Title: Receptors, circuits and neural dynamics for prediction

Authors: Sounak Mohanta[1], Mohsen Afrasiabi[1], Cameron Casey[2], Sean Tanabe[2], Michelle J. Redinbaugh[1], Niranjan A. Kambi[1], Jessica M. Phillips[1], Daniel Polyakov[2], William Filbey[2], Joseph L. Austerweil[1], Robert D. Sanders[2*]†, Yuri B. Saalmann[1*]†

Affiliations:

[1] Department of Psychology, University of Wisconsin - Madison

[2] Department of Anesthesiology, University of Wisconsin - Madison

*Correspondence to: saalmann@wisc.edu or robert.sanders@sydney.edu.au

† Authors contributed equally

Summary: Learned associations between stimuli allow us to model the world and make predictions, crucial for efficient behavior; e.g., hearing a siren, we expect to see an ambulance and quickly make way. While theoretical and computational frameworks for prediction exist, circuit and receptor-level mechanisms are unclear. Using high-density EEG, Bayesian modeling and machine learning, we show that trial history and frontal alpha activity account for reaction times (a proxy for predictions) on a trial-by-trial basis in an audio-visual prediction task. Predictive beta feedback activated sensory representations in advance of predicted stimuli. Low-dose ketamine, a NMDA receptor blocker – but not the control drug dexmedetomidine – perturbed predictions, their representation in higher-order cortex, and feedback to posterior cortex. This study suggests predictions depend on alpha activity in higher-order cortex, beta feedback and NMDA receptors, and ketamine blocks access to learned predictive information.

Keywords: Predictive coding, Ketamine, NMDA receptors, Frontal cortex, Feedback.

Introduction

The classical view of sensory processing focuses on feedforward information transmission from the sensory organs to higher-order cortex, to generate representations of the world [1,2]. However, growing evidence of expectations strongly influencing perception and behavior [3–5] suggests that the brain actively predicts incoming sensory information. This is captured in a radically different view of sensory processing, called predictive coding (PC), where the brain uses generative models to make inferences about the world [6–10], possibly even to support conscious experience [11–13]. PC proposes that these models, based on prior sensory experiences, are represented at higher-order levels of a cortical hierarchy. The model predictions are transmitted from higher-order to lower-order cortex along feedback connections. Any mismatch between feedback predictions and observed sensory evidence generates an error signal, which is transmitted along feedforward connections, to update models in higher-order cortex [14–16]. The updated model is used in the next iteration to generate new predictions. This process of Bayesian updating aims to optimize beliefs about the sensory world. However, the neural representation of predictions is unclear.

N-methyl-D-aspartate receptors (NMDARs) may play a key role in PC. Theoretical work on PC [17] has proposed that higher levels of a cortical hierarchy transmit top-down predictions through NMDAR-mediated signaling. Consistent with this proposal, NMDARs have been shown to modulate higher-order (frontal) cortical excitability [18–21], be enriched in superficial and deep cortical layers where feedback connections terminate [22], and contribute to feedback activity [23]. However, there is a lack of experimental evidence linking NMDARs and prediction itself. Ketamine, a NMDAR blocker [24], can reduce error signals, measured as auditory mismatch negativity (MMN) in oddball paradigms [25–27]. But because the MMN reflects the mismatch between predictions and observed sensory evidence, it is difficult to dissect if ketamine's effect

on the MMN is due to ketamine influencing feedback predictions, feedforward sensory evidence or error signaling directly, any of which can reduce the MMN. Further, prediction is inferred in oddball paradigms; there is no behavioral measure of prediction. Hence, a paradigm that incorporates a behavioral readout of predictions and separates predictions from other PC mechanisms is required to demonstrate NMDARs' contribution to predictions and their neural representation.

To test circuit and receptor-level mechanisms of prediction, we recorded 256-channel EEG of subjects performing an audio-visual delayed match-to-sample task (Figure 1a). The task design separates predictions (generated during the delay period) from error processing (after image onset), which are not readily separable in oddball paradigms [25–27]. Subjects performed the task before, during, and after recovery from sub-hypnotic dosing of ketamine, targeted to concentrations that modulate NMDARs, or the control drug dexmedetomidine (DEX), an α_2 adrenergic receptor agonist, selected to account for changes in arousal and modulation of hyperpolarization-activated cyclic nucleotide channels (HCN-1, which mediate ketamine's anesthetic effects [28]). We found that subjects responded quickest to highly predictive sounds, reaction times (RTs) thus serving as a behavioral readout of predictions. Classification analysis using a support vector machine demonstrated predictions increased the signal-to-noise ratio (SNR) in frontal cortex and, based on hierarchical drift diffusion modeling (HDDM) [29], frontal alpha power predicted RTs on a trial-by-trial basis. Moreover, frontal cortex transmitted predictions to posterior cortex using beta frequencies, measured as Granger causal influence, activating sensory representations of predicted images before their presentation. Because ketamine, but not DEX, blocked these neural representations of predictions and their behavioral advantage, this study suggests a key role for frontal and posterior alpha activity, beta feedback and NMDARs in prediction.

Results

Predictions improved RTs. Subjects initially learned paired associations (A1-V1, A2-V2, A3-V3) between 3 sounds (A1, A2, A3) and 3 images (V1, V2, V3) through trial-and-error, allowing us to modulate separate feedforward (auditory to frontal) and feedback (frontal to visual) pathways. During learning, each sound and image had equal probability (33%) of appearing in any given trial, preventing subjects developing any differential predictions due to stimulus frequency. Following presentation of both stimuli, subjects reported if the sound and image matched or not. To manipulate subjects' predictions during subsequent testing, we varied the probability of an image appearing after its associated sound. This probability was different for each sound: 85% chance of V1 after A1; 50% chance of V2 after A2; and 33% chance of V3 after A3 (Figure 1a). Thus, A1 was highly predictive (HP), A2 was moderately predictive (MP), and A3 was not match predictive (NP).

We hypothesized that increasing the predictive value of the sound would allow subjects to make better predictions about the upcoming image, enabling quicker responses (HP<MP<NP) in match trials. RTs were faster when sounds had greater predictive value (ANOVA, N=15, P<0.00001) (Figure 1b). This result was further validated in parallel psychophysics experiments, where we controlled for possible match bias [30] (ANOVA, N=25, P=0.0001; Figure S1a). We found RTs could not be explained by speed-accuracy trade-offs, as subjects were most accurate for HP, followed by MP and NP sounds (ANOVA, N=25, P=0.001, Figure S1b).

Ketamine blocked fast RTs to predictive sounds. If predictions are mediated by NMDARs, then ketamine should block predictions; i.e., subjects administered a sub-anesthetic dose of ketamine should be unable to exploit the differential predictive value of each sound, causing the

linear trend in RTs (HP<MP<NP sounds) to disappear. Regression analysis for prediction (HP, MP, NP) X drug condition (before ketamine, under ketamine) revealed, in addition to a main effect of prediction, a significant interaction effect (minimum effective plasma concentration of 0.2 μg/ml; ANOVA, N=15, main effect of prediction: P=0.0009, interaction of prediction X drug condition: P=0.005). The interaction effect confirmed that, under ketamine, subjects no longer showed a linear correlation between the predictive value of sounds and RT (Figure 1c). This was not due to low accuracy as subjects' average accuracy was 77.8% under ketamine (85.7% without ketamine, and 81.0% under DEX). Effects were specific to NMDAR manipulation, as DEX did not block predictions. Subjects under DEX showed only the main effect of prediction and no interaction of prediction and drug condition (before DEX, under DEX) (ANOVA, N=14, main effect of prediction: P=0.0002, interaction of prediction X drug condition: P=0.398). Under DEX, the linear correlation between the predictive value of sounds and RT was intact (Figure 1e), similar to the baseline condition. Results were not due to sedation as, subjects were more alert under ketamine than dexmedetomidine (ketamine average modified observer's assessment of alertness/sedation (OAA/S) score of 4.85 compared to 3.33 under DEX (5, awake – 1, unresponsive)). After recovery from ketamine (2-4 hours after ending ketamine administration, depending upon subject's recovery), subjects once again showed a linear correlation between the predictive value of sounds and RT (Figure 1d). Regression analysis of prediction (HP, MP, NP) X drug condition (before ketamine, after recovery from ketamine) showed a significant main effect of prediction but no interaction (ANOVA, N=12, 3 subjects excluded post-ketamine due to vomiting; main effect of prediction: P<0.00001, interaction of prediction X drug condition: P=0.32). We also ran two confirmatory analyses. First, a repeated measures regression for prediction (HP, MP, NP) X drug condition (before drug, under ketamine, under DEX), and found a significant main effect of prediction (ANOVA, P<0.00001) as well as a significant interaction of prediction and drug condition (P=0.0009), showing that

ketamine blocked the linear correlation of RT with predictive strength, compared to that before the drug and under DEX. Second, to demonstrate that the linear trend in RTs vanishes under ketamine, in comparison to before ketamine and after recovery, we ran another repeated measure regression analysis for prediction (HP, MP, NP) X drug condition (before, under, after ketamine). We found a significant main effect of prediction and a significant interaction of prediction and drug condition (ANOVA, main effect of prediction: P=0.0001, interaction of prediction X drug condition: P=0.01), revealing indeed that ketamine blocked the linear correlation of RTs with predictive strength, compared to before the drug and after recovery. Overall, our results demonstrate that subjects used predictive information to enhance behavioral performance and these predictions depend on NMDARs.

Predictions depended on strength of causal power. To identify what information in the trial history subjects base predictions on, we used a HDDM [29], with two possible choices from our task (correct/incorrect for match trials). In the HDDM, evidence accumulates (drift process) from a starting point to two boundaries (one for each choice). It stops when it reaches a boundary, which determines the choice (and RT) for the trial (Figure 1f). The starting point is determined by the predictive value of the sound, which may be biased towards one of the choices (reflected in a bias parameter; *z*). A drift process that starts with a larger bias will reach the decision boundary quicker, resulting in a faster RT. We tested two possibilities: (i) did subjects generate predictions by keeping track of simple co-occurrences of each sound-image pair (correlation); or (ii) did subjects not only track the occurrence of each image with its paired sound but also unpaired sounds. The latter case (ii) helps subjects eliminate weak/conditional relations between particular sounds and images, in order to conceive a cause-effect structure between stimuli. As a predictor of the bias parameter in the HDDM, we calculated the transitional

probability (how often a particular image follows the sound only; i.e., correlation) and causal power [31] (amount of evidence that a sound "causes" a particular image, as opposed to chance; i.e., causation in a statistical sense) of the sound-image association, updated each trial (Figure 1g). Specifically, we estimated the posterior probability density of the regression coefficient (β_1; Figure 1f) to determine the relationship between bias and either transitional probability or causal power. We found causal power (deviance information criterion, DIC=-3197) predicted RTs better than transitional probabilities (DIC=-1795), i.e., causal power better captured the basis of prediction generation (option (ii) above). Further, bias was positively correlated with causal power (P=0.04; Figure 1h). This suggests that more predictive sounds (higher causal power) create a larger bias before the onset of the visual image and, as a result, decisions are reached more quickly, which translates to quicker RTs.

