Supplementary Information

Supplementary Methods

Supplementary Figures related to:

Supplementary Tables

Supplementary Methods

1: Details on task versions, related to Figure 1.

During the experiment, each participant performed two task versions of the experiment that differed in the type of predictor stimuli. The two versions were performed on different days. In one version, faces represented advisors that predicted a target accurately or less accurately. In a second version, fruits represented the target that fell with a small or large distance from a tree. Differences between versions aimed to compare social and non-social conditions. Both versions only differed in their framing but were matched in statistical properties and counterbalanced across subjects. Here, reported results are averaged across conditions as they do not differ between versions. The depiction of the experimental design summarises key features that were common to both social and non-social conditions.

2: Alternative computational models, related to Figure 2, Extended Data Figure 1 and 2.

This section relates to the Method: Alternative computational models section and to Extended Data Figure 1 and 2. Here, we describe in detail how these alternative computational models were derived. We constructed two alternative computational models: a) a Bayesian model with informative priors (Extended Data Figure 1) and b) a reinforcement learning model (RL) tracking payoff history (Extended Data Figure 2).

Bayesian model with informative priors

There is a possibility that by selecting different predictors across time, participants form a representation of the likelihood of sigma values that can be associated with predictors. In consequence, participants' beliefs at a block start would not be represented by a uniform prior distribution but rather by a prior distribution that had been informed by past observations. To test this, we developed a Bayesian model with priors that have been informed by past observations, we refer to these as informative priors (Extended Data Figure 1A). Ultimately, we compare the model fit between a Bayesian model with informative and uniform priors, i.e. referred to as the original model used in the current study.

We modified the Bayesian model to include block-wise priors that reflect the previous history of all observations (irrespective of predictor). First, to obtain informative priors for each block start, we constructed a separate Bayesian model that did not differentiate between predictor identity but instead learnt about all observations as derived from a single distribution ('average predictor learner'). We used a uniform distribution for the first block for all participants. The subsequent prior distributions at the end of each block differed across participants as the shape of the distribution was dependent on individual choice behaviour and on the number of observations within a block.

Second, we combined these informative priors with the original Bayesian model used in the manuscript. In other words, at the beginning of each block, we set the prior distribution for each predictor to the posterior of the average predictor learner from the end of the previous block (Extended Data Figure 1A, bottom illustration). Then, however, each predictor's distribution was updated according to the observations specific to each predictor (referred to as adaptive model, green colours).

We derived estimates of accuracy and uncertainty and repeated all behavioural analyses of interest. We replicated all effects of interest even when using informative priors (accuracy, t(23) = 4.7, p<0.001, d=0.96, 95% confidence interval=[0.91 2.3]; uncertainty, t(23) = 1.2, p= 0.25, d=-0.24, 95% confidence interval=[-0.77 0.21]; uncertainty x block time, t(23) = 6, p<0.001, d=1.2, 95% confidence interval=[0.83 1.73]; accuracy x block time, t(23) = 2.6, p= 0.015, d=-0.54, 95% confidence interval=[-1.1 -0.13]) (Extended Data Figure 1B). Next, we compared the model fit between the adaptive and original Bayesian model (Extended Data Figure 1C). We show that a Bayesian model using uniform priors has a better fit to choice behaviour than the adaptive Bayesian model. One reason might be that a uniform prior provides more flexibility for estimates to converge towards their true value across time.

It is important to keep in mind, that for behavioural and neural analyses, variables are constructed in relative terms, for example as the difference between left and right predictors or the difference between chosen and unchosen predictors. Hence, their relative values rather than their absolute values determine choice behaviour. When changing prior belief distributions, only the absolute values change while the relative values keep the same proportions (Extended Data Figure 1D). For this reason, our key results remain unchanged when modifying initial block-wise priors.

Next, we used a descriptive analysis to show that prior beliefs do not show a major change across blocks (Extended Data Figure 1E). We used the confidence interval size at each block start averaged across four predictors as an index of prior beliefs and compared it across all (six) blocks. A one-way ANOVA applied to the first encounters across six blocks (red line) shows that participants increase their CI size across time (Mauchly's test indicated a violation of equal variances: $x^2(14)=37.4$,p=0.001, therefore we used Greenhouse Geiser test: F(2.9,66.23) = 17.15, p<0.001, $\eta^2=0.9$). This effect however is mainly driven by the difference between the first and the remaining blocks. Excluding the first block from the analysis did not show credible evidence for a change of CI sizes during the first encounters across blocks (Mauchly's test indicated a violation of equal variances: $x^2(9)=23$,p=0.006, therefore we used Greenhouse Geiser: F(2.6,92) = 2.5, p =0.076, $\eta^2=0.71$, Bayes $\text{factor}_{10}=0.91$, error%=0.36). Participants performed practice trials before the actual experiment during which they encountered very accurate predictors. Although participants were instructed that these observations were not reflective of the real experiment, nevertheless beliefs prior to the first block might have been impacted by observations made during these practice trials. In conclusion, analysis of the CI does not support the contention that participants narrow their initial expectation about the predictors over the course of the experiment. This suggests further that a very broad initial prior is a plausible assumption for the Bayesian modelling and suggests why a Bayesian model using a uniform prior is a better model fit for choice behaviour.

