal system in the brain through the activity of the LC. The LC likely gets prediction error input from higher regions (e.g. frontal areas) and in turn broadcasts surprise signals across the entire brain. Future work with better temporal resolution will need to determine the direction of neural signals between the interconnected neuromodulatory centers, other subcortical structures, and cortical areas that subtend a hierarchy of deviance mechanisms.

Acknowledgments

This work was funded by the ANR #18-CE37-0010-01 CONFI-LEARN and by the ERC StG NEURAL-PROB #947105 to FM. The MRI facility was also partly supported by the European Union (FEDER-2007–2013, agreement #91–2015–004).

Competing interests

There are no competing interests in this study.

Materials and methods

Participants

Twenty participants (9 females) aged between 20 and 36 years (mean = 27.00, SD = 4.60) were enrolled in the experiment. This protocol was approved by a national ethics committee (Comité de Protection des Personnes Ile de France 3, approval #2018-A03195-50). Participants receive monetary compensation for their participation (80€ for 2 hours). They were right-handed based on self-report and had normal or corrected-to-normal vision.

Stimuli and task

The task included 4 sessions of 11 minutes each and was run using Octave (version 6.1.0) and the Psychtoolbox (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007) in the scanner. It was the same task as used in Bekinschtein et al. (2009). Stimuli are short auditory tones composed by 3 sinusoidal tones resulting in either a low-pitched sound (stimulus A composed by 350, 700, and 1400 Hz sinusoides) or a high-pitched sound (stimulus B: 500 Hz, 1000 Hz, and 2000 Hz sinusoides). Stimuli were presented in a sequence of patterns separated by pauses. A pattern consisted in four identical tones and a fifth that could be either the same (xxxxx; within-pattern standard, i.e. *local standard*) or different (xxxxY; within-pattern deviant, i.e. *local deviant*). The assignment of tones and patterns were counterbalanced across blocks (block of AAAAA and AAAAB vs. BBBBB and BBBBA). The duration of each tone was 50 ms and pattern duration was 650 ms with an inter-pattern interval of 500ms. During the habituation phase, participants were first exposed to only one pattern. During the test phase, participants were presented with either the same pattern as during habituation in 80% of the cases (*frequent pattern*) or with the other pattern in 20% of the cases (*rare pattern*). Figure 1A depicts a schematic representation of the task

Each session included 2 blocks in counterbalanced order: one where the habituation pattern was a local standard pattern (denoted xxxxx block) and one where it was the local deviant pattern (denoted xxxxY block). One block included 135 patterns (22 rare patterns and 113 frequent patterns including the habituation phase). During the task participants had to count the number of rare patterns.

MRI data collection and preprocessing

MRI data were acquired on a 3 Tesla scanner (Siemens, Prisma) with a 64-channel coil. In order to maximize the signal-to-noise ratio in LC, we acquired partial-brain functional echo planar images (EPI) images centered on the brainstem and oriented perpendicular to the floor of the fourth ventricle (and thus, main axis of the LC). We used the following parameters: TR =

1.25 s, TE = 30 ms, flip angle = 65°, 28 interleaved slices with a slice thickness of 3 mm and a multiband factor of 2. In-plane resolution was 2.0x2.0 mm. The encoding phase direction was from anterior to posterior. To estimate distortions, we acquired two volumes with opposite phase encoding directions. One volume was in the anterior to posterior direction (AP) and the other was in the other direction (PA), with TR = 4,800 ms, TE = 54 ms.

Two partial-brain Turbo Spin Echo (TSE) structural images, sometimes referred to as neuromelanin-sensitive (Chen et al., 2014; Sasaki et al., 2006) were acquired: one centered on the LC (e.g., DeGee et al., 2017; Keren et al., 2015) and others centered on the SN/VTA. Images were acquired with an in-plane resolution of 0.7x0.7 mm and reconstructed at 0.35x0.35 (TR = 675 ms, TE = 12 ms). We acquired 14 slices per TSE, slice thickness was 2 mm, oriented perpendicular to the floor of the fourth ventricle. We also acquired a whole-brain structural T1 image with an MPRAGE sequence for anatomical co-registration and the delineation of the IC and the SC with in-plane resolution of 1x1 mm and a slice thickness of 1 mm (TR = 2,300 ms, TE = 2.98 ms).

All preprocessing steps relied on SPM12 (Wellcome Trust Center for Neuroimaging, University College London) except the TOPUP correction that relied on FSL, using the python/FSL and python/SPM interfaces afforded by Nipype (https://doi.org/10.5281/zenodo.596855). Slice-timing correction was referenced to the middle of each TR. Volumes were realigned onto the first volume of each session, and then onto the first volume of the first session. We also performed a TOPUP correction that estimates the susceptibility field using the AP/PA volumes and unwraps EPI images. Different coregistrations were made for different types of analyses. For those in native space analyses, EPI images were coregistered with the TSE images (either with the one centered on the LC to extract LC data, or the one centered on the SN/VTA to extract SN/VTA data) or with the T1 image (to extract IC and SC data). For normalized space analyses, the T1 image was first coregistered to the TSE image before normalization performed using the standard SPM template in the Montreal Neurological Institute (MNI) space.

Physiological data collection and preprocessing

During the task, we recorded cardiac rhythm with a pulse oximeter and respiration with a belt. We modelled physiological signals using FSL PNM (Brooks et al., 2013) that creates physiological regressors for each slice of each volume. We selected estimates for the reference slice used in the slice-timing correction. We defined orders for each component as follows: 4 for the cardiac component, 3 for the respiratory component, and 1 for the interaction between the two. The total number of regressors modeling physiological signals was 18. One participant had no physiological recordings due to technical issues.

Pupil size data collection and preprocessing

Pupil size was also recorded during scanning using an MRI-compatible EyeLink 1000 system. On raw data we performed the following preprocessing steps: (1) add a margin of 50 ms before and after the blinks detected by the EyeLink system, (2) interpolate the signal linearly within each blink, (3) low-pass filter (5 Hz) the data, (4) epoch the data within -0.5 to 3 s relative to each stimulus onset, (5) exclude epochs with a total blink duration exceeding 20% of the data. It is difficult to measure pupil size in the MRI scanner due to the distance between the eyes and the camera, the use of a mirror, and the partial occlusion by the antenna around the participant's head. We excluded 8 participants for whom pupil size data was available on less than 20% of epochs. The number of participants included in the analyses related to pupil size was therefore 12.

