Neuromodulatory systems partially account for the topography of cortical networks of learning under uncertainty

Alice Hodapp, Florent Meyniel
Cognitive Neuroimaging Unit, NeuroSpin (INSERM-CEA), University of Paris-Saclay, Gif-sur-Yvette, France
GHU Paris,Psychiatrie et Neurosciences, Hôpital Saint-Anne, Institute for Neuromodulation, Paris, France

Abstract

Human learning in a dynamic and stochastic environment relies on computational variables such as confidence and surprise. If the learning process is shaped by neuromodulation, then the spatial distribution of receptors and transporters across the brain could put constraints on the spatial distribution of learning-related neural activity. Here, using fMRI data from four probabilistic learning studies and a Bayesian ideal observer model, we reveal a strong spatial invariance across tasks for the functional correlates of confidence, and to a lesser extent, surprise. Using 20 PET receptor/transporter density maps, we then show that this invariance could be partly explained by the chemoarchitecture of the cortex. We identified multiple receptors and transporters whose distribution aligned with the spatial distribution of neural activity in the cortex. While many of these receptors/transporters are in line with previous proposals of neuromodulation of learning, the results also revealed novel associations that can be targeted in experimental studies.

Keywords: neuromodulation; learning; confidence; surprise; fMRI; receptors; Bayesian inference; PET; brain mapping

Introduction

In an unstable and stochastic environment, we constantly have to decide whether to change our probabilistic expectations ("How long will it approximately take me to get to work?") in the light of new observations. Deviations from our expectation (e.g., the train ride took longer than usual) could simply reflect random variations (e.g., another train blocking the tracks or temporary signal failure) or instead reflect an important change (e.g., a new train schedule) requiring an update of our estimates. Bayesian inference provides an ideal and general solution to this dilemma: observations that are more surprising (because they deviate more from learned predictions) should be used to update predictions more strongly, but this update should be modulated by the confidence about the prediction. Surprise and confidence therefore both regulate how much is learned from a single observation. Interestingly, both human behavior (Peterson & Beach, 1967; Gallistel et al., 2014) and neural data (e.g., McGuire et al., 2014; Meyniel, 2020; O'Reilly et al., 2013; Sedley et al., 2016) show signatures of this Bayesian learning.

On a neural level, how much is learned from an observation might depend on neuromodulation. Neuromodulators have been linked to different cognitive functions (Buhot, 1997; Puig et al., 2015; Robbins & Arnsten, 2009). Specifically, major neuromodulator systems such as dopamine, acetylcholine, norepinephrine and serotonin, have been proposed to implement the regulatory roles assigned to surprise and confidence in learning under uncertainty (Doya, 2002, 2008; Yu & Dayan, 2005). These neuromodulatory systems project from brain stem nuclei to the basal ganglia, the cerebellum, and the cortex. They influence neurons and circuits by changing e.g. synaptic strength, excitability of neurons, and networks (e.g., Celada et al., 2013; Puig & Gulledge, 2011; Salgado et al., 2016) depending notably on receptors (Goulas et al., 2021; Hansen et al., 2022; Shine, 2019). While many influential theories regarding the neuromodulation of learning exist, the role of receptors and transporters has received limited attention so far.

Neurotransmitter receptors and transporters vary widely not only in terms of their affinity, timescales, and downstream effects on neuronal excitability, but also in terms of their distributions across the brain (Goulas et al., 2021; Shine, 2019; Zilles & Palomero-Gallagher, 2017) which shapes structural and functional organization (Bo et al., 2024; Hansen et al., 2022; Komorowski et al., 2022; Pedersen et al., 2024). The hypothesis we test in this study is that, if neuromodulation is a domain general mechanism of learning, then the heterogeneous distribution of receptors and transporters across the brain should put constraints on the spatial distribution of neural activity corresponding to the computational variables of learning. While the spatial organization of neural activity is constrained by multiple neurobiological factors that have been studied previously (e.g., connectivity, architecture, function; Glasser et al., 2016), the focus of this study is on the role of neurotransmitter receptors and transporters.

Here, to test our hypothesis, we use fMRI data from different tasks of learning under uncertainty and PET receptor and transporter density maps to ask:

(1) Is there invariance in the topographical organization of learning-related effects?

(2) Can this invariance be explained by patterns in receptor density and therefore neuromodulatory activity?

(3) Can we identify receptors of interest specifically for confidence and surprise?

Methods

We first quantify the extent of invariance of the topographical organization of learning-related effects. To the best of our knowledge, no meta-analysis maps of these computational variables exist, we therefore first computed effect maps associated with confidence and surprise from a few datasets from

which we could obtain the raw data in order to ensure consistent analysis methods across studies, and we quantified the overlap of these maps across tasks. We then related learning-related subject data to receptor and transporter distributions in the brain. Similar approaches have been employed before to examine the relationship between receptors and brain function more globally (Bo et al., 2024; Hansen et al., 2022; Komorowski et al., 2022; Luppi et al., 2023), typically using large scale group-level maps of different cognitive functions.

fMRI datasets

To quantify the invariance of effect maps, we used 4 fMRI datasets of learning under uncertainty that differed along the axes of sensory modality (visual vs auditory), sequence structure (Bernoulli vs transition probabilities), and estimate to learn (probability vs magnitude). Studies 1-3 were probability learning tasks (Figure 1A), and Study 4 was a learning and decision-making task (Figure 1B). In all tasks, the latent variable to learn could change suddenly and unpredictably.

Study 1 comprises 7T fMRI data from Bounmy et al. (2023). 26 participants were presented with a visual sequence composed of two Gabor patches drawn from a Bernoulli process. Occasionally, participants reported their estimate for one of the stimuli as well as their confidence about this estimate.

Study 2 consists of yet unpublished 3T data in which 56 participants were presented with an auditory sequence of two tones. The tones were determined by a Bernoulli process and participants were occasionally asked to indicate their estimate for the probability of a sound and their confidence in this estimate. This study did not provide full brain coverage, as it focused on subcortical structures. Yet, we study the cortex here.

Study 3 was previously published in Meyniel & Dehaene (2017) and consists of 3T data from 21 participants learning the transition probabilities underlying items in a sequence. Both auditory and visual sequences were used. Participants occasionally reported their confidence in their current estimates.