Reinforcement learning model tracking payoff history

Next, we compared the original Bayesian model to a reinforcement learning model (RL) (Extended Data Figure 2). We constructed an RL that learnt about the payoff history associated with each predictor. The payoff scheme used in this experiment reflects the participants' beliefs in the accuracies and certainties associated with chosen predictors, however it may itself exert an additional independent effect on behaviour and neural activity. For every subject, we fitted a standard RL model to each predictor's payoff history to estimate the expected value associated with that predictor at a given trial. On every trial t, the expected value associated with a predictor was updated using a prediction error (PE) based learning rule with a learning rate α as a free parameter:

$$\text{Value}_{(t+1,\ \text{predictor})} = \text{value}_{(t,\ \text{predictor})} + \alpha * \text{PE}_{(t,\ \text{predictor})}$$

The PE was calculated according to the difference between the payoff and the value estimate for the specific predictor at the current trial t:

$$\text{PE}_{(t+1,\text{predictor})} = \text{payoff}_{(t,\ \text{predictor})} - \text{value}_{(t,\ \text{predictor})}$$

For behavioural analyses, we calculated a decision variable (DV) for each trial that reflected the difference in the expected values between the left and right predictors:

$$\text{DV}_{\text{left-right predictors}} = \text{value}_{(\text{left predictor})} - \text{value}_{(\text{right predictor})}$$

We used a softmax function with an inverse temperature ß to calculate the probability of making a leftwards choice on each trial:

$$\text{P(left predictor)} = 1/\ (1+ \exp(-ß * \text{DV}_{\text{left-right predictors}}))$$