Definition of Regions of Interest (ROIs) and preprocessing

We delineated by hand for each participant ROIs in native space using the TSE images for the LC and SN/VTA (see Figure 1B for an example for one participant), and the fourth ventricle) and the T1 image (for the IC and the SC). All masks were resampled to match the EPI resolution resulting in a probabilistic mask that was then transformed into a binary mask. Threshold probability of being part of the ROI was 0.05. We extracted time series from the EPI images using these masks. Anatomical landmark for the BF, the RN and to a lesser extent the hippocampus are less reliable in TSE and T1 images, thus, we used anatomical atlases in normalized space (maps from Zaborszky et al., 2008 for the BF; the Harvard Ascending Arousal Network atlas from Edlow et al., 2012 for the RN; the Harvard-Oxford cortical and subcortical structural atlases in FSL for the hippocampus). For comparisons between native and normalized space, we also used an anatomical atlas for the LC (Keren et al., 2015). Finally we selected 5 cortical ROIs: the anterior and posterior superior temporal gyri (aSTG and pSTG), the primary visual and auditory areas (V1 and A1), and the ventro-medial prefrontal cortex (vmPCF). The aSTG and pSTG masks came from the Harvard-Oxford cortical and subcortical structural atlases in FSL, A1 and V1 from the Talairach atlas in FSL (Lancaster et al., 2000; 2007), and the vmPFC from the following identifier from NeuroVault (https://identifiers.org/neurovault.image:18650). For each ROI, we preprocessed the signal by high-pass filtering (1/128 Hz).

Epoch-based analyses of fMRI signals

We performed epoch-based analyses on fMRI time-series extracted from each ROI. We first linearly regressed out potential confounding variables (movement parameters, the time-series extracted from the fourth ventricle4, and physiological regressors), and z-scored the residual signal per session. This signal was then upsampled (factor 1000, linear interpolation) and data was then epoched around each stimulus onset (time window: -2 s to 12) for each

participant. Then, the baseline signal was subtracted from each epoch using a time window of -2 s to 0 s.

Finite Impulse Response (FIR) analyses

FIR analyzes model a number of successive post-stimulus time steps that allow to take into account stimuli that are presented to the participant during the time window of interest, controlling for potential superposition of effects. As for epoch-based analyses, the predefined time-window was from 0 sec to 12 sec around the onset of patterns and we added additional regressors (movement parameters, the V4 time-series, and physiological regressors) in our model. For these analyses, the fMRI signal was upsampled with a factor of 5. FIR analyses make no assumptions about the hemodynamic response. We only modelled the effect of rare patterns. At the group level, we tested whether the parameter estimates for these patterns differed from 0 by using a one sample cluster permutation test (cluster-forming and cluster-level alphas of 0.05, two-tailed tests, 10,000 permutations).

Generalized Linear Model (GLM) analyses

As for FIR analyses, GLM-based analyses control for potential superposition of effects but assume the hemodynamic response to be canonical. One GLM was estimated on time-series per ROIs. The design matrix included the 4 types of patterns convolved with the canonical hemodynamic response function (HRF in SPM) as well as additional regressors corresponding to movement parameters, time series in the fourth ventricle, and physiological regressors. Parameters (betas) were estimated at the subject level with an auto-regressive AR(1) model. We then computed the difference in parameter estimates between rare and frequent patterns, and tested for its significance (against 0) at the group-level using a t-test.

Correction for multiple comparisons across time

As epoch-based analyses (of fMRI signals and pupil size) and FIR analyses require multiple comparisons across peri-stimulus times, family-wise error (FWE) correction for multiple comparisons was computed using a cluster-based permutation test (cluster-forming and cluster-level alphas of 0.05, two-tailed tests, 10,000 permutations) with the 'mne' package in Python (Gramfort et al., 2013).

Supplementary results

Effect of global deviance detection in all ROIs

Figure S1 shows the group-level effect of global deviance for each ROI (corresponding statistics can be found in Table 1 in the manuscript). All regions showed an increase in fMRI activity for rare patterns compared to frequent patterns (which remained significant when corrected for multiple comparisons across time) except for the BF, the hippocampus, and the ventral medial prefrontal cortex.