Study 4 consists of yet unpublished 3T data from a magnitude (reward) learning task. 56 participants performed a sequence of choices in a two-armed bandit task trying to maximize cumulative reward, which required learning the underlying reward level for two choices with independently evolving reward distributions. They were presented with a reward between 1 and 100 points for the chosen option. Subjects were asked occasionally to make explicit reports of their reward level estimate as well as confidence about this estimate.

Processing and analysis

For every dataset we normalized the data to the MNI space. Time series in each voxel were detrended, high-pass filtered (at 1/128 Hz), z-scored per experimental run and smoothed with a 5mm FWHM Gaussian kernel. Working with the raw data rather than previously published group results allowed us to apply identical regression models in each study, with the aim to compute effects maps corresponding to confidence and surprise derived from an ideal observer model. Effect maps correspond to the regression coefficients from a general linear model (GLM) obtained with Nilearn v0.11.1 (Abraham et al., 2014).

For all studies, the GLM modeled stimulus onset with surprise and confidence as parametric modulations. It also included the questions asked in the task as well as several corresponding parametric modulations (reaction times, the ideal observer's prediction and confidence as well as the subject's prediction and confidence in response to the question, if applicable) and motion parameters. Additionally, the probability learning tasks included predictability (defined as entropy) and the current prediction about what the stimulus would be, and the reward learning task included expected reward difference between chosen and unchosen option as stimulus onset modulations derived from the ideal observer.

Bayesian model of learning

The learning-related measures were derived from an ideal observer model, which is principled and served as a reference across subjects and studies. The ideal observer model was implemented with a well-known solution based on a Hidden Markov Model (Behrens et al., 2007; Gallistel et al., 2014). In short, the model provides optimal estimates of the changing quantity of interest (item probability or reward level) θ_t given the observations so far, $y_{1:t}$. This is achieved by iteratively combining the likelihood of the current observation and a prior:

$$p(\theta_{t+1} \mid y_{1:t+1}, \upsilon) \propto p(y_{t+1} \mid \theta_{t+1}) \int p(\theta_t \mid y_{1:t}, \upsilon) \times p(\theta_{t+1} \mid \theta_t, \upsilon)\, d\theta_t$$

The learner assumes that θ can change unexpectedly with some prior probability υ, as captured by $p(\theta_{t+1} \mid \theta_t, \upsilon)$, in which case it is resampled uniformly. The volatility υ, the transition matrix and prior probability or reward levels (when applicable) correspond to the generative structure of the task. The prediction of the next stimulus is the average of the posterior distribution:

$$p(y_{t+1} \mid y_{1:t}, \upsilon) = \iint p(y_{t+1} \mid \theta_{t+1})\, p(\theta_{t+1} \mid \theta_t) \times p(\theta_t \mid y_{1:t}, \upsilon)\, d\theta_{t+1}\, d\theta_t$$

We formalized the confidence about the prediction for the next stimulus as log precision of the distribution in the probability learning task

$$-\log(\mathrm{var}(\theta_{t+1} \mid y_{1:t}, \upsilon))$$

and as one minus the maximum a posteriori probability in the decision task (this difference does not matter in practice). We quantified surprise as log improbability of the observation (Shannon, 1948):

$$-\log_2(p(y_{t+1} \mid y_{1:t}, \upsilon))$$

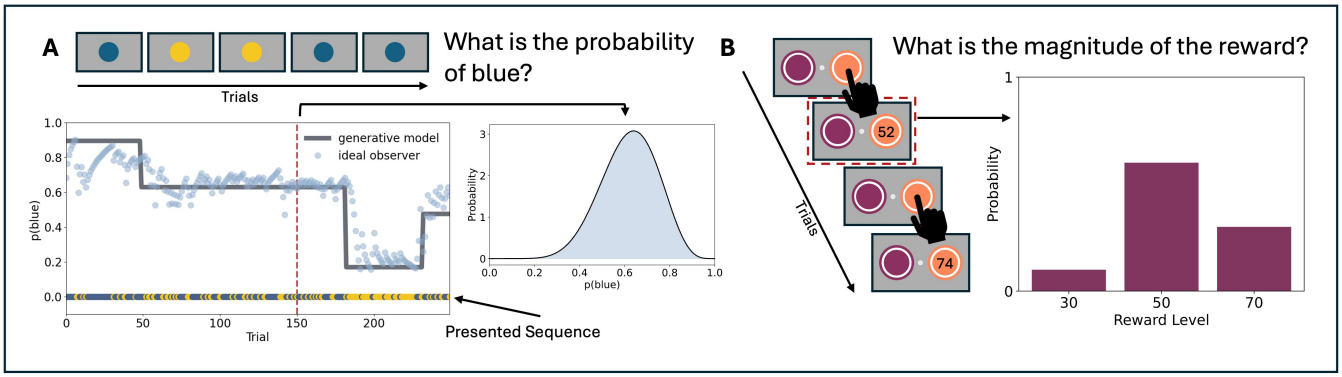

Figure 1: **Experiments**. **A**: Example probability learning task (similar to Study 1). Participants are presented with a sequence of two stimuli (blue and yellow). The generative probability p(blue) changed unpredictably (see gray line), the blue dots represent the estimates of an ideal observer. The right plot shows the model posterior over p(blue) at the trial indicated by the red line. **B**: Reward learning task. At each trial subjects choose between two options and observe the outcome (reward) of the selected option, drawn from a Gaussian distribution. The latent mean reward levels underlying the two options could change unexpectedly between 3 reward levels (30, 50, 70). The right plot shows an example posterior probability following an observation.

Overlap of effect maps

To evaluate the invariance of effect maps of confidence and surprise, we combined the individual regression coefficients in a group level t-test against 0 for each study separately. The resulting statistical maps were thresholded at $p<0.001$ and corrected for multiple comparisons across voxels at the cluster level with $p_{FWE}<0.05$ (Monte Carlo based cluster correction in Nilearn). We then plotted the spatial overlap across tasks for the negative and positive clusters separately (Figure 3B/E) and correlated group level maps (Figure 3C/F) for confidence (or surprise) between all studies for an illustration of spatial invariance.