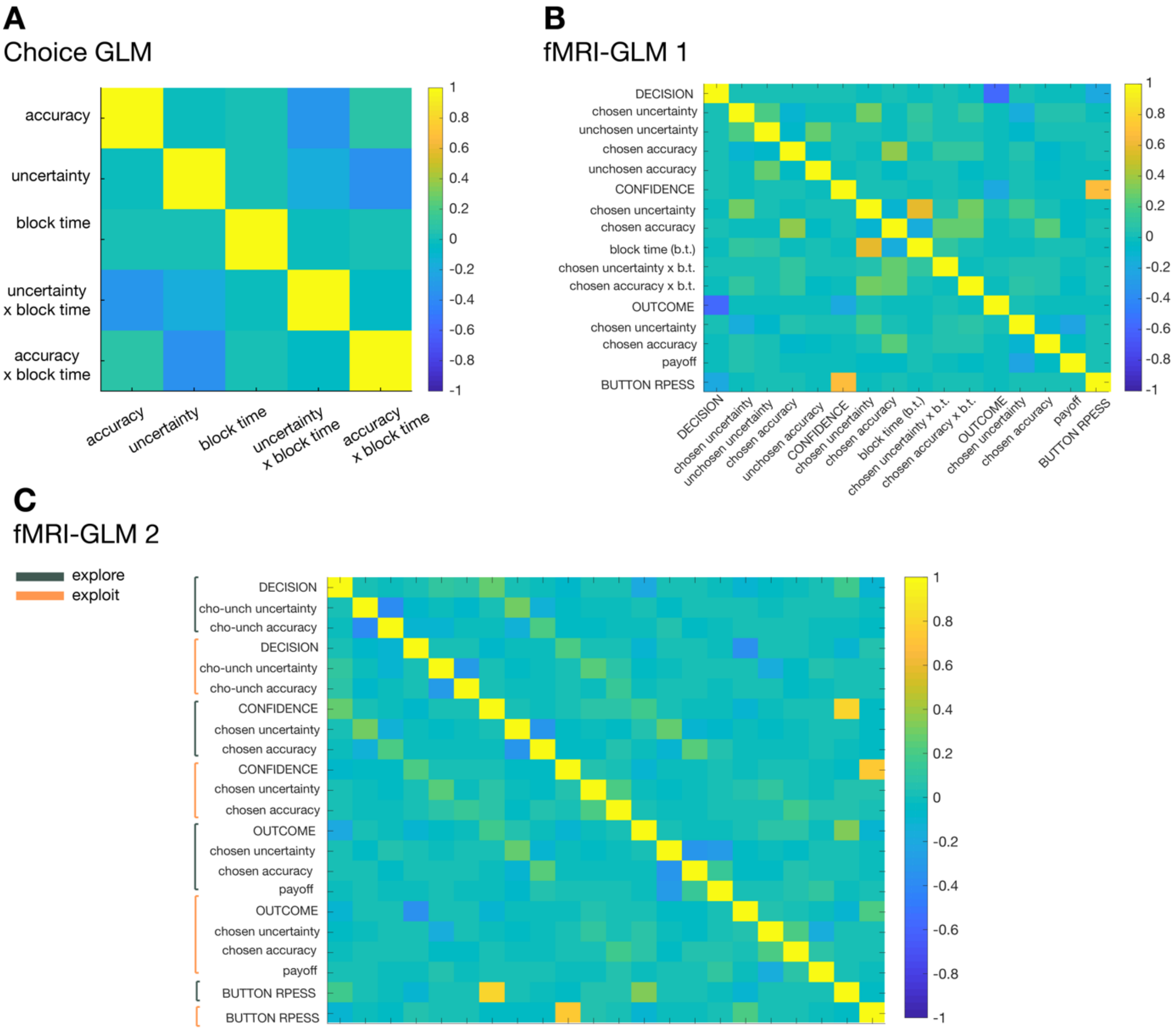
Note for neural analyses, the DV was calculated between the chosen and unchosen predictors in their respective expected value.
First, we included the RL-derived value difference between predictors on a given trial into our main GLM to investigate whether the other effects of interest remained stable. We replicated all effects of interest when controlling for RL value difference (accuracy, t(23) = 5.5, p<0.001, d=1.1, 95% confidence interval=[0.48 1.1]; uncertainty, t(23) = -3.1, p= 0.0049, d=-0.63, 95% confidence interval=[-0.75 -0.15]; uncertainty x block time, t(23) = 5.2, p<0.001, d=1.1 , 95% confidence interval=[0.49 1.13]; accuracy x block time, t(23) = -6.8, p<0.001, d=-1.4, 95% confidence interval=[-0.84 -0.44] and RL value difference, t(23) = 11.9, p<0.001, d= 2.43, 95% confidence interval=[0.7 0.99]) (Extended Data Figure 2A). This is consistent with the relative lack of correlation between variables derived from the Bayesian model and the RL model (Extended Data Figure 2B). Next, we compared the fits between models of choice behavior for all trials (Extended Data Figure 2C). There are three possible models: a GLM applied to choice behaviour first, only including regressors from a

RL model (regressor: RL value difference, yellow) second, a model only including previously used Bayesian-derived regressors (regressors: Bayesian model/original model, grey) or third, a combination of regressors of both models (regressors: RL value difference and Bayesian model regressors). A GLM model that combined (combination shown in red bars) regressors of the RL value model and the original Bayesian model was the best fit for choice behaviour, supporting the relevance of value-based and information-based variables in explaining choice behaviour.

As the combination of both RL-derived variables and Bayesian-derived variables showed the best model fit, we repeated a whole-brain analysis including regressors from both models to apply to neural data across all trials (see Methods, fMRI-GLM1 for details on the original model, to which we added RL value difference as regressor). This analysis had two aims: first, to replicate a domain general prediction difference across trials when controlling for variance explained by the RL value difference and second, to investigate brain regions associated with RL value difference. We replicated a domain general prediction difference in vmPFC even when we incorporated the RL value difference into the GLM model (Extended Data Figure 2D; top: domain general prediction difference; cluster-corrected, MNI x/y/z-peak coordinates: [4 44 -4]; z-value: 4.05). We did not find any activation significantly cluster-corrected for RL value difference. However, the activation that was strongest was located within vmPFC (Extended Data Figure 2D; bottom panel: RL value difference; MNI x/y/z-peak coordinates: [-10 46 -2]; z-value: 2.27]. In conclusion, RL value terms complement the Bayesian model but do not substitute for the Bayesian model terms as an explanation of behaviour; participants' beliefs in the accuracy and uncertainty of a predictor explained additional variance in choice behaviour above and beyond that explained by their choice value estimates.