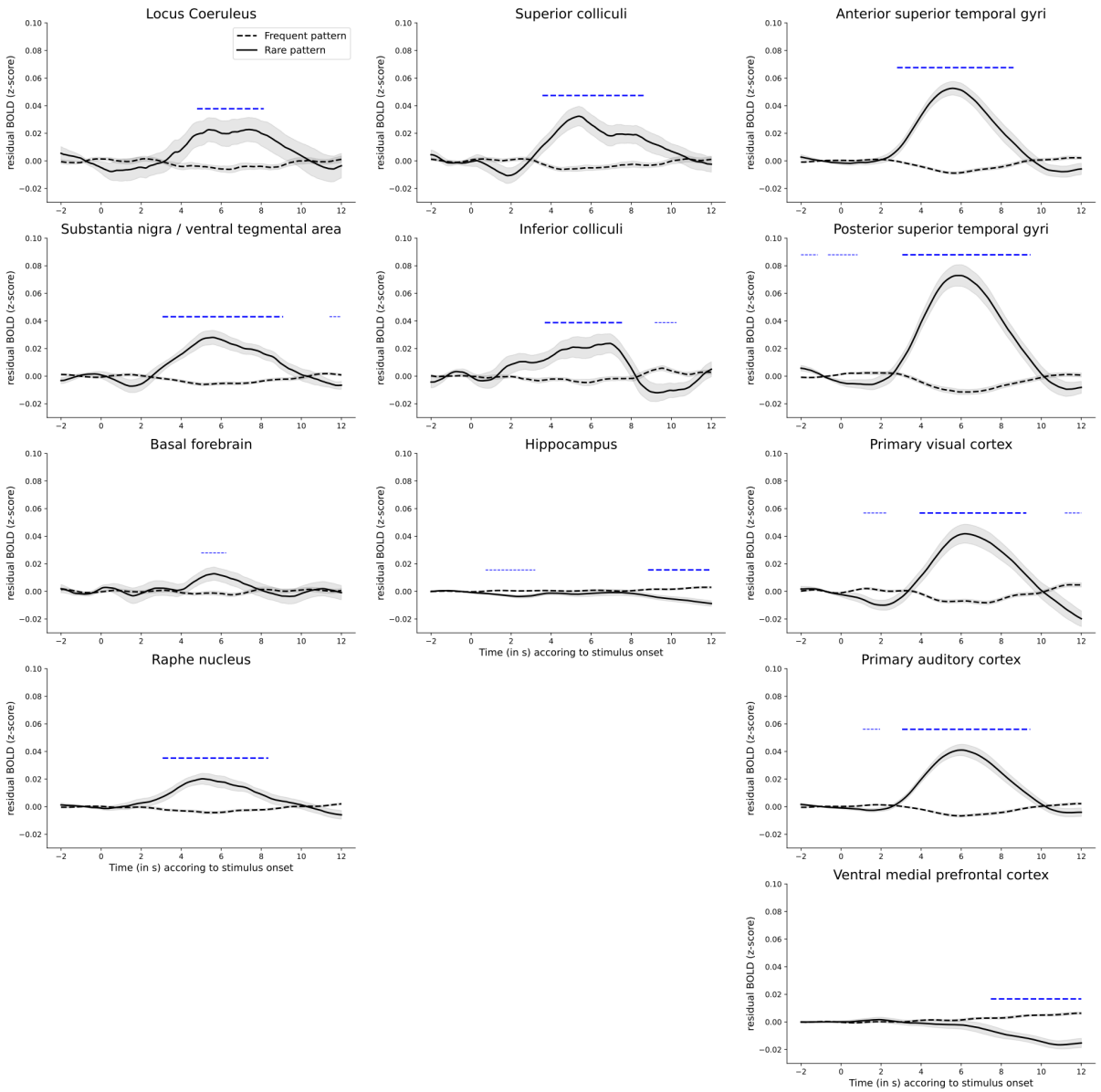

Figure S1. Time course of fMRI activity (z-score) evoked by rare patterns (bold line) and frequent patterns (dashed-line) in different brain structures. The first column shows neuromodulator nuclei, the second column other subcortical ROIs, and the third column cortical

ROIs. Horizontal dashed blue lines correspond to clusters of significant difference ($p < 0.05$) and bold dashed blue lines to significant clusters after FWE correction ($p_{FWE} < 0.05$).

Effect of global deviance using FIR analyses

Figure S2 shows the group-level effect of global deviance for each ROI and Table S1 summarizes the corresponding statistics. As for epoch-based analyses, all neuromodulator nuclei showed an increase in fMRI activity for rare patterns. As for epoch-based analyses, only the hippocampus and the ventral medial prefrontal cortex showed no effect of global deviant pattern (no difference from 0). Note that for these analyses, the significant cluster identified for the LC did not reach significance ($p_{FWE} = 0.067$) when corrected for multiple comparisons across time.

Neuromodulation-related nuclei	Other subcortical nuclei	Cortical areas
LC $t_{max} = 3.48$ $p_{max} = 0.003$ $d_{max} = 0.78$ cluster $p_{FWE} = 0.067$	**Superior colliculi** $t_{max} = 4.91$ $p_{max} < 0.001$ $d_{max} = 1.09$ cluster $p_{FWE} = 0.001$	**Anterior superior temporal gyri** $t_{max} = 10.89$ $p_{max} = < 0.001$ $d_{max} = 2.44$ cluster $p_{FWE} = < 0.001$
SN/VTA $t_{max} = 5.80$ $p_{max} < 0.001$ $d_{max} = 1.30$ cluster $p_{FWE} < 0.001$	**Inferior colliculi** $t_{max} = 4.24$ $p_{max} < 0.001$ $d_{max} = 0.95$ cluster $p_{FWE} < 0 .001$	**Posterior superior temporal gyri** $t_{max} = 12.61$ $p_{max} = < 0.001$ $d_{max} = 2.82$ cluster $p_{FWE} = <0 .001$
BF $t_{max} = 4.81$ $p_{max} < 0.001$ $d_{max} = 1.08$ cluster $p_{FWE} < 0.001$	**Hippocampus** $t_{max} = -2.38$ $p_{max} = 0.027$ $d_{max} = -0.53$ cluster $p_{FWE} = 0.276$	**Primary visual cortex** $t_{max} = 5.87$ $p_{max} = < 0.001$ $d_{max} = 1.31$ cluster $p_{FWE} = < 0.001$
RN $t_{max} = 5.25$ $p_{max} < 0.001$ $d_{max} = 1.17$ cluster $p_{FWE} < 0.001$		**Primary auditory cortex** $t_{max} = 10.08$ $p_{max} = < 0.001$ $d_{max} = 2.25$ cluster $p_{FWE} = < 0.001$
		Ventral medial prefrontal cortex $t_{max} = -2.66$ $p_{max} = 0.015$ $d_{max} = -0.60$ cluster $p_{FWE} = 0.131$

Table S1. Statistics for the effect of rare patterns in FIR analyses (max t-values, max p-values, max Cohen's d, and p-values for clusters with FWE correction), for all ROIs. When several clusters have been identified, the one with the higher max Cohen's d is reported.