We also quantified the overlap with a threshold-free, quantitative method. We correlated each of the unthresholded group level confidence effect maps for both confidence and surprise. To test whether confidence (or surprise) effect maps of different studies had a higher correlation than what would be expected by chance, we used the "spin test". We projected the maps to the spherical surface of an inflated brain, and randomly rotated them 1000 times to create null maps (for details see: Alexander-Bloch et al., 2018; Blaser & Fryzlewicz, 2016; Markello & Misic, 2021). This test is stringent because the permuted null maps retain the histogram and autocorrelation of the empirical maps.

PET receptor density maps

As a measure of receptor distribution across the cortex, we combined the PET-derived whole brain density maps from 19 unique neurotransmitter receptors and transports published in Hansen et al. (2022) with an α_2 receptor density map provided by Laurencin et al. (2023). The whole dataset spanned nine different neurotransmitter systems. All PET images were available in voxel MNI space.

Relating brain activity and receptor density maps

The focus of this study was the investigation of a possible relationship between cortical effects in fMRI and receptor density distribution (see Figure 2 for a visual illustration). For this we first fit a multiple linear regression model for each subject in each study that predicts the effect map for either surprise or confidence from all available neurotransmitter and transporter densities to evaluate full model fit. To get a measure of the average explained variance within each study, we predicted a left out subject's effect map from the model fit across all other subjects for each study/variable combination. The significance of the explained variance was assessed using the spin test. A null distribution of explained variance was constructed by repeating the same cross-validation on random rotations of the receptor/transporter maps (1000 repetitions; see section Overlap of effect maps for details). Additionally, we derived an estimate of the average explainable variance in each study/variable combination by calculating the mean R^2 across models that predicted the effect map of a left out subject from the mean map of the other subjects. Note that since the spin test used for estimating a null distribution requires surface projection, all of the measures of explained variance were calculated based on surface maps for comparability.

We first investigated the full model. In the regression model that included all 20 receptors and transporters, the significance of each parameter estimate was established in a t-test against 0 within each study. We highlighted parameter estimates that have a consistent sign across studies, and with $p_{FDR}<0.05$ in each of the studies, in figure 4AB (with plus/minus).

To identify receptors or transporters that contribute most to the model fit between receptors/transporters and the cortical effects in fMRI, we applied a dominance analysis (Azen & Budescu, 2003) on the subject level (in voxel MNI space). Dominance provides an intuitive notion of regressor impor-

tance by assigning a proportion of the model fit to each independent variable and accounts for interactions between the predictors. Specifically, it fits the same regression model on every combination and subset of independent variables (2^{p-1} regression models for p independent variables). Total dominance for a given receptor/transporter map is then calculated as the average increase in model fit across all regression models that include this receptor/transport map, compared to all other regression models. We then normalized the dominance values by the full model fit (R^2 of the model with all transporter/receptors as predictors) to make the values comparable across latent variables (confidence and surprise) and studies (see Hansen et al., 2022, for a similar approach) and report them as receptor/transporter dominance in the remainder of this paper.

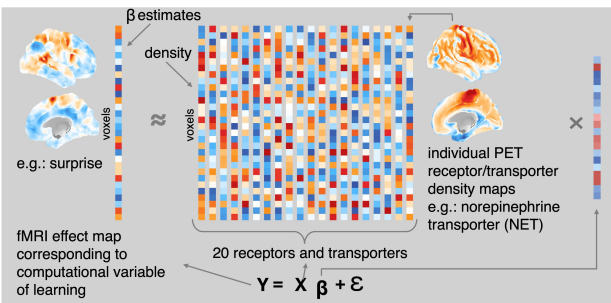

Figure 2: Analysis framework. Is there a relationship between effects in fMRI corresponding to computational variables for learning under uncertainty and receptor/transporter density distribution in the cortex?

Results

Bayesian inference accounts for subject behavior, as shown in previous studies

Previous studies and the publications underlying some of the data used here have reported that subjects are able to introspect the normative values corresponding to Bayesian learning (formalized with the ideal observer model). We first replicated this result across the studies used here by correlating subjects' reports and model estimates. Subjective confidence followed the ideal observer used in each study (Study 1: $r = .16$, $t(25) = 5.9$, $p = 3.6 \times 10^{-6}$; Study 2: $r = .18$, $t(55) = 8.8$, $p = 8.0 \times 10^{-12}$; Study 3: $r = .20$, $t(20) = 6.90$, $p = 1.1 \times 10^{-6}$; Study 4: $r = .29$, $t(55) = 14.60$, $p = 10^{-19}$) We also find a tight fit for probability estimates (when available) (Study 1: $r = .79$, $t(25) = 31.8$, $p = 9.99 \times 10^{-22}$; Study 2: $r = .69$, $t(55) = 16.8$, $p = 1.0 \times 10^{-22}$; Study 3: not available) and reward estimates (Study 4: fraction matching: $r = .61$; comparison with chance level (0.33): $t(55) = 19.20$, $p = 10^{-27}$.). For an overview of linear regression results (Table 2) and plots of the behavioral results (Figures 5 and 6), see appendix.

Spatial invariance of the effect maps of latent variables of learning across tasks

Turning to the specific goals of this study, we then quantified the extent to which effects maps are invariant across studies. Figure 3 depicts group-level z statistic maps by study for both confidence and surprise. The statistical maps for confidence (Figure 3A) are visually remarkably consistent across studies, which is supported by an overlap analysis (Figure 3B), with negative effects across all four studies in the posterior intraparietal sulcus, precuneus and anterior insula. There was also a strong overlap (up to three studies) in the middle frontal gyrus. For surprise, the cluster overlap (Figure 3E) reveals a core region with positive effects in right precentral sulcus/inferior frontal gyrus pars opercularis across all tasks and multiple other regions across probability learning tasks (see Figure 3). The effect for surprise also reveals effects in sensory cortices corresponding to the stimulus modality in the respective study, as well as bilateral effects in the anterior insula in the reward learning Study 4 (Figure 3D). Together, this suggests spatial invariance for effect maps of confidence and to some extent for surprise.

To summarize invariance quantitatively, we computed group level correlations (Figure 3C/F) between the effect maps (for either confidence or surprise) of different studies and tested their significance against a null model. The maps for confidence show high correlations across all studies, significantly higher than what would be expected from a correlation with a random map that has the same distribution and spatial autocorrelation. The effect maps of surprise from the probability learning studies (Studies 1-3) are also characterized by significant positive correlations. Meanwhile, the effect map of surprise in Study 4 (which combined reward learning and decision-making) does not reveal clear correlations with any of the other surprise maps, which is expected given differences between studies (see Discussion). For a visualization of the correlations between confidence maps and surprise maps, see supplement (Figure 7 in appendix.