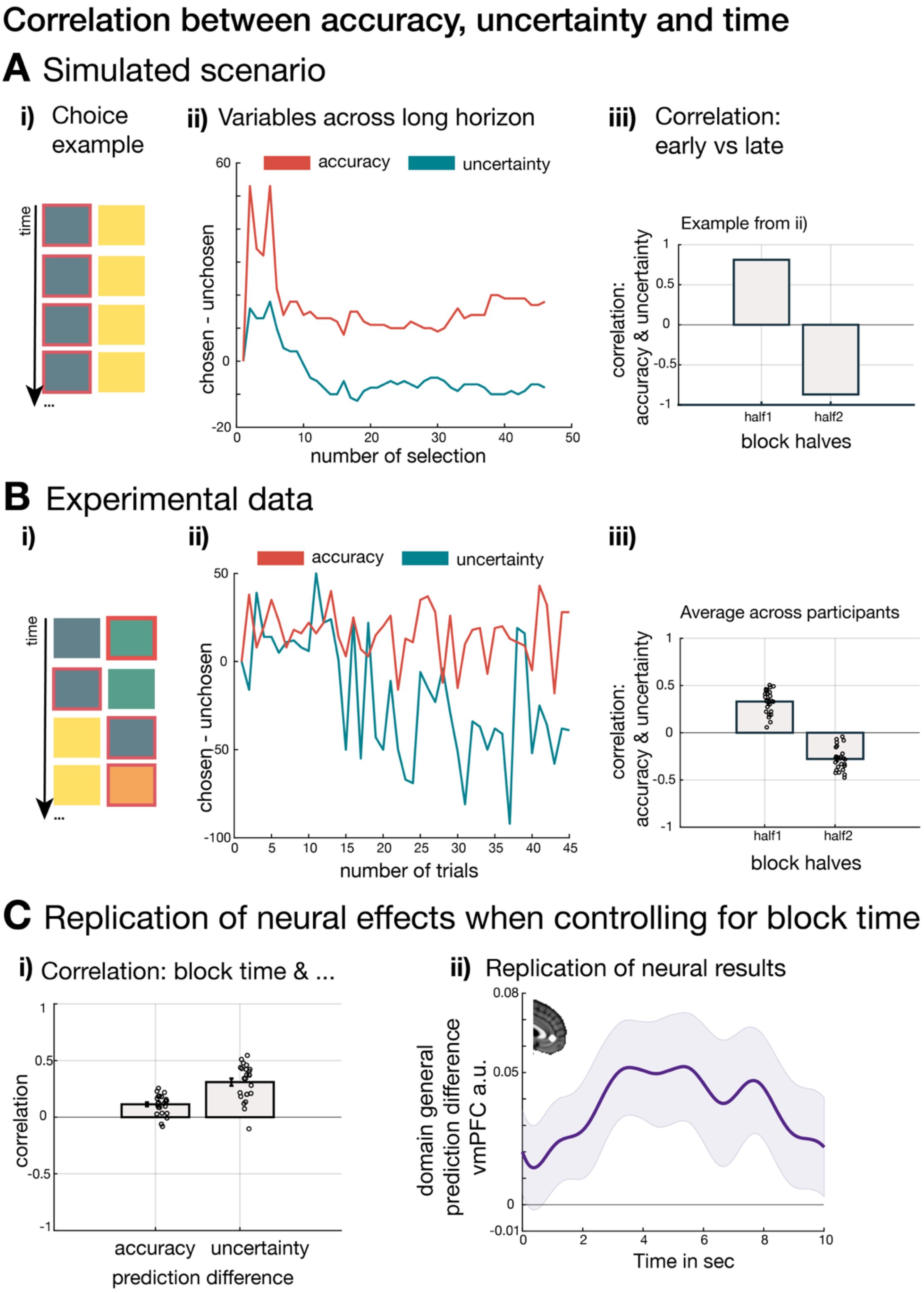
Supplementary Figures

3: Information related to Method/ experimental design



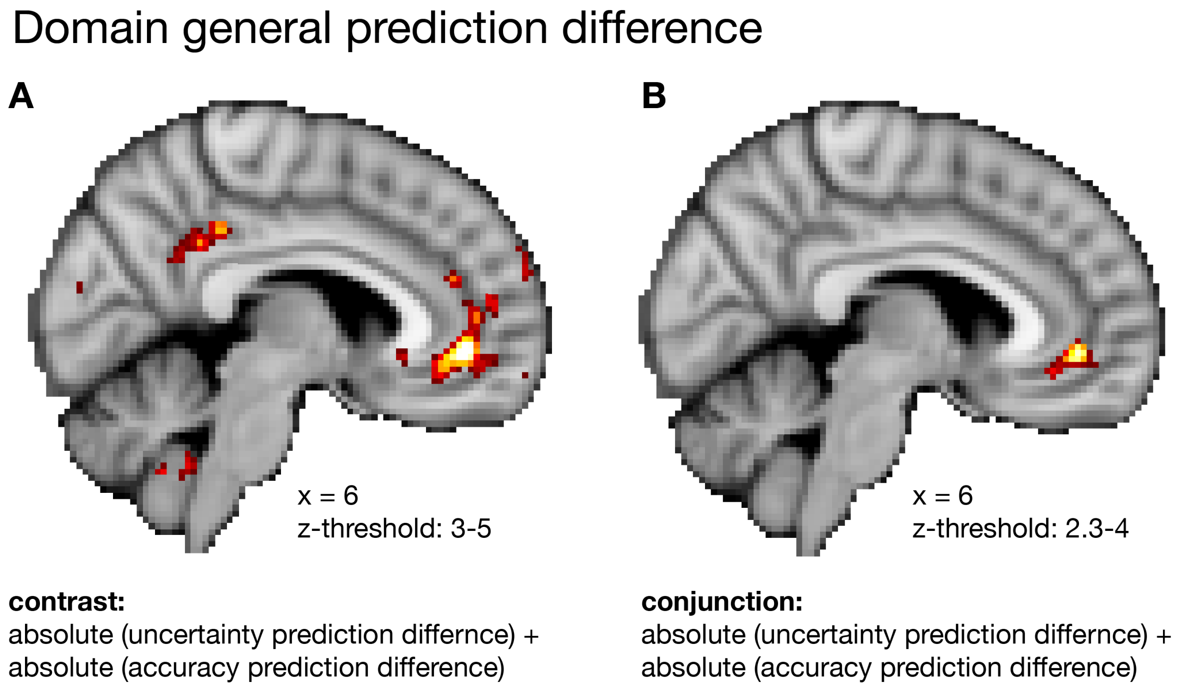
Supplementary Figure 1, related to Figure 3 and Figure 5. Correlation between variables included in behavioural and neural analyses.

Correlations represent an average across all participants: first, correlations were averaged for each subject across sessions and then averaged across subjects. All regressors were normalized before they were applied to the data. Interaction terms were treated identical to the behavioural analyses: they were normalized before and after multiplication before including them into the analysis. See methods part for further details on the specific analyses. **(A)** Regressors used for behavioural choice GLM1. **(B)** Regressors used in neural fMRI-GLM1. **(C)** Regressors used in neural fMRI-GLM2.



Supplementary Figure 2, related to Figure 4D. Correlation between accuracy, uncertainty and time. (A) The aim of the study is to show that choice behaviour and neural activity are influenced by two independent beliefs, the accuracy and uncertainty associated with an option. To be able to examine this, it is essential that variables are sufficiently independent from each other. Here, we elaborate how experimental features of the design helped us to decorrelate accuracy and uncertainty across trials within the experiment. To make the impact of experimental features clear, we first simulate a scenario during which uncertainty and accuracy estimates are highly correlated and then we show that this simulated scenario does not apply to the current study and further we show why it does not apply. **(A-i)** Simulated scenario: If participants were to be repeatedly offered the same pair of predictors and if they were to repeatedly choose the good predictor over the bad one (illustration A-i), then accuracy and uncertainty would be highly correlated across time. **(A-ii)** The simulated scenario shows the development of accuracy and uncertainty (difference between chosen and unchosen predictor) across a long horizon when repeatedly selecting the same better predictor over the same bad predictor. **(A-iii)** In such a case, accuracy and uncertainty estimates would correlate highly, as both