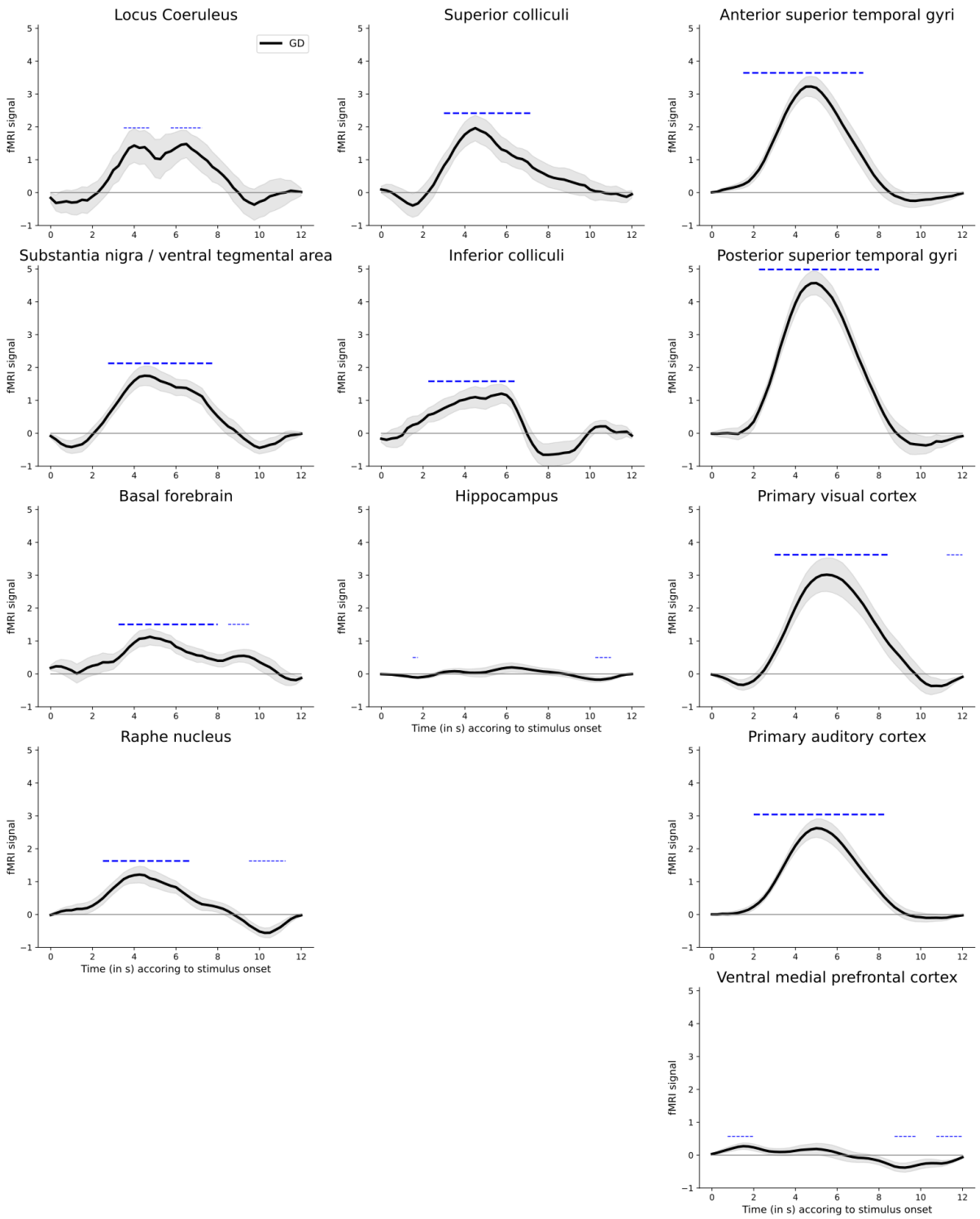

Figure S2. Time course of fMRI activity estimated using a FIR model for rare patterns. The first column shows neuromodulator nuclei, the second column other subcortical ROIs, and the third

column cortical ROIs. Error shading is standard error. Blue black dashed lines indicate significant clusters different from 0 (p_{FWE}<0.05).

Effect of global deviance using GLM analyses

Figure S3 shows the distribution of single-subject estimates for the effect of global deviance for each ROI and Table S2 summarizes the corresponding statistics. All regions showed an increase in fMRI activity for rare patterns compared to frequent patterns. Only the ventral medial prefrontal cortex, the hippocampus, and inferior colliculi showed no significant global effect.

Neuromodulation-related nuclei	Other subcortical nuclei	Cortical areas
LC t(19) = 4.29 p = 0.038 d = 0.96	**Superior colliculi** t(19) = 5.82 p = 0.002 d = 1.30	**Anterior superior temporal gyri** t(19) = 6.74 p < 0.001 z = 1.51
SN/VTA t(19) = 5.24 p < 0.001 d = 1.17	**Inferior colliculi** t(19) = 1.89 p = 0.375 d = 0.42	**Posterior superior temporal gyri** t(19) = 8.98 p < 0.001 d = 1.54
BF t(19) = 4.14 p = 0.022 d = 0.93	**Hippocampus** t(19) = 1.84 p = 0.081 d = 0.41	**Primary visual cortex** t(19) = 6.89 p < 0.001 d = 1.94
RN t(19) = 4.55 p = 0.022 d = 1.02		**Primary auditory cortex** t(19) = 8.34 p < 0.001 d = 1.86
		Ventral medial prefrontal cortex t(19) = 0.26 p = 0.797 d = 0.06

Table S2. Statistics for the difference between the effects of rare and frequent patterns in GLM analyses (t-values, p-values, and Cohen's d), for all ROIs.