In sum, the data provides evidence for a strong spatial invariance in the effects for confidence across all tasks and across surprise within the probability learning tasks.

Neuromodulation partly accounts for the spatial invariance of effect maps

We quantified the extent to which the invariance of effect maps could be accounted for by neuromodulation. Table 1 shows the average cross-validated fit of the full model (all 20 available receptors and transporter density distributions) across subjects. revThe variance accounted for by neuromodulation was sizable compared to an estimate of the explainable variance per study/variable combination. The results show that receptors and transporter density distributions could explain some of the variance in the effect maps of both confidence and surprise. For each of the study/variable combinations (except surprise effects in Study 2) the average amount of explained variance is significantly higher than what would be expected

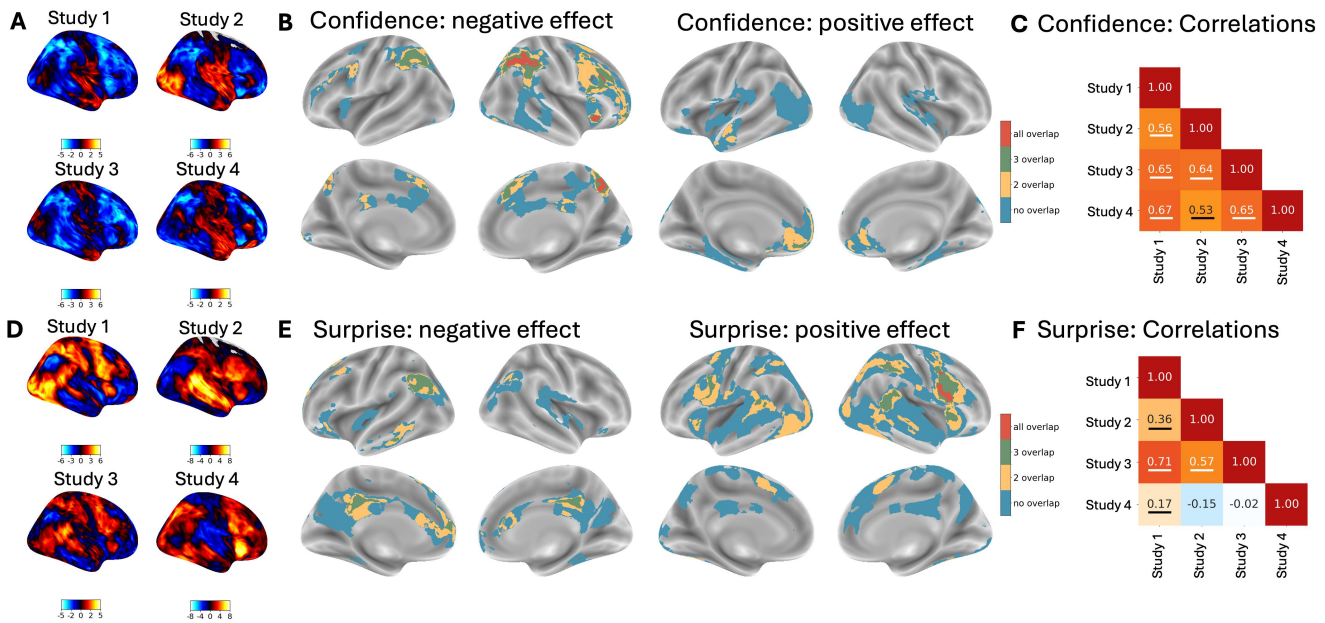

Figure 3: **Invariance of effect maps across studies.**. **A**/**D**: z statistic maps for confidence/surprise in the four studies. **B**/**E**: Brain regions where effects significantly correlated with confidence/surprise across studies. The colors correspond to the number of studies in which this region showed significant positive/negative clusters thresholded at $p<0.001$ and corrected for multiple comparisons across voxels at the cluster level with $p_{FWE}<0.05$. **C**/**F** Correlations between effect maps across studies. Underlined correlation coefficients were significant in a comparison to a null model (all $p<0.001$ at 1000 rotations)

with random receptor/transporter density maps with the same histogram and spatial autocorrelation.

Table 1: **Average model fit and explainable variance.** A significantly higher model R2 than null model is indicated in null model column: ***: p <0.001; **: p <0.01; *: p <0.05.

		Average R2 (CV)	Null model R2 (CV)	Explainable variance (max R2)	Explainable variance explained by model
Confidence	Study 1	0.036	0.022**	0.12	30%
	Study 2	0.023	0.012***	0.092	25%
	Study 3	0.036	0.016**	0.12	30%
	Study 4	0.009	0.0068*	0.067	13.4%
Surprise	Study 1	0.047	0.037**	0.25	18.8%
	Study 2	0.036	0.032	0.19	18.9%
	Study 3	0.022	0.012*	0.087	25.3%
	Study 4	0.018	0.0046**	0.094	19.1%

Identifying specific receptors/transporters contributing to the invariance of effect maps

We then identified individual receptors of interest that contributed most to the explained variance. For this we first focus on the results for confidence. Across the three probability learning studies, dominance analysis revealed a consistent pattern of receptor importance for confidence (Figure 4A). Compared to other receptors, the spatial distribution of the μ-opioid receptor (MOR) made the largest contribution to the model fit between receptors and cortical effects corresponding to confidence. In the full model, the MOR was significant across subjects in each study with a consistent negative sign. This indicates that areas with high density of MOR (see Figure 4C for its distribution) had an increased activation with higher uncertainty (i.e. negative correlation with confidence). A range of other receptors revealed consistent and significant regressors across studies (5-HT_{1B}, A_4B_2, CB_1) but contributed much less to the model fit. Contributions of receptors/transporters were fairly consistent across probability learning studies (see heatmaps in Figure 4A). For the effect maps of confidence in the reward learning study, the dominance analysis revealed a weaker pattern of receptor importance than in the probability learning studies (Figure 4D). However, MOR still arises as a receptor of interest together with other receptors.