We used the HDDM to model drug effects. If ketamine blocks predictions, all sounds will generate similar biases; i.e., there will be no correlation between bias and the predictive value of sounds. Indeed, under ketamine, the posterior probability density of β_1 was not significantly different from zero (P=0.76; Figure 1h). In contrast, under DEX, bias positively correlated with the predictive values of sounds (all sounds generating equal vs different biases, P<0.00001; Figure 1h) similar to baseline. After recovery from ketamine, the posterior probability density of β_1 was significantly greater than zero (P<0.00001; Figure 1h) confirming that subjects had again generated larger bias for more predictive sounds. The question is: (a) did subjects re-gain access to previously learned and stored predictive information (Figure S2b) or (b) did they re-learn the predictive value of each sound after recovery from ketamine (Figure S2c)? To answer this, we used the HDDM to analyze the first 30 trials for each subject after recovery (translating to approximately 10 trials for each auditory cue, for every subject). We found that, only for the

former (option (a)), bias positively correlated with causal power (P=0.03; Figure S2d). This suggests that ketamine did not erase previously learned predictive information, but rather ketamine prevented access to the predictive information.

Strength of predictions correlated with frontal alpha power. We next investigated the circuit-level mechanism of prediction. To show task-responsive EEG electrodes, we averaged the time-frequency response across all trials and all electrodes (this selection procedure does not bias towards particular predictions). This revealed a task-related increase in baseline-corrected alpha power (8-14Hz; Figure S3b). Four clusters of electrodes, right frontal (RF), right central (RC), left central (LC) and occipital (OC), showed significant modulation of delay period alpha power compared to baseline, irrespective of sound (Figure 3a). Considering alpha power as an index of neural excitability (reduced alpha equating to reduced inhibition/increased excitability) [32,33], one might expect stronger predictions to be associated with lower alpha power, reflecting greater activation of prediction-encoding neurons. Indeed, stronger predictions significantly correlated with lower delay period alpha power at the RF electrode cluster (ANOVA, N=26, P=0.002; Figure 2a-c; and Figure S3d-f). This was not due to feedforward sensory processing, as all three sounds generated similar auditory ERPs (Figure S3a). Hence, frontal alpha indexed predictions.

Ketamine blocked the correlation between predictions and frontal alpha power. Because ketamine blocked predictions, subjects under ketamine should no longer show frontal alpha differences between sounds. NMDAR blockade has been shown to increase frontal cortical excitability [18–21], reducing response selectivity and SNR [34,35]. We thus expect low-dose ketamine

to increase frontal cortical excitability irrespective of the predictive value of sounds. This would manifest as similarly low RF alpha power for all sounds. We ran a regression analysis of prediction (HP, MP, NP) X drug condition (before drug, under drug; drug either ketamine or DEX). For ketamine, subjects showed a main effect of prediction as well as a significant interaction effect (ANOVA, N=15, main effect of prediction: P=0.0002, interaction of prediction X drug condition: P=0.03). The interaction effect confirmed that, under ketamine, delay period alpha power at the RF electrode cluster was similar across sounds (Figure 2e). This was not due to a change in feedforward sensory processing under ketamine as all three sounds generated similar auditory ERPs to that before ketamine (repeated measures ANOVA, N=15, main effect of prediction: P=0.79; interaction effect of prediction X drug condition, before or under ketamine: P=0.33). Whereas, for DEX, there was no interaction effect in the regression analysis, only a main effect of prediction (ANOVA, N=11, main effect of prediction: P<0.00001, interaction of prediction X drug condition: P=0.36). RF alpha power still showed a significant linear trend (Figure 2f) similar to baseline. We also ran a confirmatory analysis, i.e., repeated measures regression for prediction (HP, MP, NP) X drug condition (before drug, under ketamine, under DEX), and found a significant main effect of prediction (ANOVA, P=0.0003) as well as a significant interaction of prediction and drug condition (P=0.0007), showing that ketamine blocked the linear correlation of RF alpha power with predictive strength, compared to that before the drug and under DEX. This suggests that NMDARs mediate prediction strength through modulation of frontal excitability (alpha).

Decodability of predictions reduced under ketamine. Increased frontal excitability does not necessarily translate to useful prediction, if it reduces SNR. We trained a decoder to measure if this hyperexcitability, reflected in low RF alpha power, still allows differential representation of

predictions. Before ketamine, the classification F score for each sound separated at 560 ms after sound onset (ANOVA, $P<0.0018$; Figure 3b). Similarly, classification F score for DEX separated at 620 ms after sound onset (ANOVA, $P<0.0067$; Figure 3d). In contrast, under ketamine, there was no separation of classification F score for each sound, and accuracy overall was lower (ANOVA, $P>0.48$; Figure 3c). The weighting of features contributing to classifier performance confirmed that, before ketamine, RF alpha power contributed most to classification accuracy (RF alpha power feature ($W_{RF\alpha}$) > other features ($W_{\sim RF\alpha}$), ANOVA, $P<0.008$; Figure 3f). This was also true under DEX ($W_{RF\alpha} > W_{\sim RF\alpha}$, ANOVA, $P<0.013$; Figure 3h). As expected, under ketamine, RF alpha power contributed little to classification accuracy ($W_{RF\alpha} > W_{\sim RF\alpha}$, ANOVA, $P>0.51$; Figure 3g). These results suggest that ketamine blocks predictions by reducing frontal SNR.

Frontal alpha power correlated with RTs on trial-by-trial basis; but ketamine perturbed the correlation. Although frontal activity correlates with predictions, we need to show that subjects use it. We used the HDDM to show RF alpha power predicts RT on a trial-by-trial basis. We calculated the posterior probability density of the regression coefficient, $\beta_{1\alpha}$ – the relationship between the bias and RF alpha power. Bias was inversely correlated with RF alpha power ($P=0.04$, Figure 3e). This suggests that for more predictive sounds, lower RF alpha power creates a larger bias, and as a result, decisions are reached quicker (quicker RT). In contrast, ketamine blocked the correlation between bias and alpha power (posterior probability distribution of regression coefficient, $\beta_{1\alpha}$, did not differ from zero, $P=0.51$; Figure 3e). This shows that subjects used frontal activity to make predictions.

Predictions activated sensory representations prior to visual stimuli; but ketamine perturbed these pre-stimulus activations. Prior work suggests that predictions can activate sensory templates prior to visual stimuli onset [36]. To test what neural circuits and receptors contribute to this, we used cross-temporal decoding to classify match (A2-V2) versus non-match (A2-V1 and A2-V3) trials for the moderately predictive scenario. We chose the moderately predictive scenario to ensure that there are a similar number of match and non-match trials. Figure 4 a-c shows the classification accuracy for the different time windows on which the classifier was trained (y-axis) and tested (x-axis) across the delay period and after visual stimulus onset. When we trained the classifier on visual stimulus-evoked activity, there was significant decoding earlier in the delay period for the "before drug" (starting 508 ms before visual stimulus onset, and remaining significant; $t(2000)>5.21$, $P<0.003$; Figure 4 a and d) and DEX (starting 476 ms before stimulus onset; $t(2000)>3.34$, $P<0.013$; Figure 4 c and d) conditions, compared to ketamine (starting 264 ms before stimulus onset; note that 275 ms before stimulus onset corresponds to the earliest that any stimulus-evoked response could contribute due to wavelet length; $t(2000)>1.12$, $P<0.048$; Figure 4 b and d). Importantly, we were able to train the classifier on the delay period activity and decode the visual stimulus for the "before drug" (starting at 524 ms before stimulus onset) and DEX (starting at 492 ms before stimulus onset) conditions. In contrast, under ketamine, our classifier decoded the visual stimulus significantly later than "before drug" and under DEX (ANOVA, $t = 2.3$, $p = 0.007$) and only just before stimulus onset (starting 376 ms before visual stimulus onset; Figure 4d). Previous functional MRI [37] and EEG [38–40] studies found visual-evoked responses to greebles in higher-order sensory cortex of the right hemisphere. Accordingly, RC alpha power contributed most to the classification accuracy for both the "before drug" (mean decrease impurity, $MDI_{RC\alpha} > MDI_{\sim RC\alpha}$, ANOVA, $P=0.003$; Figure 4e) and DEX ($MDI_{RC\alpha} > MDI_{\sim RC\alpha}$, ANOVA, $P=0.022$; Figure

4g) conditions. In contrast, under ketamine, RC alpha power did not significantly contribute to the classification accuracy ($MDI_{RC\alpha} > MDI_{\sim RC\alpha}$, ANOVA, P=0.14; Figure 4f). This suggests that NMDA receptors contribute to the pre-stimulus sensory templates activated by predictions.

Strength of predictions correlated with feedback operating at beta frequencies. To investigate how predictions generated frontally influence pre-stimulus sensory templates, we measured functional connectivity between frontal and sensory EEG sources. We first calculated sources of sensor-level, task-related increases in baseline-corrected alpha power. We found frontal cortical contributions to increased sensor-level alpha power in the right hemisphere only, as well as temporal cortical contributions consistent with previous greeble studies [37] (ROIs in Figure 5b inset showed significantly increased source power during the delay period compared to baseline; $p < 0.025$, Bonferroni corrected across all cortical AAL ROIs, to ensure robust task-related sources). Consequently, we restricted our source-level Granger causality analysis between these sources within the right hemisphere. Using non-parametric spectral Granger causality, we measured the source-level predictive feedback from the superior and medial frontal gyrus (Figure 5b, green) to inferior temporal gyrus (Figure 5b, red). The audio-visual task anatomically isolates predictions transmitted along feedback pathways to posterior visual areas during the delay period from the feedforward auditory signals. One might expect higher frontal excitability (from stronger predictions or effect of ketamine) to give rise to stronger feedback. Indeed, stronger predictions were proportionately associated with greater Granger causal influence of right frontal cortex on right inferior temporal cortex in the beta band (15-30 Hz) during the delay period (ANOVA, N=26, P=0.0001; Figure 5, a and b). This suggests predictions are disseminated along feedback connections down the cortical hierarchy prior to image onset.

Ketamine perturbed correlation between prediction strength and beta feedback. Because NMDAR blockers have been reported to perturb feedback pathways in macaques [23] and humans [41], we expected ketamine to alter feedback carrying predictions from frontal to inferior temporal cortex. We ran a regression analysis of prediction (HP, MP, NP) X drug condition (before drug, under drug; drug either ketamine or DEX). For ketamine, subjects showed a main effect of prediction as well as a significant interaction effect (ANOVA, N=15, main effect of prediction: P=0.003, interaction of prediction X drug condition: P=0.03). The interaction effect demonstrated that there was no longer a correlation between the predictive value of sounds and beta band Granger causal influence of right frontal on right inferior temporal cortex (Figure 5c); i.e., ketamine scrambled the feedback for each sound. In contrast, for DEX, while there was a main effect of prediction, the interaction effect was not significant (ANOVA, N=11, main effect of prediction: P<0.00001, interaction of prediction X drug condition: P=0.90). That is, the Granger causal influence in the beta band still differed between sounds under DEX (Figure 5d). We also ran a confirmatory analysis, i.e., repeated measures regression for prediction (HP, MP, NP) X drug condition (before drug, under ketamine, under DEX), and found a significant main effect of prediction (ANOVA, P=0.0001) as well as a significant interaction of prediction and drug condition (P=0.04), showing that ketamine blocked the linear correlation of Granger causal influence of right frontal on right inferior temporal cortex with predictive strength, compared to that before the drug and under DEX. This suggests that the predictive feedback facilitating behavior depends on NMDARs.