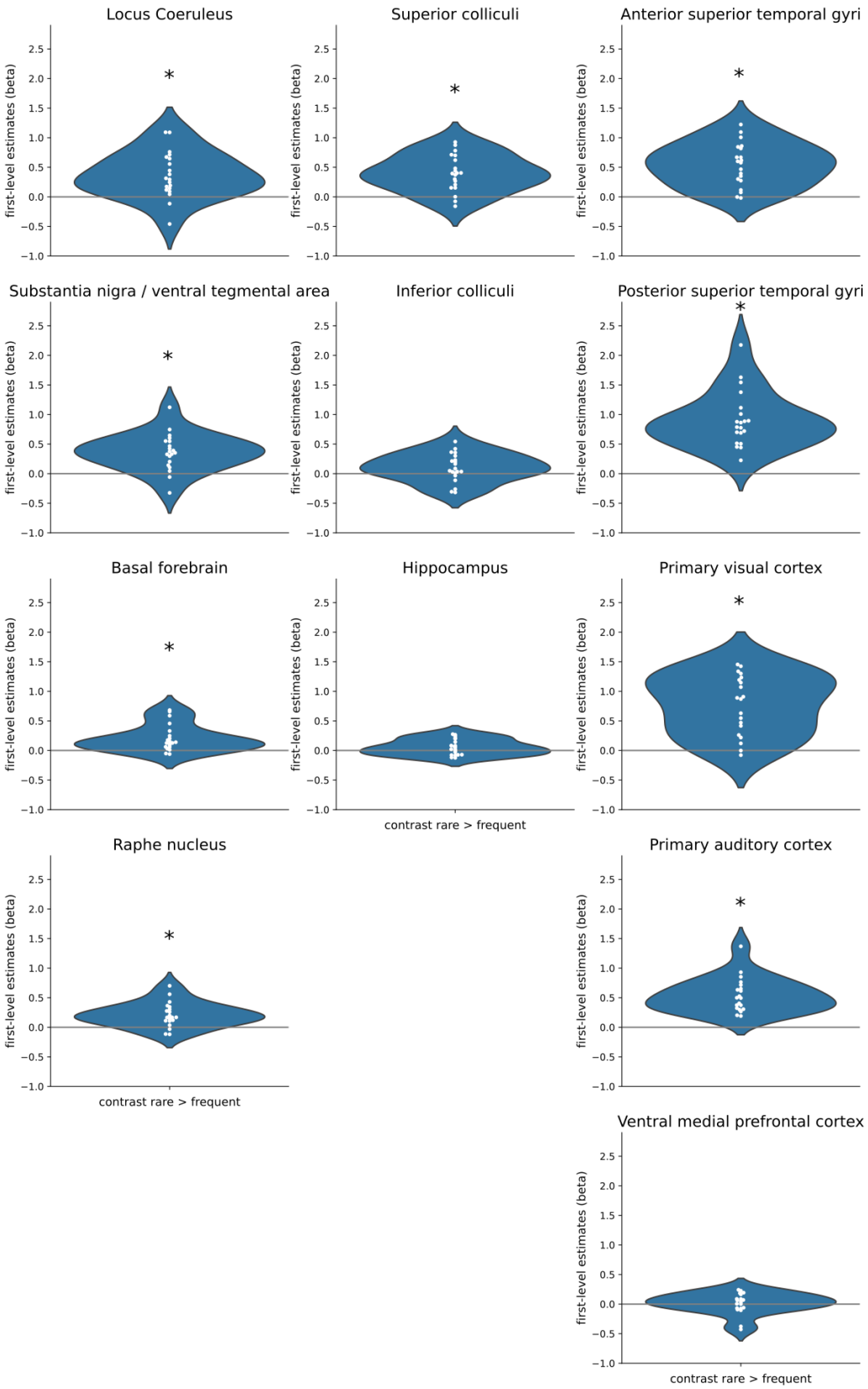

Figure S3. Single-subject estimates for the effect of rare patterns in different brain structures. The first column shows neuromodulator nuclei, the second column other subcortical ROIs, and the third column cortical ROIs. Stars indicate group-level significance ($p < 0.05$).

Individual deliantion of the LC and the SN/VTA using TSE images

Figure S4 and figure S5 present for each participant the slice in both TSE (one centered on the LC and one centered on the SN/VTA) for which each structure was the most visible.

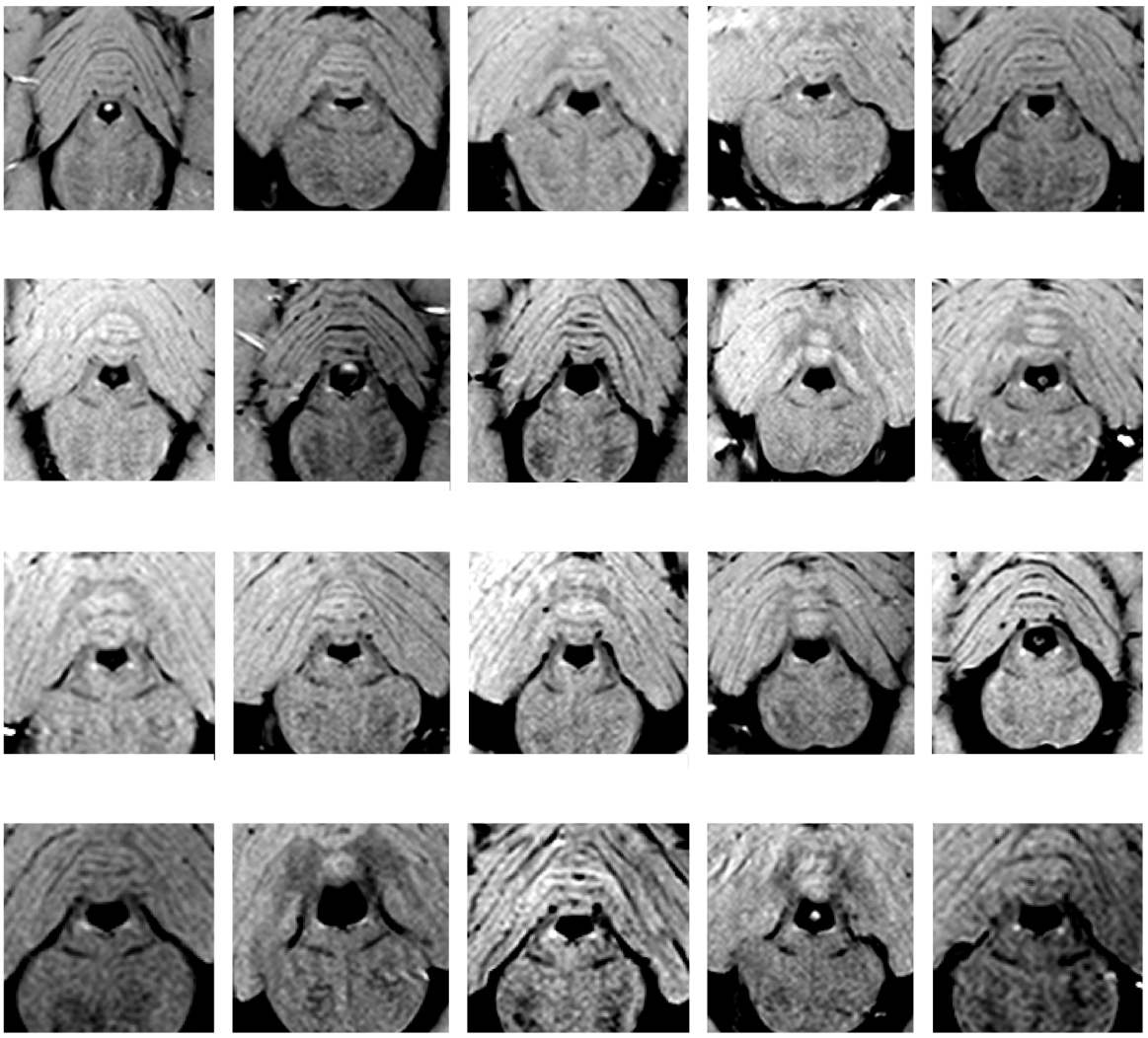

Figure S4. Anatomical images showing the LC in hypersignal for each participant.