estimates decrease across time. **(B)** However, the procedure used in our experiment was quite different. **(B-i)** During the experiment, participants were not offered the same two predictors at each trial. Instead, blocks include four predictors of different sigma values. Further, predictors were introduced at slightly different times during the experiment and the net result of this was that the set of predictors available on each trial were associated with a range of accuracy and uncertainty estimates. **(B-ii)** Changing the predictor offers across time creates a variety of prediction differences during early and late phases within a block (illustrated with data from an example participant). This can be seen in the variety in values across time. This variety is much greater than in the first scenario illustrated in A-ii. **(B-iii)** In consequence, accuracy and uncertainty prediction differences are decorrelated across time within a block and that remains true even if we examine the first and second halves of blocks separately. **(C)** In a more general sense, one could ask how variables of accuracy and uncertainty prediction differences develop across time or whether they are correlated with a linear trend, i.e. a block time variable. **(C-i)** Here, we show that both prediction differences of accuracy and uncertainty correlate respectively, r=0.12 (95% confidence interval = [-0.3 0.5]) and r=0.3 (95% confidence interval = [-0.12 0.63]) with 'block time' across participants. Hence, results depicted in the manuscript are not confounded by a 'block time' effect. **(C-ii)** To show that 'block time' does not account for the variance explained by the combination of accuracy and uncertainty prediction differences, we applied a similar GLM to fMRI-GLM1 (see Methods) across all trials including uncertainty and accuracy prediction differences and 'block time' as variables. Thereby, we controlled for variance explained by 'block time'. A time course analysis was extracted from a previous cluster-corrected vmPFC ROI and showed that when controlling for 'block time', a domain general prediction difference could still be replicated in vmPFC (leave-one-out; t(23)=3, p=0.001, d=0.76, 95% confidence interval=[0.053 0.18]). (n = 24; error bars are SEM across participants).