Discussion

Our results show NMDAR-mediated, circuit-level mechanisms of prediction and its behavioral effects. Frontal cortex represented predictions and, starting prior to image onset, transmitted them to posterior cortex in the beta band, to activate a sensory representation of the predicted image. Stronger predictions enabled quicker responses, and the HDDM showed subjects generated their predictions based on the strength of causal power, a measure of prior knowledge. In contrast, ketamine prevented quick responses to predictive stimuli, as well as subjects from using the strength of causal power to generate predictions. At the circuit level, ketamine blocked predictions by reducing frontal alpha power to the same low level prior to all images (reducing SNR), leading to undifferentiated feedback and perturbed pre-stimulus sensory activations. Overall, it suggests that NMDARs normally sharpen representations of predictions in frontal and posterior cortex, to enable PC. The data are less supportive of the classical view of sensation, with its reliance on feedforward processing to reconstruct images because one might have expected little systematic difference in behavioral and neural measures for different predictive conditions.

The initial predictive auditory stimulus will activate auditory pathways, leading to the generation of a prediction of the subsequent visual image by a higher-order, multi-modal area. Complex auditory stimuli, like the trisyllabic greeble names in our task, are represented in auditory lateral belt and parabelt cortex (Kaas et al. 1999). Belt and parabelt regions are connected with a number of multi-modal areas in superior temporal and prefrontal cortex [43,44], where there are multisensory neurons responding to both vocalizations and images [45,46]. Recent work [47] suggests that early sensory fusion occurs in temporal or parietal multimodal areas, whereas more flexible weighting and integration of sensory signals for adaptive behavior occurs in frontal multimodal areas. Our results are consistent with this, but go further by showing that the frontal

multimodal areas are the source of predictive multimodal signals, which, in addition to the enhancement of sensorimotor processing shown here, may be useful for communication and language processing [48]. Interestingly, the right frontal cortex was the prediction source in our study and right posterior cortex showed pre-stimulus sensory activations of the predicted greeble. This is consistent with previous studies suggesting a possible right hemispheric bias for the processing of greebles [37,49,50], and possibly faces more generally [51,52].

Intracortical laminar recordings in animal studies of sensory and attentional processing suggest that feedforward signaling operates at gamma frequencies, whereas feedback signaling operates at lower frequencies [53–55]. Consistent with this, previous work on predictive processing using univariate measures has reported the involvement of various lower frequencies, including theta, alpha and beta bands [56–58]. Although there are varying reports of the direction of alpha power changes in predictive processing, this may be due to “when” or “what” is being predicted, differences in task structure (e.g., analyzing the stimulus or delay period) and differences between brain regions [56,59]. Further considering connectivity, a recent macaque study reported greater alpha and beta feedback from prefrontal to visual cortex during the presentation of more predictive visual stimuli [57]. Our finding of frontal cortex causally influencing posterior cortex in the beta band according to predictions extends this finding to humans and to the delay period in the absence of sensory stimulation, as well as provides support for PC models that incorporate a key role for oscillatory activity more generally [56,60,61].

NMDARs are located in both superficial and deep cortical layers [22], potentially allowing NMDARs to modulate the representation of predictions in deep layers, as proposed in certain

PC models [60], and predictive feedback signaling to superficial and/or deep layers. NMDAR blockade in humans has been shown to modulate frontal cortical excitability [20,21], and our decoding analyses suggest that this NMDAR-related change in excitability reduces the SNR of prediction representations. This is consistent with macaque experiments showing that NMDAR blockade reduces the SNR in frontal cortex during the working memory period of an anti-saccade task [34]. In our study, NMDAR blockade also perturbed predictive feedback, consistent with macaque experiments showing NMDARs contribute to feedback signaling [23]. Taken together, these results suggest NMDARs influence both the representation of predictions in higher-order areas and predictive signaling to lower-order areas.

Our results have broader clinical and scientific relevance. Ketamine is a promising treatment for depression [62]. Possible mechanisms of rapid anti-depressant actions of ketamine include blocking lateral habenula bursting activity [63] and activating the mammalian target of rapamycin (mTOR) signaling pathway [64]. Our results point to an additional antidepressant mechanism for ketamine. Depression is associated with negative predictions about upcoming personal events [65,66]. By blocking predictions, ketamine may reduce the negative [67,68] bias, to ameliorate depressive symptoms [69]. Finally, it has been proposed that generative models create virtual realities that support conscious experience [11–13]. That subjects' predictions in our study could be blocked, and error signaling reduced, without impairing consciousness imposes constraints on PC as a theory of consciousness.

Acknowledgements:

We thank G. Lupyan, R.A. Pearce and B.R. Postle for useful discussions. NIH grants R01MH110311 (Y.S.), R01NS117901 (Y.S and R.S.), K23AG055700 (R.S.) and R01AG063849 (R.S.) supported this work

Author contributions:

S.M., M.R., N.K., J.P., R.S. and Y.S. designed study; S.M., C.C., S.T., D.P., W.F., R.S., and Y.S. performed research; S.M., M.A., C.C., S.T., J.A., R.S. and Y.S. analyzed data; S.M., R.S. and Y.S. wrote paper; S.M., M.A., C.C., S.T., M.R., N.K., J.P, D.P., W.F., J.A., R.S., and Y.S. edited paper.

Declaration of interests:

Authors declare no competing interests.

MAIN FIGURES AND LEGENDS

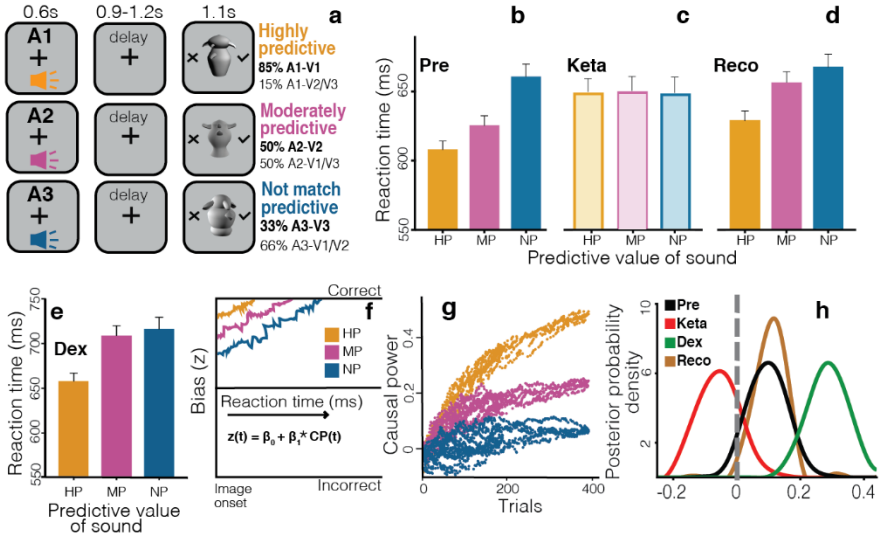

Figure 1. Ketamine blocked fast RTs to predictive sounds. See also Figure S1 and S2.(a) We manipulated subjects' predictions by changing the probability of an image appearing after its associated sound in an audio-visual delayed match-to-sample task. Population RT (+SE) of 15 subjects (b) before (Pre), (c) under (Keta), and (d) after recovery (Reco) from ketamine. (e) Population RT (+SE) of 14 subjects under dexmedetomidine (DEX). (f) HDDM of RTs. Bias (z) calculated for each trial (t) using causal power (CP). β_1 determines relationship between z and CP. (g) Population CP values across time for pre-ketamine (Pre) testing. (H) Posterior probability density of β_1 for different drug conditions.

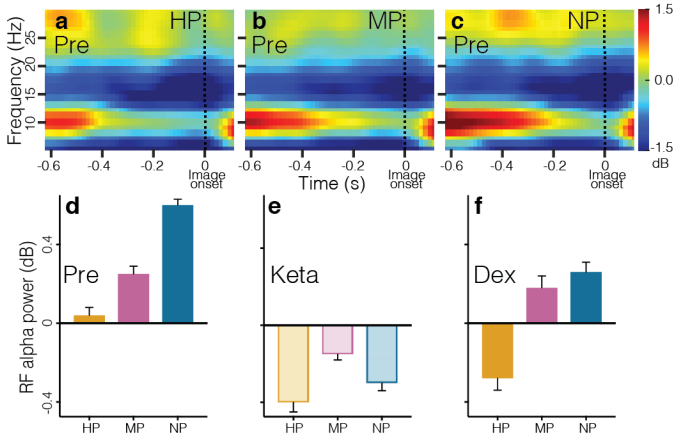

Figure 2. Ketamine blocked correlation between prediction strength and right frontal alpha power. See also Figure S3. Population time-frequency decomposition of right frontal electrode cluster (RF) before drug administration (Pre; a-c), for highly predictive (HP; a), moderately predictive (MP; b), and not-match predictive (NP; c) sounds. Power calculated in 0.55 s sliding windows, with window at 0 s representing interval -0.275 s to +0.275 s. Population average RF alpha power in delay period for (d) Pre, (e) ketamine (Keta) and (f) Dexmedetomidine (Dex).

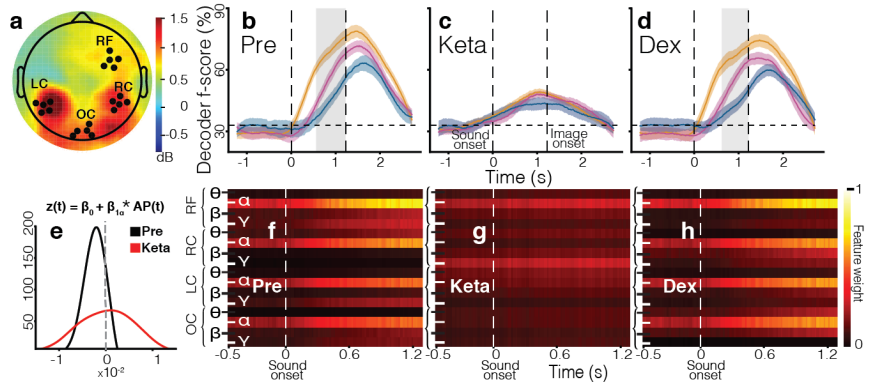

Figure 3. Ketamine blocked decoding of predictions. (a) Electrode clusters in right frontal (RF), right central (RC), left central (LC), and occipital (OC) cortex showing significant modulation in delay period alpha power. Average across all trials (HP, MP and NP). F score using support vector machine (SVM) to decode predictive value (highly predictive, HP; moderately predictive, MP; or not-match predictive, NP) of sound, based on time-frequency power spectrum (b) before drug administration (Pre), (c) under ketamine (Keta), and (d) under dexmedetomidine (Dex). Dashed vertical line at 0 s denotes sound onset. Dashed vertical line at 1.225 s denotes earliest possible time window including image onset. Dashed horizontal line signifies level of decoding expected by chance. Gray shaded areas indicate significant zone of separation between HP, MP and NP, prior to image onset. (e) Posterior probability density of $\beta_{1\alpha}$. Bias (z) calculated for each trial (t) using alpha power (AP). $\beta_{1\alpha}$ determines relationship between z and AP. Feature weights from SVM decoder for (f) Pre, (g) Keta, and (h) Dex. θ, α, β, and γ indicate theta (5-7 Hz), alpha (8-14 Hz), beta (15-30 Hz), and gamma (30-45 Hz) bands respectively.