For surprise effect maps in probability learning studies, the density of the norepinephrine transporter (NET; see Figure 4C for its distribution) made the largest contribution to the fit between receptors and cortical effects corresponding to surprise (Figure 4B). However, a strong contribution was mostly visible in studies 1 and 3. For Study 2 there was less clear dominance in receptor contribution. The NET coefficient in the full model had a consistent positive sign and was significant across participants within all three studies. This implies that areas with high density of NET had an increased activation

with higher surprise. Again, a range of other receptors revealed consistent and significant regressors across studies (5-HT_6, D_1, D_2, CB_1) but contributed much less to the model fit. The results differ for the reward learning task. For surprise effect maps in this task, MOR and NET density made the largest contributions to a model of effects in fMRI (Figure 4D). Specifically, regions with strong effects for increased surprise seem characterized by a lower density of NET and higher density of MOR.

Discussion

Learning under uncertainty is regulated by computational variables such as confidence and surprise. Here we examined (1) if there is invariance of the fMRI effect maps related to these variables across probabilistic learning tasks and (2) if this spatial pattern corresponds to the distribution of specific receptors across the brain.

We find that the cortical effects corresponding to confidence are strongly invariant across the different tasks. The confidence effect map shows substantial overlap and strong correlations not only across modalities (visual vs auditory) and complexity of the to be learned statistic (Bernoulli vs transition probabilities) within probability learning, but also with the magnitude (reward) learning task. While confidence (or inversely, uncertainty) drives learning rate in probability learning, it has a reduced role in magnitude learning, where learning rate is increased mostly by change point probability (Foucault & Meyniel, 2024). The effect maps of confidence were nevertheless similar between probability and magnitude learning, while this finding needs to be replicated using different magnitude learning tasks, it suggests that the neural activity corresponding to confidence is to some extent domain general (for the same question in metacognition, see: Fleming, 2024; Mazancieux et al., 2020).

This invariance of the confidence effect maps can be partly explained by the spatial distribution of specific receptors and transporters across the cortex. The relationship between neurotransmitter receptors/transporters and confidence effect maps is consistent across probability learning tasks and subjects (see Figure 4). Receptors related to neuromodulators previously discussed in the context of learning under uncertainty (Doya, 2002, 2008; Yu & Dayan, 2005) were significant and consistent in their effect across studies, notably 5-HT_{1b} (serotonin) and A_4B_2 (acetylcholine). Unexpectedly however, across studies, the μ-opioid receptor (MOR) density revealed the strongest contributions to the mapping of cortical effects to receptor distribution. MOR receptor activation has complex downstream effects; in the context of learning, the relationship with the inhibition of the locus coeruleus through endogenous opioids (Curtis et al., 2001, 2012) and indirect disinhibition of dopamine neurons (Gysling & Wang, 1983; Johnson & North, 1992) may be most relevant. The relationship between confidence representations and the MOR receptor is not established in the literature and provides an avenue for further research. While the contribution of individual receptors is less clear in the reward learning study, it demonstrated a similar trend. The consistent alignment of the confidence effect maps to receptor density across different tasks suggests that the invariant representation of confidence might to some extent be a consequence of a domain general neuromodulatory process.

Cortical effects corresponding to surprise also revealed some invariance across probability learning tasks. The effects overlapped in core regions and was additionally influenced by the sensory modality. The spatial distribution of surprise can to some extent be explained by the receptor and transporter distribution in the brain. Of special interest is the norepinephrine transporter, since it contributed most to the mapping between effects for surprise and receptor/transporter densities. The effect of NET density is significant and consistent in direction across studies. Its contribution is less strong in Study 2, potentially because of the partial brain coverage and the strong activation of modality specific (temporal) anatomical regions. The NE transporter is responsible for the reuptake of extracellular norepinephrine and its importance in probability learning suggested by our results is much aligned with influential theories on probabilistic learning (Yu & Dayan, 2005). Other receptors that were significant across studies and with a consistent direction of the effect, were also in line with neuromodulatory systems suggested to play a role in the mediation of learning under uncertainty, 5-HT_6 (serotonin), and D_1 and D_2 (dopamine) in particular (Doya, 2002, 2008).

The surprise effect maps in the reward learning study reveal a different pattern for both the spatial pattern and the alignment with receptors and transporters. The effect maps of probability learning and reward learning studies overlapped only in a core region in right precentral sulcus/inferior frontal gyrus pars opercularis, and otherwise differed in their topography corresponding to surprise. This difference remains to be replicated in other reward learning studies, but some key differences between the probability and reward learning studies analyzed here might explain this difference in effects. In reward learning tasks, the error has a meaningful sign (i.e., higher or lower than expected), making (Shannon) surprise less relevant for the task. Additionally, the task combined learning and decision-making. Last, the two computational variables of interest, confidence and surprise, were also more strongly correlated in the reward learning task than in the probability learning tasks. This could have contributed to the strong similarity of the surprise map to the confidence map that is observed for the reward learning study. Some of the spatial distribution of the effect map could be explained by the spatial distribution of receptor and transporter density. While the pattern of receptor importance is much less clear compared to the probability learning studies, NET again arises as a receptor of interest. However, in line with the difference in spatial effects compared to probability learning, the NE transporter density effect seems to be inverse (i.e. less transporter in areas with large surprise effect).