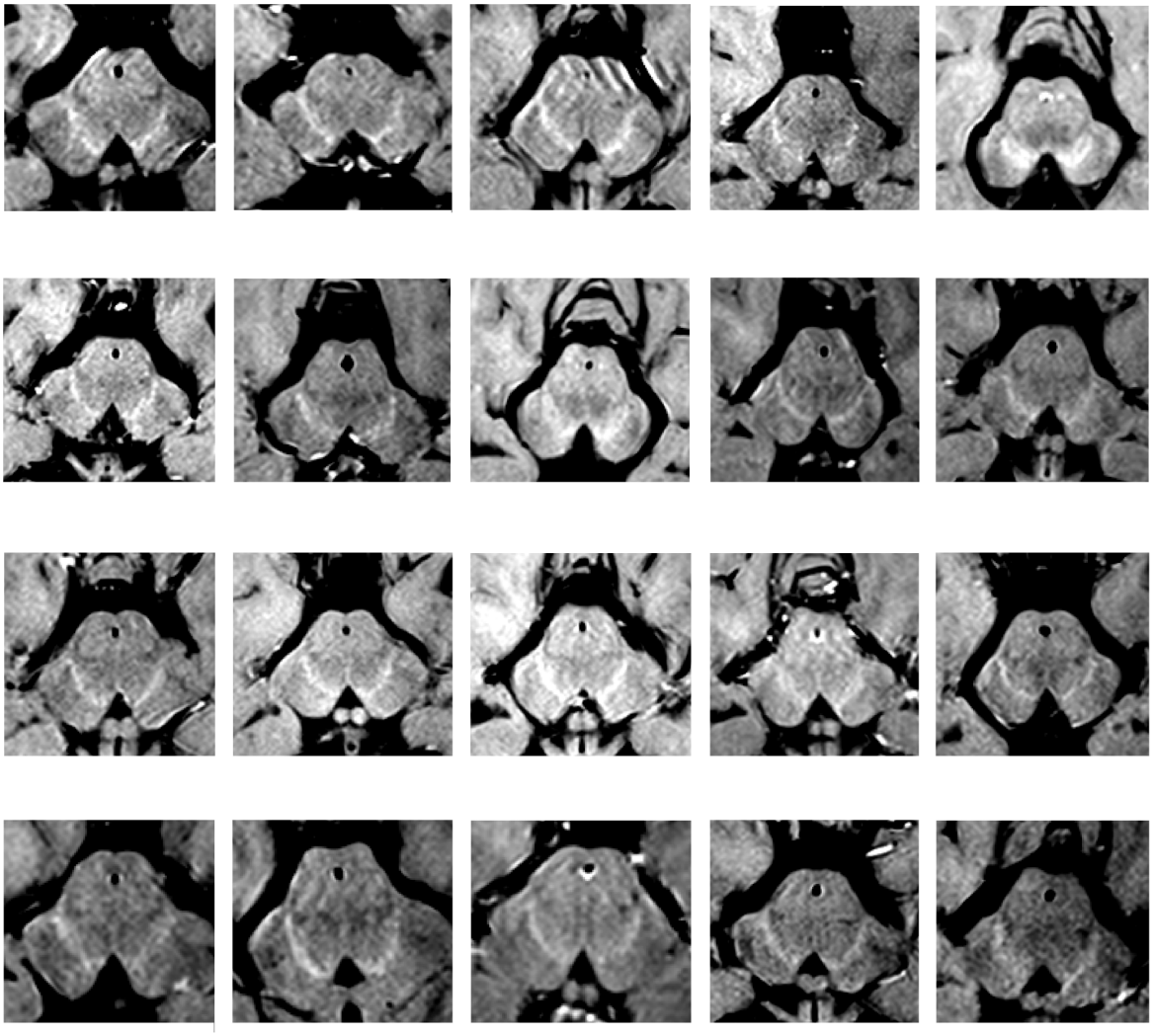

Figure S5. Anatomical images showing the SN/TVA in hypersignal for each participant.

References

Ahnaou, A., Biermans, R., & Drinkenburg, W. H. I. M. (2018). Cholinergic mechanisms of target oddball stimuli detection: The late "P300-like" event-related potential in rats. *Neural Plasticity*, *2018*, 4270263.

Anderson, L. A., & Malmierca, M. S. (2013). The effect of auditory cortex deactivation on stimulus-specific adaptation in the inferior colliculus of the rat. *The European Journal of Neuroscience*, *37*(1), 52–62.

Arnsten, A. F., & Goldman-Rakic, P. S. (1984). Selective prefrontal cortical projections to the region of the locus coeruleus and raphe nuclei in the rhesus monkey. *Brain Research*, *306*(1-2), 9–18.

Astafiev, S. V., Snyder, A. Z., Shulman, G. L., & Corbetta, M. (2010). Comment on "Modafinil Shifts Human Locus Coeruleus to Low-Tonic, High-Phasic Activity During Functional MRI" and "Homeostatic Sleep Pressure and Responses to Sustained Attention in the Suprachiasmatic Area." *Science*, *328*(5976), 309–309.

Aston-Jones, G., & Cohen, J. D. (2005). An integratice theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), 403–450.

Aston-Jones, G., Rajkowski, J., & Kubiak, P. (1997). Conditioned responses of monkey locus coeruleus neurons anticipate acquisition of discriminative behavior in a vigilance task. *Neuroscience*, *80*(3), 697–715.

Aston-Jones, G., Rajkowski, J., Kubiak, P., & Alexinsky, T. (1994). Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *14*(7), 4467–4480.

Ayala, Y. A., & Malmierca, M. S. (2012). Stimulus-specific adaptation and deviance detection in the inferior colliculus. *Frontiers in Neural Circuits*, *6*, 89.

Bassareo, V., De Luca, M. A., & Di Chiara, G. (2002). Differential Expression of Motivational Stimulus Properties by Dopamine in Nucleus Accumbens Shell versus Core and Prefrontal

Cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *22*(11), 4709–4719.

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., De Weerd, P., Kennedy, H., & Fries, P. (2015). Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. *Neuron*, *85*(2), 390–401.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences*, *106*(5), 1672–1677.

Bellet, M. E., Gay, M., Bellet, J., Jarraya, B., Dehaene, S., van Kerkoerle, T., & Panagiotaropoulos, T. I. (2021). Prefrontal neural ensembles encode an internal model of visual sequences and their violations. In *bioRxiv* (p. 2021.10.04.463064). https://doi.org/10.1101/2021.10.04.463064

Brown, S. B. R. E., van der Wee, N. J. A., van Noorden, M. S., Giltay, E. J., & Nieuwenhuis, S. (2015). Noradrenergic and cholinergic modulation of late ERP responses to deviant stimuli. *Psychophysiology*, *52*(12), 1620–1631.