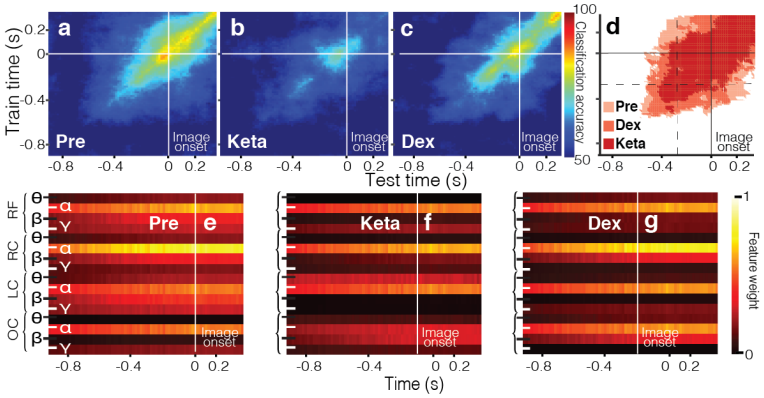

Figure 4. Ketamine perturbed pre-stimulus representations of predicted images. (a-c) Cross-temporal classification accuracy based on the time-frequency power spectrum for match (A2-V2) versus non-match (A2-V1 and A2-V3) trials of the moderately predictive scenario. (a) Before drug administration (Pre), (b) under ketamine (Keta), and (c) under dexmedetomidine (Dex). Vertical and horizontal white solid lines denote image onset. (d) Time periods for Pre in a (light pink), Keta in b (red), and Dex in c (dark pink), showing significantly above chance classification accuracy. White area not significant. Vertical and horizontal solid black lines denote image onset. Considering the wavelet length, the vertical and horizontal dashed black lines denote time before which no stimulus-evoked response could contribute. (e)-(g) Feature importance from decoder contributing to diagonal pixels' accuracy in (a-c) for (e) Pre, (f) Keta, and (g) Dex. θ, α, β, and γ indicate theta (5-7 Hz), alpha (8-14 Hz), beta (15-30 Hz), and gamma (30-45 Hz) bands respectively.

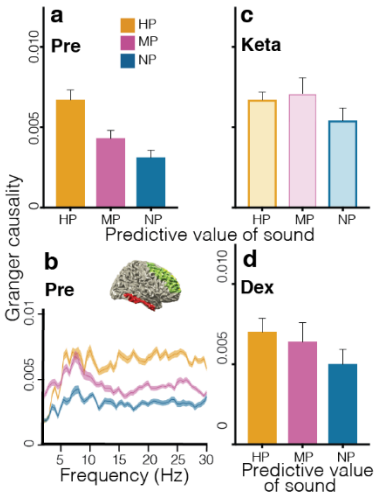

Figure 5. Ketamine perturbed correlation between prediction strength and beta feedback. (a) Population average non-parametric Granger causal influence from right frontal to right inferior temporal cortex in beta band (15-30 Hz) during delay period (starting 900 ms before image onset). Highly predictive, HP; moderately predictive, MP; not-match predictive, NP. (b) Population average non-parametric spectral Granger causal influence during delay period. Inset shows regions of significantly increased source power during delay period compared to baseline ($P < 0.025$, Bonferroni corrected across all cortical AAL regions) after beamforming source localization. Green, superior and middle frontal gyrus; Red, inferior temporal gyrus. (c) and (d) Population average non-parametric Granger causal influence from right frontal to right inferior temporal cortex in beta band during delay period for (c) ketamine (Keta) and (d) dexmedetomidine (Dex).

METHODS

RESOURCE AVAILABILITY

Lead Contact and Materials Availability

This study did not generate new unique reagents. Further information and requests for resources, equipment and experimental methods should be directed to and will be fulfilled by the Lead Contact, Yuri Saalmann (saalmann@wisc.edu).

Data and Code Availability Statement

The datasets/code supporting the current study have not been deposited in a public repository but are available from the corresponding authors on reasonable request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The University of Wisconsin-Madison Health Sciences and Social Sciences Institutional Review Boards (IRBs) approved experiments. 29 participants (14 female) performed the psychophysics predictive coding experiment. We excluded data from four subjects as their performance accuracy was below 50%. 17 additional participants (six female) for dexmedetomidine and 15 participants (five female) for ketamine took part in the pharmacology predictive coding experiments (mean age = 22.35 years; SD = 2.82 years). Of the 15 participants who performed ketamine experiments, the first 12 had to participate in the dexmedetomidine experiments first as per IRB requirements, whereas the final three participated in ketamine experiments first. The different orders of drug experiments yielded similar results. Three participants also performed in saline control experiments, which yielded results similar to the pre-drug baseline. Three participants (from 17 dexmedetomidine participants) were excluded for low accuracy (accuracy less than 50%). Participants who performed the psychophysics predictive coding experiment did

not take part in the pharmacology experiments. We obtained informed consent from all participants.

METHOD DETAILS

Stimuli

For our psychophysics experiments, we used biomorphic visual stimuli from Michael Tarr's lab (http://wiki.cnbc.cmu.edu/Novel_Objects). These are known as greebles. Figure 1a shows examples presented to participants. We used three gray-scale greebles for each psychophysics session, and each greeble was personified with a name. We used novel sounds (trisyllabic nonsense words) for the greeble names – e.g., "Tilado", "Paluti", and "Kagotu" – from Saffran et al. [70]. The sounds were generated using the Damayanti voice in the "text to speech" platform of an Apple MacBook. To avoid differences in the salience of stimuli, greeble images have similar size (13 degrees of visual angle in height and 8 degrees in width), number of extensions and mean contrast, and greeble names have the same number of syllables and sound level (80 dB SPL).

For the pharmacology experiments, we generated three new triplets of greebles. To control for saliency, each participant rated greeble salience for each of four (three plus one from the psychophysics experiment) triplets, i.e., the participant identified whether any of the greebles in a triplet stand out compared to the other two greebles from the same triplet. We proceeded to use triplets which the participant rated all three greebles as being equally salient. We then named each of these greebles with a new trisyllabic nonsense word.

Audio-visual delayed match-to-sample task

Each trial of the task involves the sequential presentation of a sound (trisyllabic nonsense word) followed by a greeble image. We refer to stimuli using the following notation: A1, A2 and A3 correspond to each of the three sounds used (A for auditory) and V1, V2 and V3 correspond to each of the greebles used (V for visual). Using this notation, audio-visual stimulus sequences containing the matching name and greeble are A1-V1, A2-V2 and A3-V3. Audio-visual stimulus sequences containing a non-matching name and greeble are A1-V2, A1-V3, A2-V1, A2-V3, A3-V1 and A3-V2. We pseudo-randomized names for greebles (i.e., matching sounds and images) across subjects.

Learning Phase. During the first phase of the task, participants learn the association between the sounds and images (i.e., names of the greebles) through trial-and-error, by performing a match/non-match (M/NM) task. This phase is called the "learning" phase. Each trial starts with blank blue screen (R=35, G=117, B=208, 200ms duration; shown as gray in Figure 1a). After that, a black fixation cross (size 1.16 degrees of visual angle; jittered 200-400 ms) is presented followed by a sound, a greeble name voiced by the computer (600 ms duration). After a jittered delay period (900-1,200 ms duration), a greeble image (until a participant responds or 1,100 ms duration, whichever is earliest) was presented on the monitor screen, as well as two symbols (√ and **X**) to the left and right of the greeble (9.3 degrees of visual angle from screen center). These symbols indicated participants' two response options: match (√) or non-match (**X**). The symbol location, left or right of the greeble image, corresponded to the left or right response button, respectively: left and right arrow keys of a computer keyboard in the psychophysics experiments; and left and right buttons of a mouse in the pharmacology experiments. We randomly varied the symbols' locations relative to the greeble image to minimize motor preparation (i.e., on some trials, a match response required a left button press and, on other trials, a match response required a right button press). In the learning phase, each greeble

name and image had 33% probability of appearing in any given trial. This is to prevent subjects from developing any differential predictions about the greebles due to greeble name or image frequency, during the learning phase.

To address possible same-different biases, e.g., quicker reaction times (RTs) for match trials [30] we introduced a control called "inversion trials" in psychophysics experiments, to minimize the expectation of M/NM trials which, in itself, might otherwise contribute to participants' responses. In these inversion trials, participants had to respond whether the greeble image presented on screen is inverted (the appropriate response button, left/right, indicated on screen by the left/right location of a red arrow pointing down) or upright (yellow arrow pointing up). Participants did not know the type of trial in advance; the trial type was only revealed by the symbols to the left and right of screen center at the onset of the greeble image (i.e., √ and **X** signal M/NM trials, whereas downward red arrow and upward yellow arrow signal inversion trials). 50% of the total number of trials in the learning phase were inversion trials and the rest were M/NM trials. Because participants cannot specifically prepare in advance for M/NM trials due to the random presentation of inversion and M/NM trials, there should be minimal confounding of RTs with a bias towards match responses.

Testing phase. Once participants show above 80% accuracy for the M/NM trials in the learning phase of the task, they move on to the "testing phase" (1,000 trials for the psychophysics experiments; Figure 1a). During the testing phase, we manipulated predictions by changing the probability of a greeble appearing after its learnt name. This probability is different for each greeble name and image. That is, in the testing phase, when a participant hears A1, there is 85% chance of V1 being shown (highly predictive; HP); when a participant hears A2, there is 50% chance of V2 being shown (moderately predictive; MP); and when a participant hears A3, there is a 33% chance of V3 being shown (not-match predictive; NP). This allows participants to

make stronger predictions about the identity of the upcoming visual image after hearing A1, than after hearing A2 or A3, for instance.

Inversion trials also consisted of half the total trials in the testing phase of the psychophysics experiments. We randomly presented all the trial types (M/NM trials and inversion trials) to the participants. The testing phase of the task had approximately equal match and non-match trials to avoid response bias. The testing phase also had an approximately equal number of trials for each greeble image, and its corresponding name had approximately equal probability of being voiced, to control for stimulus familiarity.

Causal Strength and Transitional Probability

We quantified the relationship between a sound (name) and its paired image (greeble) using the strength of causal induction: given a candidate cause *C* (sound) how likely is the effect *E* (i.e., how likely is it followed by its paired image). We will represent variables *C* and *E* with upper case letters, and their instantiations with lower case letters. Hence, $C = c+ / E = e+$ indicates that the cause/effect is present, and $C = c- / E = e-$ indicates that the cause/effect is absent (for brevity, we will shorten variables equal to outcomes, such as $C = c+$ or $C = c-$ as simply $c+$ or $c-$, respectively). The evidence for a relationship can be encoded as a 2 X 2 contingency table for each sound, as in Table S1 (black letters), where $\mathrm{N(c+, e+)}$ represents the number of trials in which the effect occurs in the presence of the cause, $\mathrm{N(c-, e+)}$ represents the number of trials in which the effect occurs in the absence of the cause and so on. Applied to our study, e.g., *C* could be hearing sound A1, and *E* viewing the paired greeble V1. For this case, $\mathrm{N(c+, e+)}$ would be the number of trials V1 follows A1; whereas $\mathrm{N(c-, e+)}$ would be the number of trials V1 follows A2 or A3. The full contingency table for the “Highly Predictive” auditory cue A1 and its paired greeble V1 is shown in Table S1 (in green letters). There are analogous contingency tables for the other two auditory cues and their paired greebles.