Note that overall, the average R^2 for the full model and

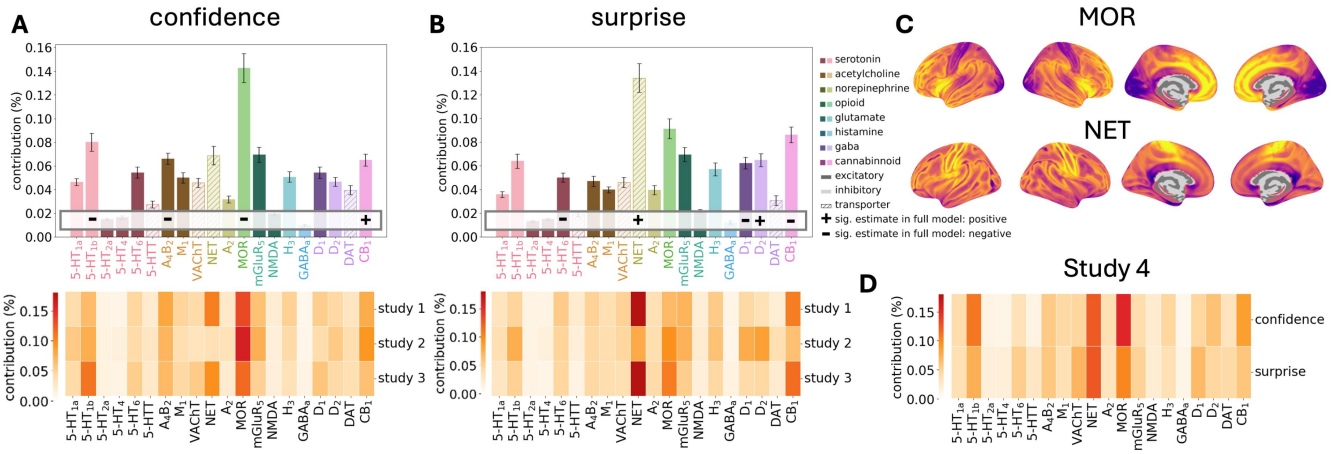

Figure 4: **Predicting fMRI effects from receptor distributions**. Dominance analysis revealed the contribution of each receptor and transporter to the model fit. The percent contribution of each receptor reflects the variables dominance normalized by the model fit (R2). The top row in **A** (confidence) and **B** (surprise) depicts the mean percent contribution across all subjects from all probability studies (studies 1-3). Error bars reflect the SEM. The signs across the bottom of the bar plot indicate consistent signs and significance across subjects in each study in the full regression model. The heat maps illustrate the dominance results by study. **C**: Example receptor density distributions across the cortex for MOR and NET. **D**: Dominance results for surprise and confidence for the reward learning study (Study 4).

the contribution (dominance) of receptors of interest are lower here than in comparable studies (Bo et al., 2024; Hansen et al., 2022; Komorowski et al., 2022). This is because the results reported here use single subject (voxel-wise) effect maps, which allows us to get a measure of the variability across subjects (see error bars in Figure 4), as opposed to the use group-level (often by region) effect maps in other studies, which increases the smoothness of maps, and thus, the R^2.

We now discuss some limitations of our work. We restricted the results here to the cortex, due to differences in PET radioligand sensitivity between cortical and subcortical structures (Elsinga et al., 2006; Hansen et al., 2022; Van De Bittner et al., 2014). Along the same lines, while the PET dataset provided by Hansen et al. (2022) (with the addition of the α_2 receptor by Laurencin et al., 2023) is the largest dataset of receptor densities to date, it only presents a selection of all receptors. The use of PET images as our source of receptor densities has the disadvantage of a low spatial resolution, resulting in a relatively smooth image which could inflate the relationships found, and making it difficult to perform an analysis at a finer spatial resolution than the whole brain. Another downside of the used PET images is that they consist of average maps from a different group of participants. Our focus on invariant effect maps make our between-subjects approach (across fMRI and PET) valid; yet, using subject-specific PET and fMRI data in this study could strengthen the results. Our approach assumes a linear effect of receptors and transporters, which is a conservative assumption that might not hold in the brain. It is therefore important to clarify that we do not provide evidence for the absence of a relationship between certain receptors/transports and effects in fMRI (this relationship could be nonlinear, and not detected here), only that we find a strong linear relationship for some receptors and transporters. Looking into nonlinear effects is a promising future perspective. Lastly, a key limitation of the current study is its correlational nature. However, this allowed us to evaluate a large range of receptors and transporters corresponding to multiple neuromodulatory systems and generate hypotheses which can be tested in future experimental work.

Conclusion

We demonstrated that cortical activity related to latent variables of learning under uncertainty shows similar effects across different tasks, specifically for representations of confidence and to a lesser extent for surprise. This invariance across tasks might be partly explained by the distribution of receptors across the cortex. While cortical activity is likely a result of overlapping patterns of receptor densities, we identified clear receptors of interest for future studies. Some of the receptors suggest neuromodulatory projections that have previously been outlined in the learning literature, but crucially some results suggest novel relationships and highlight the potential of generating hypotheses by investigating the role of receptors.

Acknowledgements

This work was supported by funding from the European Research Council (ERC grant 947105) to F.M. α_2 receptor density map © Copyright CRNL, CERMEP – Imagerie du vivant, www.cermep.fr and Hospices Civils de Lyon. All rights reserved

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., … Varoquaux, G. (2014, February). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. doi: 10.3389/fninf.2014.00014

Alexander-Bloch, A. F., Shou, H., Liu, S., Satterthwaite, T. D., Glahn, D. C., Shinohara, R. T., … Raznahan, A. (2018, September). On testing for spatial correspondence between maps of human brain structure and function. *NeuroImage*, *178*, 540–551. doi: 10.1016/j.neuroimage.2018.05.070

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, *8*(2), 129–148. doi: 10.1037/1082-989X.8.2.129

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007, September). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. doi: 10.1038/nn1954

Blaser, R., & Fryzlewicz, P. (2016, January). Random rotation ensembles. *J. Mach. Learn. Res.*, *17*(1), 126–151.

Bo, K., Kraynak, T. E., Kwon, M., Sun, M., Gianaros, P. J., & Wager, T. D. (2024, March). A systems identification approach using Bayes factors to deconstruct the brain bases of emotion regulation. *Nature Neuroscience*, 1–13. doi: 10.1038/s41593-024-01605-7

Bounmy, T., Eger, E., & Meyniel, F. (2023, March). A characterization of the neural representation of confidence during probabilistic learning. *NeuroImage*, *268*, 119849. doi: 10.1016/j.neuroimage.2022.119849

Buhot, M.-C. (1997, April). Serotonin receptors in cognitive behaviors. *Current Opinion in Neurobiology*, *7*(2), 243–254. doi: 10.1016/S0959-4388(97)80013-X

Celada, P., Puig, M. V., & Artigas, F. (2013, April). Serotonin modulation of cortical neurons and networks. *Frontiers in Integrative Neuroscience*, *7*. doi: 10.3389/fnint.2013.00025

Curtis, A. L., Bello, N. T., & Valentino, R. J. (2001, July). Evidence for Functional Release of Endogenous Opioids in the Locus Ceruleus during Stress Termination. *The Journal of Neuroscience*, *21*(13), RC152-RC152. doi: 10.1523/JNEUROSCI.21-13-j0001.2001