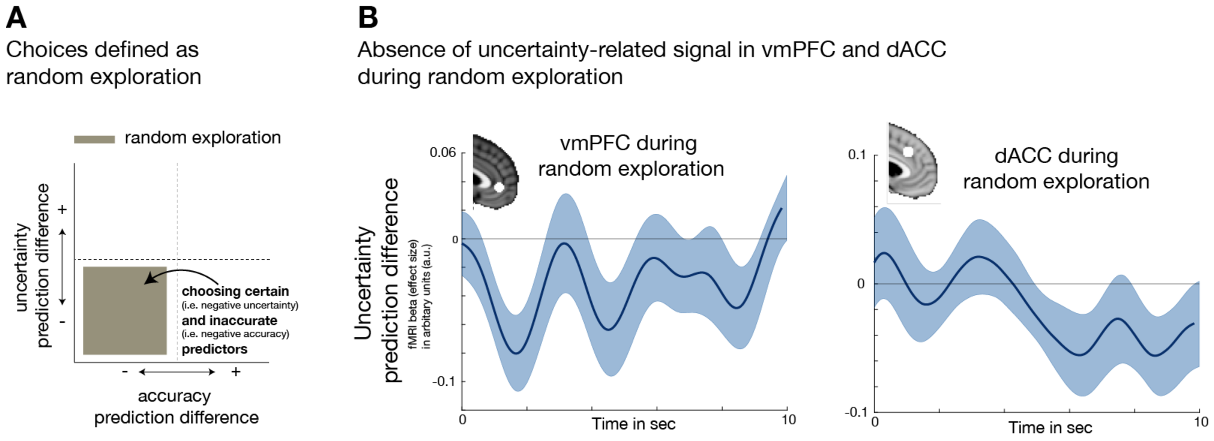
4: Information related to neural results



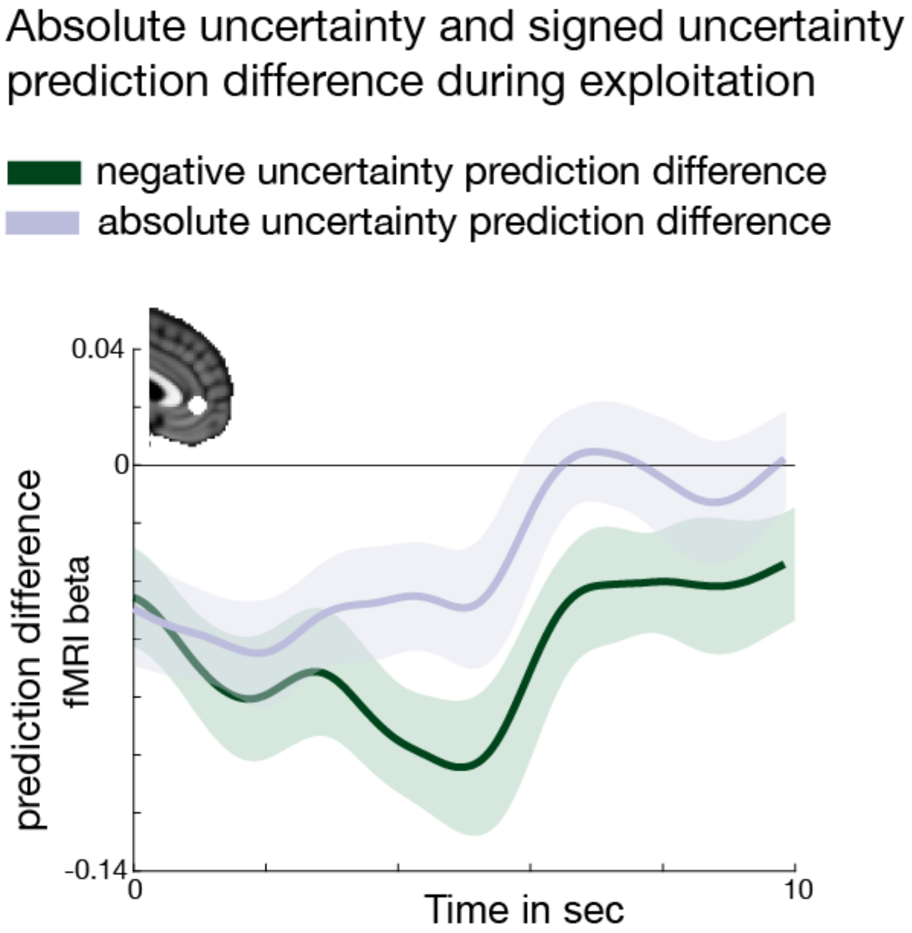
contrast:
absolute (uncertainty prediction differnce) +
absolute (accuracy prediction difference)