Based on these contingency tables, we calculated three different measures of causal relationship (ΔP, causal power and causal support) for each trial. ΔP and causal power assume that *C* causes *E*. ΔP reflects how the probability of *E* changes as a consequence of the occurrence of the *C*. Causal power corresponds to the probability that an effect *E* happened because of cause *C* in the absence of all other causes. Whereas causal support evaluates whether or not a causal relationship actually exists and calculates the strength of that relationship. To do this, causal support estimates the evidence for a graphical model with a link between *C* and *E* against one without a link. For example, let us consider the graphs denoted by Graph 0 and Graph 1 in Figure S1d (adapted from [31]). There are three variables in each graph: cause *C*, effect *E* and background cause B. In Graph 0, *B* causes *E*, but *C* has no relationship to either *B* or *E*. In Graph 1, both *B* and *C* cause *E*. While calculating ΔP and causal power Graph 1 is assumed, whereas causal support compares the structure of Graph 1 to that of Graph 0. Causal support is defined as the evidence provided from data D in favor of Graph 1, P(D | Graph 1), over Graph 0, P(D | Graph 0), which can be calculated by the following equation:

$$Causal\ support = \log \frac{\mathrm{P(D|Graph\ 1)}}{\mathrm{P(D|Graph\ 0)}} \qquad (1)$$

We calculated causal support using freely available Matlab code from [31]. ΔP and causal power were calculated using the following formulas:

$$\Delta P = \frac{N(c+,e+)}{N(c+,e+)\ +\ N(c+,e-)} - \frac{N(c-,e+)}{N(c-,e+)\ +\ N(c-,e-)} = P(e+|c+) - P(e+|c-) \quad (2)$$

$$\textit{Causal power} = \frac{\Delta P}{1-P(e+|c-)} \qquad (3)$$

We compared these three measures of causal relationship with the transitional probability (i.e., a comparison between causation and correlation). We calculated the transitional probability of each greeble (V) given prior presentation of its paired sound (A), using the equation below:

$$\text{Transitional probability} = \frac{N(c+,e+)}{N(c+)} \qquad (4)$$

For each subject, we used the causal relationship value of each condition (HP, MP and NP) at the end of a testing phase as the starting values of the next testing phase. For example, the starting values of causal relationship for the "under drug" testing phase were equal to the causal relationship values at the end of the "pre/baseline" testing phase. Similarly, the starting values for the "after recovery" testing phase were equal to the causal relationship values at the end of the "under drug" phase (Figure S2a and b). We tested if subjects (i) regained access to already learned and stored predictive information, after they recovered from ketamine dosing, or (ii) re-learned the predictive information. We mimicked hypothesis (ii) by forcing the starting values of causal relationship (causal power here) "after recovery" to be zero (Figure S2c) instead of starting values equal to the causal relationship at the end of the "under drug" phase (hypothesis (i); Figure S2b).

Pharmacology Experiments

To manipulate participants' predictions, we administered two drugs, ketamine and dexmedetomidine, each on a separate day, with at least one month intervening. The fixed order of dexmedetomidine in the first session, and ketamine in the second session, for the first 12 ketamine subjects was IRB-imposed in their consideration of safety profiles of the different medications (registered on NCT03284307); whereas the final three ketamine subjects were not administered dexmedetomidine in a prior session. All subjects were healthy and aged between

18-40 years old without contraindication to study drugs. We also acquired EEG data throughout pharmacology experiments (see section “EEG Recording” below), to measure electrophysiological activity during predictive coding. A typical pharmacology experiment consisted of three segments: (a) pre-drug baseline, (b) under drug influence (dexmedetomidine or ketamine), and (c) after recovery. During the pre-drug baseline, participants performed the learning phase (200 trials), then the first testing phase (400 trials). Under drug influence, participants performed the second testing phase (200 trials). Under ketamine, participants also performed a third testing phase (200 trials; see ketamine dosing for details). After recovery, they performed the last testing phase (400 trials). As a drug control, we tested three participants during saline administration. The saline results were similar to the pre-drug baseline. Due to a protocol-limited maximum time under drug influence, and the need to acquire sufficient M/NM trials for EEG analyses, the pharmacology experiments did not include inversion trials. All other aspects of the task in pharmacology experiments were the same as that in psychophysics experiments. For each of the two pharmacology experiments involving a particular participant, we used three new greeble name and image pairs to rule out any possible contribution of long-term memory.

Dexmedetomidine Dosing

We intravenously administered a 0.5 mcg/kg bolus over 10 minutes, followed by 0.5 mcg/kg/h infusion (MedFusion4000 pump at mcg/kg/hr). Participants performed the testing phase (under drug influence) during this infusion time, corresponding to stable drug levels according to the pharmacokinetic model for dexmedetomidine by Hannivoort et al. [71]. We targeted a plasma concentration of dexmedetomidine that is associated with mild sedation (modified observer’s assessment of alertness/sedation (OAA/S) of 4 [72]) to control for non-specific sedative effects, including hyperpolarization-activated cyclic nucleotide channel (HCN-1) effects [73]. The actual

sedation achieved was on average slightly deeper than anticipated (modified OAA/S median 3, IQR 2) with a mean plasma concentration of 0.8 (SD 0.33) ng/ml [72].

Ketamine Dosing

In initial experiments, we tested two doses of ketamine to target the lowest plasma concentration that would modulate NMDARs in the relevant concentration range (<1microMolar [74]). The first dose corresponded to intravenously administered 0.25 mg/kg ketamine over 10 minutes, followed by 30 mg/h infusion, corresponding to 0.4 microMolar (Rugloop software using a Harvard 22 pump). Prior to testing subjects, lack of nystagmus and visual disturbance was confirmed in all participants. Participants performed the testing phase (under drug influence) during this infusion time, corresponding to stable drug levels according to the pharmacokinetic model for ketamine by Domino et al. [75,76]. A second dose was tested with a second bolus of 0.25 mg/kg ketamine over 10 minutes, followed again by 30 mg/h infusion. Testing again was completed once stable plasma concentrations of approximately 0.8 microMolar were achieved. Ketamine blocked predictions at this second level of ketamine dosing, equating to a minimum plasma concentration of 0.2 μg/ml; range tested: 0.2-0.3 μg/ml. As we found the effective ketamine dose to be 0.2 μg/ml in our first seven subjects, we targeted that plasma concentration for our remaining subjects. We report data for the 0.2 μg/ml dosing in the manuscript. Three subjects were excluded from “after recovery” testing due to vomiting.

Monitoring

Subjects were monitored during drug exposure according to the American Society of Anesthesiologists guidelines, including electrocardiogram, blood pressure and oxygen saturation. We monitored arousal level using the modified OAA/S scale [77].

EEG Recording

We performed high-density EEG recordings using a 256 channel system (including NA 300 amplifier; Electrical Geodesics, Inc., Eugene, OR). After applying the EEG cap with conductive gel (ECI Electro-Gel), we adjusted electrodes so that the impedance of each electrode was within 0-50 kiloohms. We checked electrode impedance before the experiment started, and again before drug administration. Using Net Station, we sampled EEG signals at 250 Hz and, off-line, bandpass filtered between 0.1 Hz and 45 Hz.

EEG Preprocessing

We combined pre-drug baseline data from both ketamine and dexmedetomidine experiments (baseline RT data showed similar results for both drugs) but "under drug" analyses were performed separately for each drug. We performed offline preprocessing and analysis using EEGLAB [78]. First, we extracted data epochs -1,500 ms to 3,000 ms relative to the onset of the sound and -3000 ms to 800 ms relative to the onset of the image, for each trial. We then visually inspected each epoch and excluded noisy trials (around 5% of the total trials). Next, we performed Independent Component Analysis (ICA) using built-in functions of EEGLAB (pop_runica.m) and removed noisy components through visual inspection. We excluded three dexmedetomidine subjects from further analysis– due to very noisy EEG data, which after cleaning left too few trials for analysis (conditions with <10 trials). Finally, we performed channel interpolation (EEGLAB function, eeg_interp.m, spherical interpolation) and re-referenced to the average reference.

Time-Frequency Decomposition

To investigate changes in EEG spectral content, we performed time-frequency decompositions of the preprocessed data in sliding windows of 550 ms using Morlet wavelets, whose frequency

ranged from 5 Hz to 45 Hz in 40 linearly spaced steps. Power for each time-frequency point is the absolute value of the resulting complex signal. We dB normalized power (dB power = 10*log10[power/baseline]) to the pre-stimulus baseline, starting 700 ms before sound onset (i.e., baseline calculated using sliding 550 ms-long wavelets, starting with the wavelet positioned from 700-150 ms before sound onset (centered on 425 ms before sound onset) and avoiding the sound-evoked response). For drift-diffusion model analysis, we calculated power spectral density and performed divisive baseline correction for each trial.

Electrode Selection

We used a data-driven approach, orthogonal to the effect of interest, to select the electrodes of interest based on the task EEG data. In the first step, we averaged power across all electrodes (aligned to image onset) and all sounds/greeble names. This revealed increased alpha power (8-14 Hz) during the delay period compared to baseline (Figure S3b). In the second step, we selected the electrode clusters that showed significant change in alpha power during the -925 ms to -275 ms time window pre-image onset, compared to baseline. This time window ensured that our delay period did not overlap with the sound or image. After cluster-based multiple comparisons correction (*6*), four different clusters showed significant modulation in alpha power during the delay period (Figure 3a): (i) right frontal (RF electrodes 4, 214, 215, 223, 224); (ii) left central (LC electrodes 65, 66, 71, 76, 77); (iii) right central (RC electrodes 163, 164, 173, 181, 182); and (iv) occipital (OC electrodes 117, 118, 119, 127, 129).

Alpha Power Calculation

For each condition, HP, MP and NP, we averaged alpha power over all five electrodes in a cluster, to calculate the average alpha power of each cluster. To best capture the delay period activity just prior to image onset, we calculated mean alpha power between -625 to -275ms (as

the wavelet window is centered around each time point, the power estimate before -625 ms and after -275 ms may contain auditory and visual stimulus-related responses, respectively) for each trial aligned to image onset. We also calculated mean delay period alpha power between 600 to 1,225 ms for each trial aligned to sound onset. To link EEG power spectral density to behavior using our drift-diffusion model analysis (see section HDDM), we calculated single-trial baseline-corrected (divisive normalization) alpha band power, aligned to image onset.

Source Space Analysis

We used FieldTrip's beamforming technique to localize sources of the sensor level alpha activity [79]. This technique uses an adaptive spatial filter to estimate activity at a given location in brain. Across all sound cues, we used the Dynamic Imaging of Coherent Sources (DICS) [80] algorithm to beamform the delay period alpha activity (window 0-900 ms before image onset). We then calculated the average source power for each region of interest (ROI) of the AAL atlas [81]. We selected ROIs that showed significant change in source power during the delay period (prior to image onset) compared to baseline ($P < 0.025$, Bonferroni corrected across all cortical AAL ROIs).

Granger Causality

We performed non-parametric spectral Granger causality analyses at the source level. Because there were no significant frontal ROIs in the left hemisphere, we restricted our analyses to the right hemisphere only. We used a covariance window of 0-900 ms prior to image onset and the Linearly Constrained Minimum Variance (LCMV) algorithm[82] to generate virtual time series for significant frontal and temporal ROIs. We averaged across the significant frontal ROIs (Superior Frontal Gyrus, medial; Superior Frontal Gyrus, dorsolateral; Middle Frontal Gyrus); and there was only one significant temporal lobe ROI (Inferior Temporal Gyrus). We calculated non-

parametric spectral granger causality between the (average) frontal and temporal ROIs using the ft_connectivityanalysis function in FieldTrip for the stable window starting 900 ms prior to image onset [83,84].