Curtis, A. L., Leiser, S. C., Snyder, K., & Valentino, R. J. (2012, March). Predator stress engages corticotropin-releasing factor and opioid systems to alter the operating mode of locus coeruleus norepinephrine neurons. *Neuropharmacology*, *62*(4), 1737–1745. doi: 10.1016/j.neuropharm.2011.11.020

Doya, K. (2002, June). Metalearning and neuromodulation. *Neural Networks*, *15*(4), 495–506. doi: 10.1016/S0893-6080(02)00044-8

Doya, K. (2008, April). Modulators of decision making. *Nature Neuroscience*, *11*(4), 410–416. doi: 10.1038/nn2077

Elsinga, P. H., Hatano, K., & Ishiwata, K. (2006, August). PET Tracers for Imaging of the Dopaminergic System. *Current Medicinal Chemistry*, *13*(18), 2139–2153. doi: 10.2174/092986706777935258

Fleming, S. M. (2024, January). Metacognition and Confidence: A Review and Synthesis. *Annual Review of Psychology*, *75*(Volume 75, 2024), 241–268. doi: 10.1146/annurev-psych-022423-032425

Foucault, C., & Meyniel, F. (2024, May). Two Determinants of Dynamic Adaptive Learning for Magnitudes and Probabilities. *Open Mind*, *8*, 615–638. doi: 10.1162/opmi_a_00139

Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The perception of probability. *Psychological Review*, *121*(1), 96–123. doi: 10.1037/a0035232

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., … Van Essen, D. C. (2016, August). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. doi: 10.1038/nature18933

Goulas, A., Changeux, J.-P., Wagstyl, K., Amunts, K., Palomero-Gallagher, N., & Hilgetag, C. C. (2021, January). The natural axis of transmitter receptor distribution in the human cerebral cortex. *Proceedings of the National Academy of Sciences*, *118*(3), e2020574118. doi: 10.1073/pnas.2020574118

Gysling, K., & Wang, R. Y. (1983, October). Morphine-induced activation of A10 dopamine neurons in the rat. *Brain Research*, *277*(1), 119–127. doi: 10.1016/0006-8993(83)90913-7

Hansen, J. Y., Shafiei, G., Markello, R. D., Smart, K., Cox, S. M. L., Nørgaard, M., … Misic, B. (2022, November). Mapping neurotransmitter systems to the structural and functional organization of the human neocortex. *Nature Neuroscience*, *25*(11), 1569–1581. doi: 10.1038/s41593-022-01186-3

Johnson, S., & North, R. (1992, February). Opioids excite dopamine neurons by hyperpolarization of local interneurons. *The Journal of Neuroscience*, *12*(2), 483–488. doi: 10.1523/JNEUROSCI.12-02-00483.1992

Komorowski, A., Murgaš, M., Vidal, R., Singh, A., Gryglewski, G., Kasper, S., … Goya-Maldonado, R. (2022). Regional gene expression patterns are associated with task-specific brain activation during reward and emotion processing measured with functional MRI. *Human Brain Mapping*, *43*(17), 5266–5280. doi: 10.1002/hbm.26001

Laurencin, C., Lancelot, S., Merida, I., Costes, N., Redouté, J., Le Bars, D., … Ballanger, B. (2023, May). Distribution of $A2$-Adrenergic Receptors in the Living Human Brain Using [11C]yohimbine PET. *Biomolecules*, *13*(5), 843. doi: 10.3390/biom13050843

Luppi, A. I., Hansen, J. Y., Adapa, R., Carhart-Harris, R. L., Roseman, L., Timmermann, C., ... Stamatakis, E. A. (2023, June). In vivo mapping of pharmacologically induced functional reorganization onto the human brain's neurotransmitter landscape. *Science Advances*, *9*(24), eadf8332. doi: 10.1126/sciadv.adf8332

Markello, R. D., & Misic, B. (2021, August). Comparing spatial null models for brain maps. *NeuroImage*, *236*, 118052. doi: 10.1016/j.neuroimage.2021.118052

Mazancieux, A., Fleming, S. M., Souchay, C., & Moulin, C. J. A. (2020, September). Is there a G factor for metacognition? Correlations in retrospective metacognitive sensitivity across tasks. *Journal of Experimental Psychology: General*, *149*(9), 1788–1799. doi: 10.1037/xge0000746

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014, November). Functionally Dissociable Influences on Learning Rate in a Dynamic Environment. *Neuron*, *84*(4), 870–881. doi: 10.1016/j.neuron.2014.10.013

Meyniel, F. (2020, June). Brain dynamics for confidence-weighted learning. *PLOS Computational Biology*, *16*(6), e1007935. doi: 10.1371/journal.pcbi.1007935

Meyniel, F., & Dehaene, S. (2017, May). Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proceedings of the National Academy of Sciences*, *114*(19), E3859-E3868. doi: 10.1073/pnas.1615773114

O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013, September). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, *110*(38), E3660-E3669. doi: 10.1073/pnas.1305373110

Pedersen, R., Johansson, J., Nordin, K., Rieckmann, A., Wåhlin, A., Nyberg, L., ... Salami, A. (2024, March). Dopamine D1-Receptor Organization Contributes to Functional Brain Architecture. *The Journal of Neuroscience*, *44*(11), e0621232024. doi: 10.1523/JNEUROSCI.0621-23.2024

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, *68*(1), 29–46. doi: 10.1037/h0024722

Puig, M. V., & Gulledge, A. T. (2011, December). Serotonin and Prefrontal Cortex Function: Neurons, Networks, and Circuits. *Molecular Neurobiology*, *44*(3), 449–464. doi: 10.1007/s12035-011-8214-0

Puig, M. V., Gulledge, A. T., Lambe, E. K., & Gonzalez-Burgos, G. (2015, October). Editorial: Neuromodulation of executive circuits. *Frontiers in Neural Circuits*, *9*. doi: 10.3389/fncir.2015.00058

Robbins, T. W., & Arnsten, A. F. T. (2009, July). The Neuropsychopharmacology of Fronto-Executive Function: Monoaminergic Modulation. *Annual Review of Neuroscience*, *32*(Volume 32, 2009), 267–287. doi: 10.1146/annurev.neuro.051508.135535