Bunzeck, N., & Düzel, E. (2006). Absolute coding of stimulus novelty in the human substantia nigra/VTA. *Neuron*, *51*(3), 369–379.

Burlingham, C. S., Ryoo, M., Roth, Z. N., Mirbagheri, S., Heeger, D. J., & Merriam, E. (2022). Task-related hemodynamic responses in human early visual cortex are modulated by task difficulty and behavioral performance. *eLife*, *11*. https://doi.org/10.7554/eLife.73018

Caldenhove, S., Borghans, L. G. J. M., Blokland, A., & Sambeth, A. (2017). Role of acetylcholine and serotonin in novelty processing using an oddball paradigm. *Behavioural Brain Research*, *331*, 199–204.

Cazettes, F., Reato, D., Morais, J. P., Renart, A., & Mainen, Z. F. (2021). Phasic Activation of Dorsal Raphe Serotonergic Neurons Increases Pupil Size. *Current Biology: CB*, *31*(1), 192–197.e4.

Chao, Z. C., Takaura, K., Wang, L., Fujii, N., & Dehaene, S. (2018). Large-Scale Cortical Networks for Hierarchical Prediction and Prediction Error in the Primate Brain. *Neuron*, *100*(5), 1252–1266.e3.

Costa, V. D., & Rudebeck, P. H. (2016). More than Meets the Eye: the Relationship between Pupil Size and Locus Coeruleus Activity [Review of *More than Meets the Eye: the Relationship between Pupil Size and Locus Coeruleus Activity*]. *Neuron*, *89*(1), 8–10. Elsevier.

Dayan, P. (2012). Twenty-five lessons from computational neuromodulation. *Neuron*, *76*(1), 240–256.

Debener, S., Makeig, S., Delorme, A., & Engel, A. K. (2005). What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research*, *22*(3), 309–321.

de Gee, J. W., Colizoli, O., Kloosterman, N. A., Knapen, T., Nieuwenhuis, S., & Donner, T. H. (2017). Dynamic modulation of decision biases by brainstem arousal systems. *eLife*, *6*. https://doi.org/10.7554/eLife.23232

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron*, *88*(1), 2–19.

Devauges, V., & Sara, S. J. (1990). Activation of the noradrenergic system facilitates an attentional shift in the rat. *Behavioural Brain Research*. https://www.sciencedirect.com/science/article/pii/016643289090118X

Dubois, M., Habicht, J., Michely, J., Moran, R., Dolan, R. J., & Hauser, T. U. (2021). Human complex exploration strategies are enriched by noradrenaline-modulated heuristics. *eLife*, *10*. https://doi.org/10.7554/eLife.59907

Dunlop, B. W., & Nemeroff, C. B. (2007). The role of dopamine in the pathophysiology of depression. *Archives of General Psychiatry*, *64*(3), 327–337.

Dürschmid, S., Edwards, E., Reichert, C., Dewar, C., Hinrichs, H., Heinze, H.-J., Kirsch, H. E., Dalal, S. S., Deouell, L. Y., & Knight, R. T. (2016). Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(24), 6755–6760.

Eckert, M. A., Keren, N. I., & Aston-Jones, G. (2010). Looking forward with the locus coeruleus.

Science E-Letter.

Faugeras, F., Rohaut, B., Weiss, N., Bekinschtein, T., Galanaud, D., Puybasset, L., Bolgert, F., Sergent, C., Cohen, L., Dehaene, S., & Naccache, L. (2012). Event related potentials elicited by violations of auditory regularities in patients with impaired consciousness. *Neuropsychologia*, *50*(3), 403–418.

Foote, S. L., Aston-Jones, G., & Bloom, F. E. (1980). Impulse activity of locus coeruleus neurons in awake rats and monkeys is a function of sensory stimulation and arousal. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(5), 3033–3037.

Frank, M. J., Santamaria, A., O'Reilly, R. C., & Willcutt, E. (2007). Testing computational models of dopamine and noradrenaline dysfunction in attention deficit/hyperactivity disorder. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *32*(7), 1583–1599.

Friston, K., Ashburner, J., Kiebel, S., Nichols, T., & Penny, W. (2007). *Statistical parametric mapping: the analysis of functional brain images* (Academic Press).

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *108 Suppl 3*, 15647–15654.

Goldstein, D. S., Holmes, C., Bentho, O., Sato, T., Moak, J., Sharabi, Y., Imrich, R., Conant, S., & Eldadah, B. A. (2008). Biomarkers to detect central dopamine deficiency and distinguish Parkinson disease from multiple system atrophy. *Parkinsonism & Related Disorders*, *14*(8), 600–607.

Harsay, H. A., Spaan, M., Wijnen, J. G., & Ridderinkhof, K. R. (2012). Error awareness and salience processing in the oddball task: shared neural mechanisms. *Frontiers in Human Neuroscience*, *6*, 246.

Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, *389*, 54–73.

Hervé-Minvielle, A., & Sara, S. J. (1995). Rapid habituation of auditory responses of locus

coeruleus cells in anaesthetized and awake rats. *Neuroreport*, *6*(10), 1363–1368.

Janitzky, K., Lippert, M. T., Engelhorn, A., Tegtmeier, J., Goldschmidt, J., Heinze, H.-J., & Ohl, F. W. (2015). Optogenetic silencing of locus coeruleus activity in mice impairs cognitive flexibility in an attentional set-shifting task. *Frontiers in Behavioral Neuroscience*, *9*, 286.

Jiang, Y., Komatsu, M., Chen, Y., Xie, R., Zhang, K., Xia, Y., Gui, P., Liang, Z., & Wang, L. (2022). Constructing the hierarchy of predictive auditory sequences in the marmoset brain. *eLife*, *11*. https://doi.org/10.7554/eLife.74653

Joshi, S., & Gold, J. I. (2020). Pupil Size as a Window on Neural Substrates of Cognition. *Trends in Cognitive Sciences*, *24*(6), 466–480.

Karoui, I. E., King, J.-R., Sitt, J., Meyniel, F., Van Gaal, S., Hasboun, D., Adam, C., Navarro, V., Baulac, M., Dehaene, S., Cohen, L., & Naccache, L. (2014). Event-Related Potential, Time-frequency, and Functional Connectivity Facets of Local and Global Auditory Novelty Processing: An Intracranial Study in Humans. *Cerebral Cortex* , bhu143.