Hierarchical Drift-Diffusion Model (HDDM)

We used a drift-diffusion model (DDM) [85], where there are two possible choices (correct/incorrect responses of the match trials) in our predictive coding task. According to this model, decision-making involves the accumulation of evidence (drift process) from a starting point to one of two (upper or lower) thresholds, representing the choices. The accumulation rate is known as the drift rate, v; and the starting point can be biased towards one of the choices (in our study, by the predictive value of the sound), reflected in a bias parameter, z. We used HDDM software (http://ski.clps.brown.edu/hddm_docs/) [29] for hierarchical Bayesian estimation of the parameters of the drift-diffusion model. The hierarchical approach allows estimates of the group level and individual subject level parameters simultaneously. Particularly with fewer trials per condition, this method has been shown to provide more reliable estimates of parameters and is less susceptible to outliers [85] than more traditional approaches to DDMs [86,87] .

To directly link the causal relationship between the sound and its paired greeble image to behavior and drift-diffusion parameters, we included the estimates of causal relationship and transitional probability as predictor variables of the bias, z, of the model. That is, we estimated posterior probability densities not only for basic model parameters, but also the degree to which these parameters are altered by variations in the psychophysical measures (ΔP, causal power, causal support and transitional probability). In these regressions, the bias parameter is given by, $z(t) = \beta_0 + \beta_1 CP(t)$, where CP is either ΔP, causal power, causal support or transitional

probability, β_0 is the intercept, and t is the trial number. Here, the slope, β_1, is weighted by the value of the psychophysical measure on that trial. The regression across trials allows us to infer how the bias changes depending on the psychophysical measure. For example, if these psychophysical measures are positively correlated to bias, then increased causal strength or transitional probability will yield faster RTs. We fit four different versions of the model: (i) ΔP Model, where bias was estimated from the ΔP and updated after each trial; (ii) Causal Power Model, where bias was estimated from the causal power and updated after each trial; (iii) Causal Support Model, where bias was estimated from the causal support and updated after each trial; and (iv) Transitional Probability Model, where bias was estimated from the transitional probability and updated after each trial. We also modeled our data where drift rate varied according to causal strength or transitional probability. We used the Deviance Information Criterion (DIC) for model comparison [88]. The DIC is a measure of model fit (i.e., lack thereof) with a penalty for complexity (i.e., the number of parameters used to fit the model to the data) [89]. Models with lower DIC are better models. Models where bias was manipulated instead of drift rate had significantly lower DIC. For the rest of our analyses, we used models where bias varied with causal strength or transitional probability and drift rate was kept constant at the default values, as this yielded the lowest DIC.

We used Markov Chain Monte Carlo chains with 20,000 samples and 5000 burn-in samples for estimating the posterior distributions of the model parameters. We assessed chain convergence by visually inspecting the autocorrelation distribution, as well as by using the Gelman-Rubin statistic, which compares between-chain and within-chain variance. This statistic was near 1.0 for the parameters, indicating that our sampling was sufficient for proper convergence. We analyzed parameters of the best model (model with lowest DIC) using Bayesian hypothesis testing, where the percentage of samples drawn from the posterior fall within a certain region

(e.g., > 0). Posterior probabilities ≥ 95% were considered significant. Please note that this value is not equivalent to p-values estimated by frequentist methods, but they can be coarsely interpreted in a similar manner.

All the model comparisons were estimated on the psychophysics data as these had the greatest number of trials per condition. This ensures robust estimation of the best model. The best-fitting model was then used to analyze data from different conditions: pre-drug baseline, under dexmedetomidine, under ketamine and recovery from ketamine. To directly link EEG power spectral density to behavior and drift-diffusion parameters, we used the HDDM, but now included the right frontal cluster power estimate (aligned to image onset) in the alpha band as the predictor variable of the bias, *z*, of the model; i.e., in the regression equation above, CP was now alpha power.

Event-related Potentials

We used ERPLAB (https://erpinfo.org/erplab) [90] to run event-related potential (ERP) analysis. First, we cleaned epoched data aligned to the sound onset for auditory ERPs using the pop_artmwppth.m function of ERPLAB with a moving window of 200 ms (2-3% of trials for each subject were excluded). We then averaged over trials to generate an average ERP for each subject. Based on previous literature [91–94], we chose channel 9 (Cz electrode) for auditory ERPs. For auditory ERPs, we calculated the N200 response (negative peak amplitude between 200-300 ms) for each sound.

EEG Signal Decoding

We used the power spectral density of EEG signals across four frequency bands, theta (5-7 Hz), alpha (8-14Hz), beta (15-30 Hz) and gamma (31-40 Hz), for the decoding analysis. We calculated the sum of squared absolute power in each frequency band for each electrode

cluster (RF, LC, RC and OC), thus generating 16 features for each trial as the input dataset. Using Scikit-learn [95] implemented in Python, for each 20 ms time bin, we trained a Support Vector Machine (SVM) model to classify the EEG data into three classes: highly predictive (HP), moderately predictive (MP) and not-match predictive (NP). We denote $y_i(t) \in \{0,1,2\}$ as the identifier of the condition at time bin t, where 0, 1, and 2 denotes HP, MP and NP respectively. The SVM classifier is implemented by a nonlinear projection of the training data $\mathbf{x}(t)$ feature space $\mathcal{X}$ into a high dimensional feature space $\mathcal{F}$ using a kernel function ϕ. So with $\phi: \mathcal{X} \longrightarrow \mathcal{F}$ being the mapping kernel, the weight vector $\mathbf{w}$ can be expressed as a linear combination of the training trials and the kernel trick can be used to express the discriminant function as

$$y(\mathbf{x}(t); \zeta(t)) = \mathrm{a}^T(t)\phi_{\mathbf{x}}(t) + b(t) = \sum_{n=1}^{N} a_n(t)\varphi(\mathbf{x}_n(t), \mathbf{x}(t)) + b(t) \qquad (5)$$

where $\zeta(t) = \{\mathrm{a}(t), b(t)\}$ is the new parameter at time bin t with $\mathrm{a}(t)$ and $\mathrm{b}(t)$ as weights and biases of the mapped features space $\mathcal{F}$. We used the radial basis function (RBF) kernel that allows nonlinear decision boundary implementation in the input space. The RBF kernel holds the elements

$$\varphi\big(\mathbf{x}_n(t), \mathbf{x}(t)\big) = \exp(-\eta(t)\|\mathbf{x}_n(t) - \mathbf{x}(t)\|^2) \qquad (6)$$

where $\eta(t)$ is a tunable parameter at each time bin. Model hypermeters consisting of regularization penalty ($C(t)$) and $\eta(t)$ were selected by grid search through 10-fold cross validation. F-score at each time bin and for each label was calculated as

$$F - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (7)$$

where

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (8)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (9)$$

As we penalized the mapped weights of the classifier at each time bin, we used normalized absolute values of the weights as a measure to deduce each feature's contribution to classify the outputs.

For the decoding of match (A2-V2) versus non-match (A2-V1 and A2-V3) trials of the moderately predictive scenario, due to the limited number of trials for each label in each condition, we used a decision tree classifier (CART) [96] with maximum depth of 16 and Gini as the impurity measure, to decide the optimal split from a root node, and subsequent splits.

Considering $\mathbf{x}(t) \in R^N$ as feature space and $y(t) \in \{0,1\}$ as label space, the decision tree recursively partitions the space such that the samples with the same labels are grouped together. Let the data at node m be represented by Q . For each candidate split $\theta = (j, t_m)$ consisting of a feature j and threshold t_m , data will be partitioned into $Q_{left}(\theta)$ and $Q_{right}(\theta)$ subsets where

$$Q_{left}(\theta) = \forall(x, y)|x_j \leq t_m$$

$$Q_{right}(\theta) = Q_{right}(\theta)\backslash Q$$

The impurity at node m can be calculated as below

$$G(Q, \theta) = \frac{n_{left}}{N_m} H\big(Q_{left}(\theta)\big) + \frac{n_{right}}{N_m} H(Q_{right}(\theta))$$

And the parameter set at node *m* can be derived as $\theta^* = \underset{\theta}{\operatorname{argmin}}\, G(Q, \theta)$. These steps will be recursed until the maximum allowable depth of the tree is reached. This process will cause the feature space to be partitioned into M regions, namely $R_1, R_2, \ldots, R_m$, and considering N_m as the number of data points in region R_m (represented by node *m*), the proportion of class *k* data points in node m can be calculated as $\hat{p}_{mk} = \frac{1}{N_m}\sum_{x_i \in R_m} \mathbb{I}(y_i = k)$ and the Gini function at that node can be defined as

$$H(X_m) = \sum_k \hat{p}_{mk}(1 - \hat{p}_{mk})$$

We used XGBoost (both tree and gradient boosting) to increase the performance of the decision tree ensemble. For cross-temporal decoding, we trained the boosted CART classifier with the data for one time period (indicated on the y axis in Figure 4 a-c) and tested it with data from a different time period (indicated on the x axis). We imposed regularization on the model to avoid overfitting. To determine the contribution of electrode clusters to the decoding, we used normalized Mean Decrease Impurity (MDI) as a measure of feature importance. MDI is defined as the total decrease in node impurity (weighted by the probability of reaching that node) averaged over all trees of the ensemble.

STATISTICAL ANALYSIS

We performed statistical analysis of trials from testing phases, and used the learning phase only to confirm that the participants learned the correct associations. To only include the trials where the causal power of HP, MP and NP trials have differentiated (Figure 1g), we excluded the first 50 trials of the testing phase in psychophysics experiments and the pre-drug baseline. We used all available trials for the testing phase of “under drug” and “after recovery”, as the causal power of HP, MP and NP trials were already differentiated from the beginning (Figure S2, a and b). For

RT analysis, we excluded RTs more/less than the mean +/- 3SD for each subject, and we used correct match trials. For delay period EEG analyses, we used both correct match and non-match trials. The data were analyzed using a linear mixed effect model with the predictive value of sound (HP, MP or NP; varying within subject) and drug condition (before drug or under drug; drug either ketamine or dexmedetomidine; varying within subject) as the independent variables. Both main effects of prediction and drug as well as their interaction were included. We used contrast analysis to model the three categories of sounds. We used a linear (-1, 0, 1) contrast as our contrast of interest, and a quadratic (-1, 2, -1) contrast as the orthogonal contrast in the analysis. We used the same linear mixed effect model only changing the dependent variable with respect to hypotheses; i.e., RTs, accuracy (correct versus incorrect), average right frontal alpha power (aligned to image onset), or average GC at beta frequency. To statistically test N200 amplitude in the auditory ERP, we used a similar linear mixed effects model with the predictive value of sound as our only predictor variable. Our study conformed to the guidelines set out by Ableson and Prentice [97] with regards to contrast analysis; i.e., we included the contrast of interest along with a paired, orthogonal contrast. Our statistical tests showed that contrasts of interest were significant while the orthogonal contrasts were not.