Salgado, H., Treviño, M., & Atzori, M. (2016, June). Layer- and area-specific actions of norepinephrine on cortical synaptic transmission. *Brain Research*, *1641*, 163–176. doi: 10.1016/j.brainres.2016.01.033

Sedley, W., Gander, P. E., Kumar, S., Kovach, C. K., Oya, H., Kawasaki, H., ... Griffiths, T. D. (2016, March). Neural signatures of perceptual inference. *eLife*, *5*, e11476. doi: 10.7554/eLife.11476

Shine, J. M. (2019, July). Neuromodulatory Influences on Integration and Segregation in the Brain. *Trends in Cognitive Sciences*, *23*(7), 572–583. doi: 10.1016/j.tics.2019.04.002

Van De Bittner, G. C., Ricq, E. L., & Hooker, J. M. (2014, October). A Philosophy for CNS Radiotracer Design. *Accounts of Chemical Research*, *47*(10), 3127–3134. doi: 10.1021/ar500233s

Yu, A. J., & Dayan, P. (2005, May). Uncertainty, Neuromodulation, and Attention. *Neuron*, *46*(4), 681–692. doi: 10.1016/j.neuron.2005.04.026

Zilles, K., & Palomero-Gallagher, N. (2017, September). Multiple Transmitter Receptors in Regions and Layers of the Human Cerebral Cortex. *Frontiers in Neuroanatomy*, *11*. doi: 10.3389/fnana.2017.00078

Appendix

Table 2: Mean results by study for a linear regression analysis comparing explicit subject reports and ideal observer estimates. Note that probability is expressed on the same scale in the ideal observer and subjects, making the regression coefficient interpretable and equal to 1 had subjects been ideal. By contrast, confidence is expressed in precision unit for study 1, 2 and 3, whose scale is in theory not bounded, whereas subject reported their confidence on a qualitative scale that was arbitrarily remapped to the unit interval for the purpose of the analysis; therefore the corresponding regression coefficient is not interpretable (and much lower than 1 due to the remapping).

		R2	**coef**	**sd**	***p***
Study 1	confidence	0.038	0.105	0.109	4.70x10-05
	prediction	0.641	0.961	0.291	3.84x10-15
Study 2	confidence	0.053	0.136	0.121	7.83x10-11
	prediction	0.563	0.724	0.324	4.53x10-22
Study 3	confidence	0.059	0.177	0.120	1.44x10-06
	prediction	not applicable (see text)			
Study 4	confidence	0.109	0.425	0.236	4.65x10-19
	prediction	different comparison (see text)			

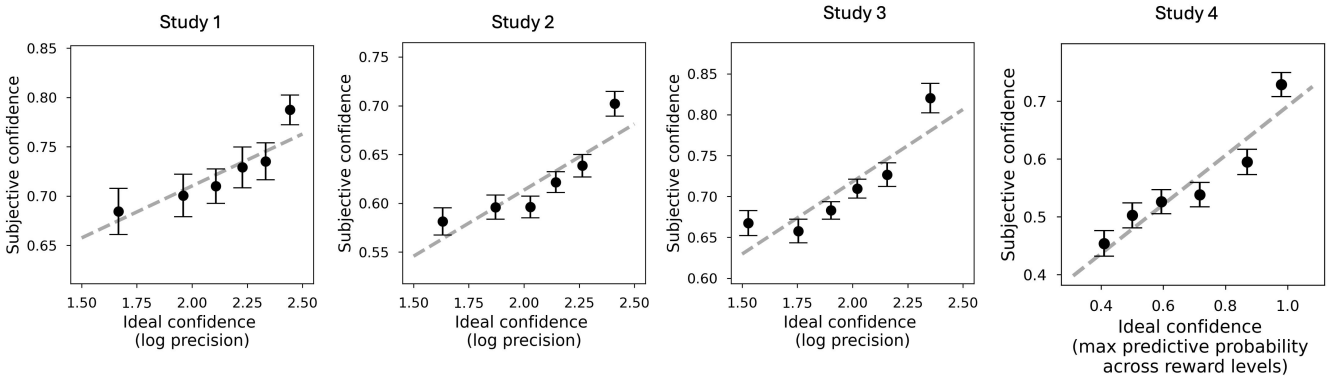

Figure 5: **Reported confidence levels sorted in 6 bins of ideal estimates**. Dots and error bars indicate the mean and S.E.M. The dashed line is a linear fit. Note that for the probability learning studies (Studies 1-3) subjective confidence (lowest and highest response bins set to 0 and 1 arbitrarily) and ideal confidence (in log SD unit) are on different scales.

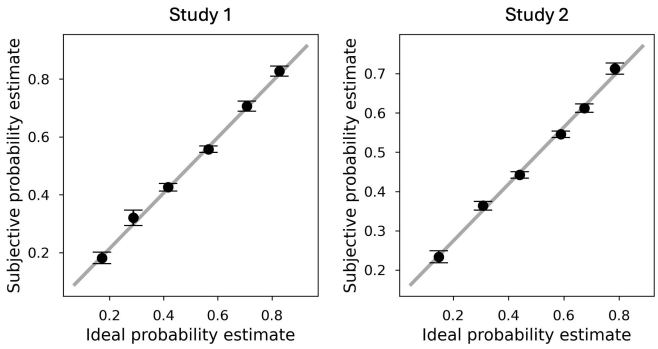

Figure 6: **Reported probability estimates sorted in 6 bins of ideal estimates**. Dots and error bars indicate the mean and S.E.M. The plain line indicates the identity. Probability reports are not available for study 3 and 4

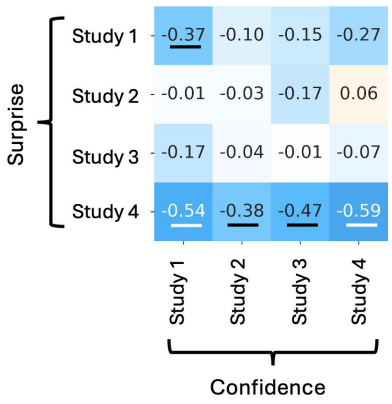

Figure 7: **Correlations between confidence and surprise maps**. Underlined correlation coefficients were significant in a comparison to a null model (spin test, see Methods; all $p<0.001$ at 1000 rotations)