Keren, N. I., Lozar, C. T., Harris, K. C., Morgan, P. S., & Eckert, M. A. (2009). In vivo mapping of the human locus coeruleus. *NeuroImage*, *47*(4), 1261–1267.

King, J. R., Faugeras, F., Gramfort, A., Schurger, A., El Karoui, I., Sitt, J. D., Rohaut, B., Wacongne, C., Labyt, E., Bekinschtein, T., Cohen, L., Naccache, L., & Dehaene, S. (2013). Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *NeuroImage*, *83*, 726–738.

Knapen, T., de Gee, J. W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and Ocular Factors Jointly Determine Pupil Responses under Equiluminance. *PloS One*, *11*(5), e0155574.

Krebs, R. M., Park, H. R. P., Bombeke, K., & Boehler, C. N. (2018). Modulation of locus coeruleus activity by novel oddball stimuli. *Brain Imaging and Behavior*, *12*(2), 577–584.

Kutlu, M. G., Zachry, J. E., Melugin, P. R., Cajigas, S. A., Chevee, M. F., Kelly, S. J., Kutlu, B., Tian, L., Siciliano, C. A., & Calipari, E. S. (2021). Dopamine release in the nucleus accumbens core signals perceived saliency. *Current Biology: CB*, *31*(21), 4748–4761.e8.

Lesicko, A. M. H., Angeloni, C. F., Blackwell, J. M., De Biasi, M., & Geffen, M. N. (2022). Corticofugal Regulation of Predictive Coding. *eLife*, *11*, 2022;11:e73289.

Liu, K. Y., Marijatta, F., Hämmerer, D., Acosta-Cabronero, J., Düzel, E., & Howard, R. J. (2017). Magnetic resonance imaging of the human locus coeruleus: A systematic review. *Neuroscience and Biobehavioral Reviews*, *83*, 325–355.

Luppi, P. H., Aston-Jones, G., Akaoka, H., Chouvet, G., & Jouvet, M. (1995). Afferent projections to the rat locus coeruleus demonstrated by retrograde and anterograde tracing with cholera-toxin B subunit and Phaseolus vulgaris leucoagglutinin. *Neuroscience*, *65*(1), 119–160.

Malmierca, M. S., Cristaudo, S., Pérez-González, D., & Covey, E. (2009). Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(17), 5483–5493.

Megemont, M., McBurney-Lin, J., & Yang, H. (2022). Pupil diameter is not an accurate real-time readout of locus coeruleus activity. *eLife*, *11*. https://doi.org/10.7554/eLife.70510

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, *35*(8), 4140–4154.

Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdés-Baizabal, C., Escera, C., & Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature Communications*, *8*(1), 2148.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–1045.

Poe, G. R., Foote, S., Eschenko, O., Johansen, J. P., Bouret, S., Aston-Jones, G., Harley, C. W., Manahan-Vaughan, D., Weinshenker, D., Valentino, R., Berridge, C., Chandler, D. J., Waterhouse, B., & Sara, S. J. (2020). Locus coeruleus: a new look at the blue spot. *Nature Reviews. Neuroscience*, *21*(11), 644–659.

Quirins, M., Marois, C., Valente, M., Seassau, M., Weiss, N., El Karoui, I., Hochmann, J.-R., & Naccache, L. (2018). Conscious processing of auditory regularities induces a pupil dilation. *Scientific Reports*, *8*(1), 14819.

Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1994). Locus coeruleus activity in monkey: phasic and tonic changes are associated with altered vigilance. *Brain Research Bulletin*, *35*(5-6), 607–616.

Reichardt, R., Polner, B., & Simor, P. (2020). Novelty Manipulations, Memory Performance, and Predictive Coding: the Role of Unexpectedness. *Frontiers in Human Neuroscience*, *14*, 152.

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, *7*, 13289.

Safaai, H., Neves, R., Eschenko, O., Logothetis, N. K., & Panzeri, S. (2015). Modeling the effect of locus coeruleus firing on cortical state dynamics and single-trial sensory processing. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), 12834–12839.

Sara, S. J., & Bouret, S. (2012). Orienting and reorienting: the locus coeruleus mediates cognition through arousal. *Neuron*, *76*(1), 130–141.

Sara, S. J., Vankov, A., & Hervé, A. (1994). Locus coeruleus-evoked responses in behaving rats: a clue to the role of noradrenaline in memory. *Brain Research Bulletin*, *35*(5-6), 457–465.

Schomaker, J., & Meeter, M. (2015). Short- and long-lasting consequences of novelty, deviance and surprise on brain and cognition. *Neuroscience and Biobehavioral Reviews*, *55*, 268–279.

Siegel, M., Donner, T. H., & Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews. Neuroscience*, *13*(2), 121–134.

Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews. Neuroscience*, *20*(10), 635–644.

Strange, B. A., & Dolan, R. J. (2007). β-adrenergic modulation of oddball responses in humans. *Behavioral and Brain Functions: BBF*, *3*(1), 1–5.

Strauss, M., Sitt, J. D., King, J.-R., Elbaz, M., Azizi, L., Buiatti, M., Naccache, L., van Wassenhove, V., & Dehaene, S. (2015). Disruption of hierarchical predictive coding during sleep. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(11), E1353–E1362.

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., & Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature*, *571*(7765), 361–365.

Swick, D., Pineda, J. A., Schacher, S., & Foote, S. L. (1994). Locus coeruleus neuronal activity in awake monkeys: relationship to auditory P300-like potentials and spontaneous EEG. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *101*(1), 86–92.

Uhrig, L., Dehaene, S., & Jarraya, B. (2014). A Hierarchy of Responses to Auditory Regularities in the Macaque Brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(4), 1127–1132.

Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and dopamine neurons in the reward/effort trade-off: a direct electrophysiological comparison in behaving monkeys. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(20), 7866–7877.

Verleger, R., & Śmigasiewicz, K. (2016). Do Rare Stimuli Evoke Large P3s by Being Unexpected? A Comparison of Oddball Effects Between Standard-Oddball and Prediction-Oddball Tasks. *Advances in Cognitive Psychology / University of Finance and Management in Warsaw*, *12*(2), 88–104.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20754–20759.

Wang, X.-J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*, *90*(3), 1195–1268.