We ran confirmatory linear regression analyses to show that the effects of ketamine differed from both before ketamine and after recovery. To demonstrate that the linear trend in RTs vanishes under ketamine, in comparison to before ketamine and after recovery, we ran a repeated measure regression analysis for prediction (HP, MP, NP) X drug condition (before, under, after ketamine). For prediction, we used a linear (-1, 0, 1) contrast as our contrast of interest, and a quadratic (-1, 2, -1) contrast as the orthogonal contrast in the analysis. For drug condition, we used a quadratic (-1, 2, -1) contrast as our contrast of interest, and a linear (-1, 0, 1) contrast as the orthogonal contrast in the analysis. We also ran confirmatory analyses to

show that the correlation between predictive strength and either RT, RF alpha power or Granger causal influence of right frontal on right inferior temporal in the beta band vanishes under ketamine, in comparison to that before ketamine and under DEX. We ran a repeated measures regression analysis for prediction (HP, MP, NP) X drug condition (before drug, under ketamine, under DEX). For prediction, we used a linear (-1, 0, 1) contrast as our contrast of interest, and a quadratic (-1, 2, -1) contrast as the orthogonal contrast in the analysis. For drug condition, we used a quadratic (-1, 2, -1) contrast as our contrast of interest, and a linear (-1, 0, 1) contrast as the orthogonal contrast in the analysis.

We used repeated measures ANOVA (Holm-Sidak corrected p-values) to test the significance of F-scores for the three differentially predictive conditions (HP, MP, NP) as well as the significance of each feature's contribution in output classification.

To test the significance of the cross-temporal decoding results, we performed a paired t-test between the 20-fold cross validated and 100 times resampled accuracy at each pixel and randomly permuted output labels model at the same pixel. The resulting p-values then were corrected for multiple comparisons (Holm-Sidak corrections).

SUPPLEMENTAL INFORMATION

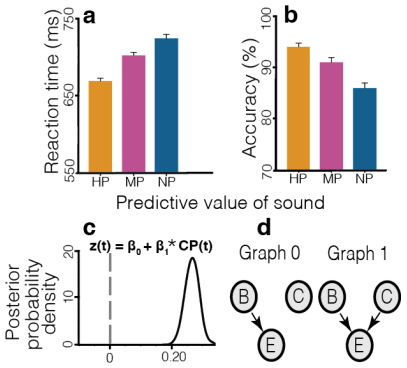

Figure. S1. Fast RTs to predictive sounds not due to speed-accuracy trade off. Related to Figure 1. Population average (25 subjects from psychophysics experiment) (a) RT and (b) accuracy for highly predictive (HP), moderately predictive (MP), and not-match predictive (NP) sounds. (c) Posterior probability density of β_1 of hierarchical drift diffusion model from psychophysics experiment. (d) Directed graphs for calculating causal support. In graph 0, B causes E, but C has no relationship to either B or E. In graph 1, both B and C cause E (adapted from [31]).

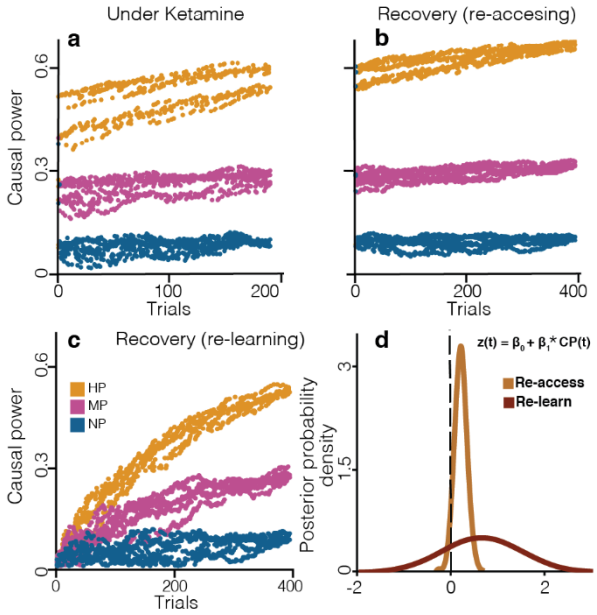

Figure S2. Ketamine blocked access to predictive information. Related to Figure 1. (a) Population causal power values for subjects under ketamine for highly predictive (HP), moderately predictive (MP), and not-match predictive (NP) sounds. (b) Population causal power values after recovery when subjects regained access to predictive information. (c) Population causal power values for hypothetical condition where subjects re-learned predictive information. (d) Positive posterior probability density of β_1 of hierarchical drift diffusion model for first 30 trials after recovery from ketamine, when subjects re-accessed predictive information (P=0.03). In contrast, the posterior probability density of β_1 for re-learning was not greater than zero (P=0.20).

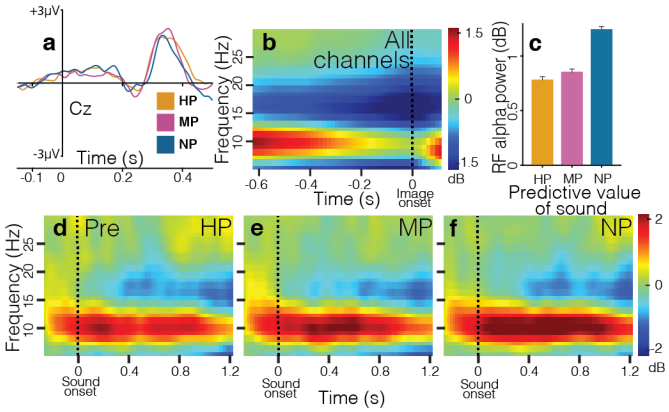

Figure S3. Early sensory processing differences cannot account for prediction strength as all three sounds generated similar auditory ERPs. Related to Figure 2. (a) Auditory ERPs at Cz electrode for highly predictive (HP), moderately predictive (MP) and not-match predictive (NP) sounds. Linear contrast of N200 for HP, MP and NP not significant (P=0.88). (b) Time-frequency plot of power averaged over all electrodes and all trials. Power calculated in 0.55 s sliding windows, with window at 0 s representing interval -0.275 s to +0.275 s. Plot aligned to image onset. (c) Population average RF alpha power (+SE) during delay period; aligned to sound onset. Time-frequency decomposition of right frontal electrode cluster (RF) before drug administration (Pre) for HP (d), MP (e) and NP (f) sounds. Power calculated in 0.55 s sliding windows, with window at 0 s representing interval -0.275 s to +0.275 s. Plots aligned to sound onset.

	Effect Present (e^+) *Effect Present (V)*	Effect Absent (e^-) *Effect Absent*
Cause Present (c^+) *Cause Present (A)*	$N(c^+,e^+)$ *A1-V1*	$N(c^+,e^-)$ *A1-V2, A1-V3*
Cause Absent (c^-) *Cause Absent*	$N(c^-,e^+)$ *A2-V1, A3-V1*	$N(c^-,e^-)$ *A2-V2, A2-V3, A3-V2, A3-V3*

Table S1. 2 X 2 contingency table for each sound. Related to Figure 1. Generic contingency table for all sounds in black. $N(c^+,e^+)$ represents the number of trials in which the effect occurs in the presence of the cause, $N(c^-,e^+)$ represents the number of trials in which the effect occurs in the absence of the cause, $N(c^+,e^-)$ represents the number of trials in which the cause occurs but not the effect, and $N(c^-,e^-)$ represents the number of trials in which the cause and effect are absent. 'Cause' used in statistical sense. In green, example contingency table for sound A1 where N(c+,e+) are the number of trials V1 follows A1 (A1-V1); whereas N(c-,e+) would be the number of trials V1 follows A2 or A3 (A2-V1 or A3-V1) and so on.

REFERENCES

1. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nat Neurosci* **2**, 1019–25 (1999).
2. Serre, T., Oliva, A. & Poggio, T. A feedforward architecture accounts for rapid categorization. *PNAS* **104**, 6424–6429 (2007).
3. Summerfield, C. *et al.* Predictive codes for forthcoming perception in the frontal cortex. *Science* **314**, 1311–4 (2006).
4. Körding, K. P. & Wolpert, D. M. Bayesian integration in sensorimotor learning. *Nature* **427**, 244–247 (2004).
5. Lange, F. P. de, Heilbron, M. & Kok, P. How Do Expectations Shape Perception? *Trends in Cognitive Sciences* **22**, 764–779 (2018).
6. Dayan, P., Hinton, G. E., Neal, R. M. & Zemel, R. S. The Helmholtz machine. *Neural Comput* **7**, 889–904 (1995).
7. Spratling, M. W. A review of predictive coding algorithms. *Brain Cogn* **112**, 92–97 (2017).
8. Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* **2**, 79–87 (1999).
9. Mumford, D. On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern* **66**, 241–51 (1992).
10. Friston, K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci* **11**, 127–38 (2010).
11. Hobson, J. A. & Friston, K. J. Waking and dreaming consciousness: Neurobiological and functional considerations. *Progress in Neurobiology* **98**, 82–98 (2012).
12. Seth, A. K., Suzuki, K. & Critchley, H. D. An Interoceptive Predictive Coding Model of Conscious Presence. *Front. Psychol.* **2**, (2012).

13. Hobson, J. A. & Friston, K. J. Consciousness, dreams, and inference: The Cartesian theatre revisited. *Journal of Consciousness Studies* **21**, 6–32 (2014).
14. Issa, E. B., Cadieu, C. F. & DiCarlo, J. J. Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. *Elife* **7**, (2018).
15. Gordon, N., Koenig-Robert, R., Tsuchiya, N., van Boxtel, J. J. & Hohwy, J. Neural markers of predictive coding under perceptual uncertainty revealed with Hierarchical Frequency Tagging. *Elife* **6**, (2017).
16. Egner, T., Monti, J. M. & Summerfield, C. Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream. *J. Neurosci.* **30**, 16601–16608 (2010).
17. Friston, K. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* **360**, 815–36 (2005).
18. Homayoun, H. & Moghaddam, B. NMDA receptor hypofunction produces opposite effects on prefrontal cortex interneurons and pyramidal neurons. *J Neurosci* **27**, 11496–500 (2007).
19. Homayoun, H. & Moghaddam, B. Orbitofrontal cortex neurons as a common target for classic and glutamatergic antipsychotic drugs. *Proc Natl Acad Sci U S A* **105**, 18041–6 (2008).
20. Rosch, R. E., Auksztulewicz, R., Leung, P. D., Friston, K. J. & Baldeweg, T. Selective Prefrontal Disinhibition in a Roving Auditory Oddball Paradigm Under N-Methyl-D-Aspartate Receptor Blockade. *Biol Psychiatry Cogn Neurosci Neuroimaging* **4**, 140–150 (2019).
21. Murray, J. D. *et al.* Linking microcircuit dysfunction to cognitive impairment: effects of disinhibition associated with schizophrenia in a cortical working memory model. *Cereb Cortex* **24**, 859–72 (2014).