conjunction:
absolute (uncertainty prediction differnce) +
absolute (accuracy prediction difference)

Supplementary Figure 3, related to Figure 4A. Domain general prediction difference across all trials.
A "domain general prediction difference" correlated with activation in vmPFC. (**A**) The effect was tested, firstly, by calculating the mean across both absolute contrasts of uncertainty (Figure 4A-i) and accuracy prediction differences (Figure 4A-ii). (**B**) Secondly, we tested the effect as conjunction of shared variance between these two absolute contrasts. See Methods section for more details. (n=24; whole-brain effects were family-wise error cluster corrected with $z > 2.3$ and $p < 0.05$, see Table S17 for cluster peak coordinates. Z-thresholds indicated in figure panels are used for visualisation only.

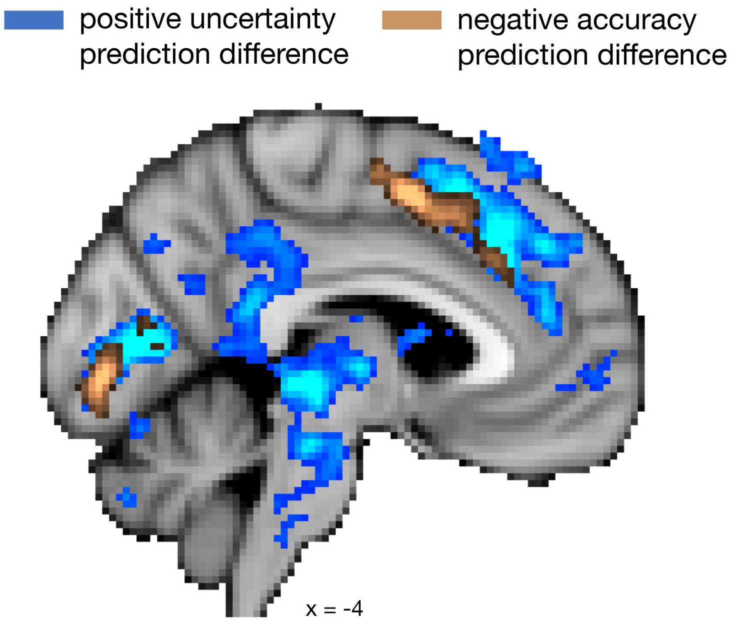


Supplementary Figure 4, related to Figure 4,5. Absence of uncertainty prediction difference during random exploration in vmPFC and dACC.
We showed that vmPFC and a more widespread network centred on dACC represents uncertainty prediction difference during directed exploration (Figure 5). Here, we show that this neural signature is unique to directed exploration and is absent during random explorative selections between predictors. **(A)** Selections during random explorative phases were defined as being directed towards inaccurate (negative accuracy prediction difference) and less uncertain predictors (negative uncertainty prediction difference). **(B)** We show that during random exploration, neither vmPFC (leave-one-out test, t(23) = -0.47, p= 0.64, d=-0.096, 95% confidence interval = [-0.09 0.05], Bayes factor$_{10}$ = 0.24, error %=0.034) nor dACC (leave-one-out test, t(23) = -1.1, p= 0.28, d=-0.23, 95% confidence interval = [-0.11 0.032], Bayes factor$_{10}$ = 0.37, error %=1.21e-4) represented an uncertainty prediction difference. (n = 24; error bars are SEM across participants).