22. Rosier, A. M., Arckens, L., Orban, G. A. & Vandesande, F. Laminar distribution of NMDA receptors in cat and monkey visual cortex visualized by [3H]-MK-801 binding. *J Comp Neurol* **335**, 369–80 (1993).
23. Self, M. W., Kooijmans, R. N., Super, H., Lamme, V. A. & Roelfsema, P. R. Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proc Natl Acad Sci U S A* **109**, 11031–6 (2012).
24. Zorumski, C. F., Izumi, Y. & Mennerick, S. Ketamine: NMDA Receptors and Beyond. *J. Neurosci.* **36**, 11158–11164 (2016).
25. Javitt, D. C., Steinschneider, M., Schroeder, C. E. & Arezzo, J. C. Role of cortical N-methyl-D-aspartate receptors in auditory sensory memory and mismatch negativity generation: implications for schizophrenia. *Proc Natl Acad Sci U S A* **93**, 11962–7 (1996).
26. Schmidt, A. *et al.* Mismatch negativity encoding of prediction errors predicts S-ketamine-induced cognitive impairments. *Neuropsychopharmacology* **37**, 865–875 (2012).
27. Corlett, P. R., Honey, G. D. & Fletcher, P. C. Prediction error, ketamine and psychosis: An updated model. *J Psychopharmacol* **30**, 1145–1155 (2016).
28. Chen, X., Shu, S. & Bayliss, D. A. HCN1 channel subunits are a molecular substrate for hypnotic actions of ketamine. *J. Neurosci.* **29**, 600–609 (2009).
29. Wiecki, T. V., Sofer, I. & Frank, M. J. HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Front Neuroinform* **7**, 14 (2013).
30. Lupyan, G. & Thompson-Schill, S. L. The evocative power of words: activation of concepts by verbal and nonverbal means. *J Exp Psychol Gen* **141**, 170–86 (2012).
31. Griffiths, T. L. & Tenenbaum, J. B. Structure and strength in causal induction. *Cogn Psychol* **51**, 334–84 (2005).

32. Jensen, O. & Mazaheri, A. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front Hum Neurosci* **4**, 186 (2010).
33. Lange, J., Oostenveld, R. & Fries, P. Reduced occipital alpha power indexes enhanced excitability rather than improved visual perception. *J Neurosci* **33**, 3212–20 (2013).
34. Ma, L., Skoblenick, K., Seamans, J. K. & Everling, S. Ketamine-Induced Changes in the Signal and Noise of Rule Representation in Working Memory by Lateral Prefrontal Neurons. *J. Neurosci.* **35**, 11612–11622 (2015).
35. Skoblenick, K. & Everling, S. NMDA antagonist ketamine reduces task selectivity in macaque dorsolateral prefrontal neurons and impairs performance of randomly interleaved prosaccades and antisaccades. *J. Neurosci.* **32**, 12018–12027 (2012).
36. Kok, P., Mostert, P. & Lange, F. P. de. Prior expectations induce prestimulus sensory templates. *PNAS* **114**, 10473–10478 (2017).
37. Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P. & Gore, J. C. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci* **2**, 568–573 (1999).
38. Krusemark, E. A. & Li, W. From Early Sensory Specialization to Later Perceptual Generalization: Dynamic Temporal Progression in Perceiving Individual Threats. *Journal of Neuroscience* **33**, 587–594 (2013).
39. Moulson, M. C., Balas, B., Nelson, C. & Sinha, P. EEG Correlates of Categorical and Graded Face Perception. *Neuropsychologia* **49**, 3847–3853 (2011).
40. Rossion, B., Kung, C.-C. & Tarr, M. J. Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proc Natl Acad Sci U S A* **101**, 14521–14526 (2004).

41. Vlisides, P. E. *et al.* Neurophysiologic Correlates of Ketamine Sedation and AnesthesiaA High-density Electroencephalography Study in Healthy Volunteers. *Anesthes* **127**, 58–69 (2017).
42. Kaas, J. H., Hackett, T. A. & Tramo, M. J. Auditory processing in primate cerebral cortex. 7.
43. Romanski, L. M. *et al.* Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* **2**, 1131–1136 (1999).
44. Kaas, J. H. & Hackett, T. A. Subdivisions of auditory cortex and processing streams in primates. *PNAS* **97**, 11793–11799 (2000).
45. Diehl, M. M. & Romanski, L. M. Responses of prefrontal multisensory neurons to mismatching faces and vocalizations. *J Neurosci* **34**, 11233–43 (2014).
46. Hwang, J. & Romanski, L. M. Prefrontal neuronal responses during audiovisual mnemonic processing. *J Neurosci* **35**, 960–71 (2015).
47. Cao, Y., Summerfield, C., Park, H., Giordano, B. L. & Kayser, C. Causal Inference in the Multisensory Brain. *Neuron* **102**, 1076-1087.e8 (2019).
48. Kuperberg, G. R. & Jaeger, T. F. What do we mean by prediction in language comprehension? *Lang Cogn Neurosci* **31**, 32–59 (2016).
49. Tarr, M. J. & Gauthier, I. FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat Neurosci* **3**, 764–769 (2000).
50. Brants, M., Wagemans, J. & Op de Beeck, H. P. Activation of fusiform face area by Greebles is related to face similarity but not expertise. *J Cogn Neurosci* **23**, 3949–3958 (2011).

51. Júnior, R. D. M., Sousa, B. M. & Fukusima, S. Hemispheric specialization in face recognition: from spatial frequencies to holistic/analytic cognitive processing. (2014) doi:10.3922/J.PSNS.2014.4.09.

52. Tsao, D. Y., Schweers, N., Moeller, S. & Freiwald, W. A. Patches of face-selective cortex in the macaque frontal lobe. *Nature Neuroscience* **11**, 877–879 (2008).

53. Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J. & Desimone, R. Laminar differences in gamma and alpha coherence in the ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11262–11267 (2011).

54. van Kerkoerle, T. *et al.* Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14332–14341 (2014).

55. Bastos, A. M. *et al.* Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* **85**, 390–401 (2015).

56. Arnal, L. H. & Giraud, A.-L. Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences* **16**, 390–398 (2012).

57. Bastos, A. M., Lundqvist, M., Waite, A. S., Kopell, N. & Miller, E. K. Layer and rhythm specificity for predictive routing. *PNAS* (2020) doi:10.1073/pnas.2014868117.

58. Han, B. & VanRullen, R. The rhythms of predictive coding? Pre-stimulus phase modulates the influence of shape perception on luminance judgments. *Scientific Reports* **7**, 43573 (2017).

59. Klimesch, W. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences* **16**, 606 (2012).

60. Bastos, A. M. *et al.* Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711 (2012).

61. Alamia, A. & VanRullen, R. Alpha oscillations and traveling waves: Signatures of predictive coding? *PLoS Biol* **17**, e3000487 (2019).

62. Aan Het Rot, M., Zarate, C. A., Jr., Charney, D. S. & Mathew, S. J. Ketamine for depression: where do we go from here? *Biol Psychiatry* **72**, 537–47 (2012).

63. Yang, Y. *et al.* Ketamine blocks bursting in the lateral habenula to rapidly relieve depression. *Nature* **554**, 317–322 (2018).

64. Li, N. *et al.* mTOR-dependent synapse formation underlies the rapid antidepressant effects of NMDA antagonists. *Science* **329**, 959–964 (2010).

65. Strunk, D. R., Lopez, H. & DeRubeis, R. J. Depressive symptoms are associated with unrealistic negative predictions of future life events. *Behav Res Ther* **44**, 861–882 (2006).

66. Alloy, L. B. & Ahrens, A. H. Depression and pessimism for the future: biased use of statistically relevant information in predictions for self versus others. *J Pers Soc Psychol* **52**, 366–378 (1987).

67. Bruder, G. E. *et al.* EEG alpha measures predict therapeutic response to an SSRI antidepressant: Pre and post treatment findings. *Biol Psychiatry* **63**, 1171–1177 (2008).

68. Muthukumaraswamy, S. D. *et al.* Evidence that Subanesthetic Doses of Ketamine Cause Sustained Disruptions of NMDA and AMPA-Mediated Frontoparietal Connectivity in Humans. *J Neurosci* **35**, 11694–706 (2015).

69. Lyons, T. & Carhart-Harris, R. L. More Realistic Forecasting of Future Life Events After Psilocybin for Treatment-Resistant Depression. *Front. Psychol.* **9**, (2018).

70. Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–8 (1996).

71. Hannivoort, L. N. *et al.* Development of an Optimized Pharmacokinetic Model of Dexmedetomidine Using Target-controlled Infusion in Healthy Volunteers. *Anesthesiology* **123**, 357–367 (2015).

72. Colin, P. J. *et al.* Dexmedetomidine pharmacokinetic-pharmacodynamic modelling in healthy volunteers: 1. Influence of arousal on bispectral index and sedation. *Br J Anaesth* **119**, 200–210 (2017).

73. Yang, Y., Meng, Q., Pan, X., Xia, Z. & Chen, X. Dexmedetomidine produced analgesic effect via inhibition of HCN currents. *Eur. J. Pharmacol.* **740**, 560–564 (2014).

74. Morris, P. J. *et al.* Synthesis and N-Methyl-D-aspartate (NMDA) Receptor Activity of Ketamine Metabolites. *Org Lett* **19**, 4572–4575 (2017).

75. Domino, E. F. *et al.* Ketamine kinetics in unmedicated and diazepam-premedicated subjects. *Clin. Pharmacol. Ther.* **36**, 645–653 (1984).

76. Domino, E. F. *et al.* Plasma levels of ketamine and two of its metabolites in surgical patients using a gas chromatographic mass fragmentographic assay. *Anesth. Analg.* **61**, 87–92 (1982).

77. Chernik, D. A. P. *et al.* Validity and Reliability of the Observer's: Assessment of Alertness/Sedation Scale: Study with: Intravenous Midazolam. *Journal of Clinical Psychopharmacology* **10**, 244–251 (1990).

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

79. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational*

Intelligence and Neuroscience vol. 2011 e156869 https://www.hindawi.com/journals/cin/2011/156869/ (2010).

80. Gross, J. *et al.* Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *PNAS* **98**, 694–699 (2001).
81. Tzourio-Mazoyer, N. *et al.* Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage* **15**, 273–289 (2002).
82. Van Veen, B. D., van Drongelen, W., Yuchtman, M. & Suzuki, A. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans Biomed Eng* **44**, 867–880 (1997).
83. Wang, X., Chen, Y. & Ding, M. Estimating Granger causality after stimulus onset: A cautionary note. *NeuroImage* **41**, 767 (2008).
84. Seth, A. K., Barrett, A. B. & Barnett, L. Granger Causality Analysis in Neuroscience and Neuroimaging. *J. Neurosci.* **35**, 3293–3297 (2015).
85. Ratcliff, R. & McKoon, G. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput* **20**, 873–922 (2008).
86. Voss, A. & Voss, J. A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology* **52**, 1–9 (2008).
87. Vandekerckhove, J. & Tuerlinckx, F. Diffusion model analysis with MATLAB: a DMAT primer. *Behavior research methods* **40**, 61–72 (2008).
88. Gelman, A. *et al. Bayesian Data Analysis*. (Chapman and Hall/CRC, 2013). doi:10.1201/b16018.

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

90. Lopez-Calderon, J. & Luck, S. J. ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front. Hum. Neurosci.* **8**, (2014).

91. Rossion, B. *et al.* The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport* **11**, 69–74 (2000).

92. Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J. & Crommelinck, M. Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychol Sci* **13**, 250–257 (2002).

93. Winkler, I., Denham, S. & Escera, C. Auditory Event-related Potentials. in *Encyclopedia of Computational Neuroscience* (eds. Jaeger, D. & Jung, R.) 1–29 (Springer, 2013). doi:10.1007/978-1-4614-7320-6_99-1.

94. Ceponiene, R. *et al.* Event-related potentials associated with sound discrimination versus novelty detection in children. *Psychophysiology* **41**, 130–141 (2004).

95. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

96. The Elements of Statistical Learning - Data Mining, Inference, and Prediction, Second Edition | Trevor Hastie | Springer. https://www.springer.com/gp/book/9780387848570.

97. Abelson, R. P. & Prentice, D. A. Contrast tests of interaction hypothesis. *Psychological Methods* **2**, 315–328 (1997).