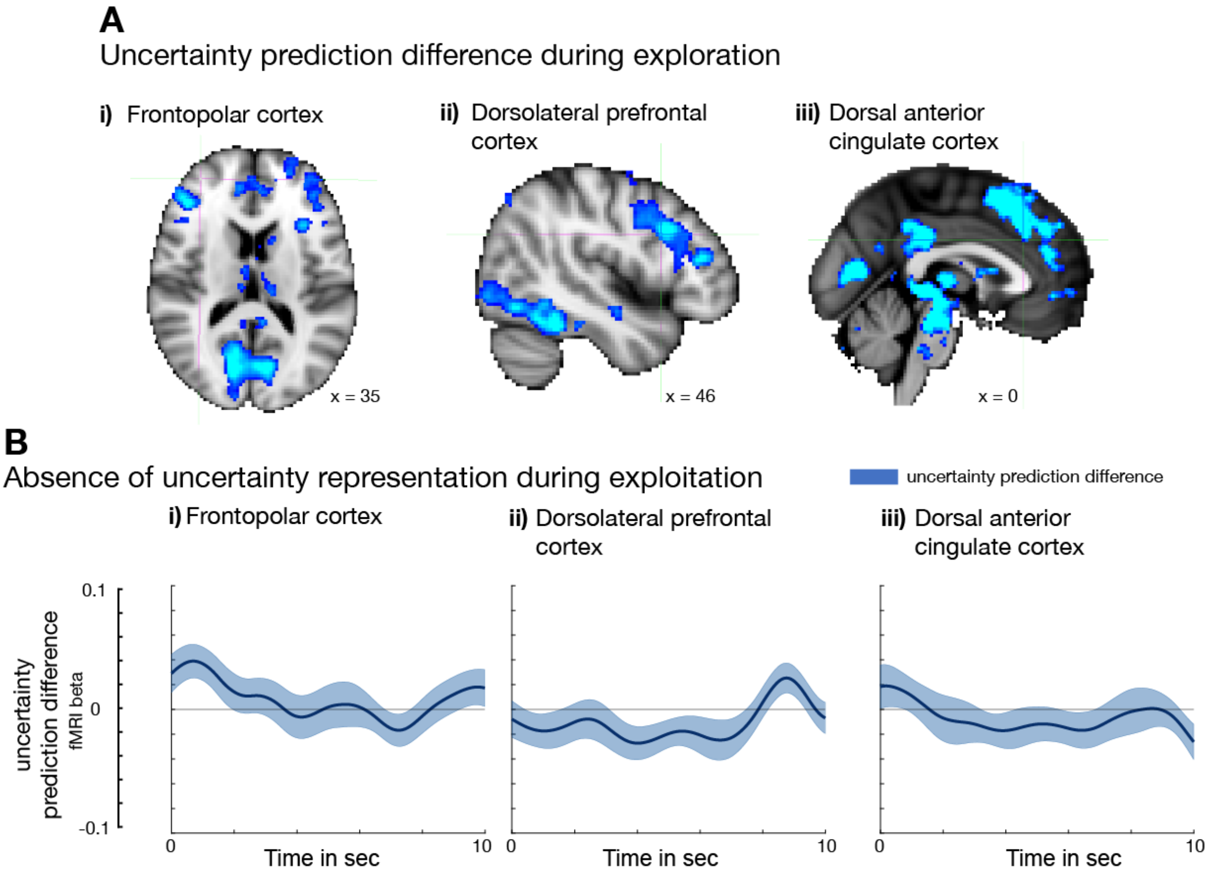


Supplementary Figure 5, related to Figure 4C. Polarity change of uncertainty between behavioural modes cannot be explained by an absolute uncertainty prediction difference.

We have shown that vmPFC represents the uncertainty between predictors with opposing signs during exploration (positive uncertainty prediction difference) and exploitation (negative uncertainty prediction difference). However, it is possible to construct an alternative account of the observed sustained uncertainty activation across behavioural modes, according to which it reflects an "absolute uncertainty prediction difference" within both modes. To test for this alternative, we included the absolute uncertainty prediction difference as an additional regressor into the previously used ROI-analysis (fMRI-GLM2). During exploration, both the absolute uncertainty and uncertainty prediction differences are identical as in both cases, trials are defined by a positive uncertainty prediction difference, i.e. choosing the more uncertain predictor. Therefore, we analysed the exploitation phase to dissociate between the two alternative explanations. Trials within exploitation are defined by both, positive and negative uncertainty prediction differences and therefore the polarity switch should not be affected by the additional regressor of an absolute uncertainty prediction difference. We can replicate the negative uncertainty prediction difference during exploitation when controlling for the absolute uncertainty prediction difference ($t(23) = -4.5$, $p<0.001$, $d=-0.92$,95% confidence interval=[-1.56 -0.06]; $n = 24$; error bars are SEM across participants).

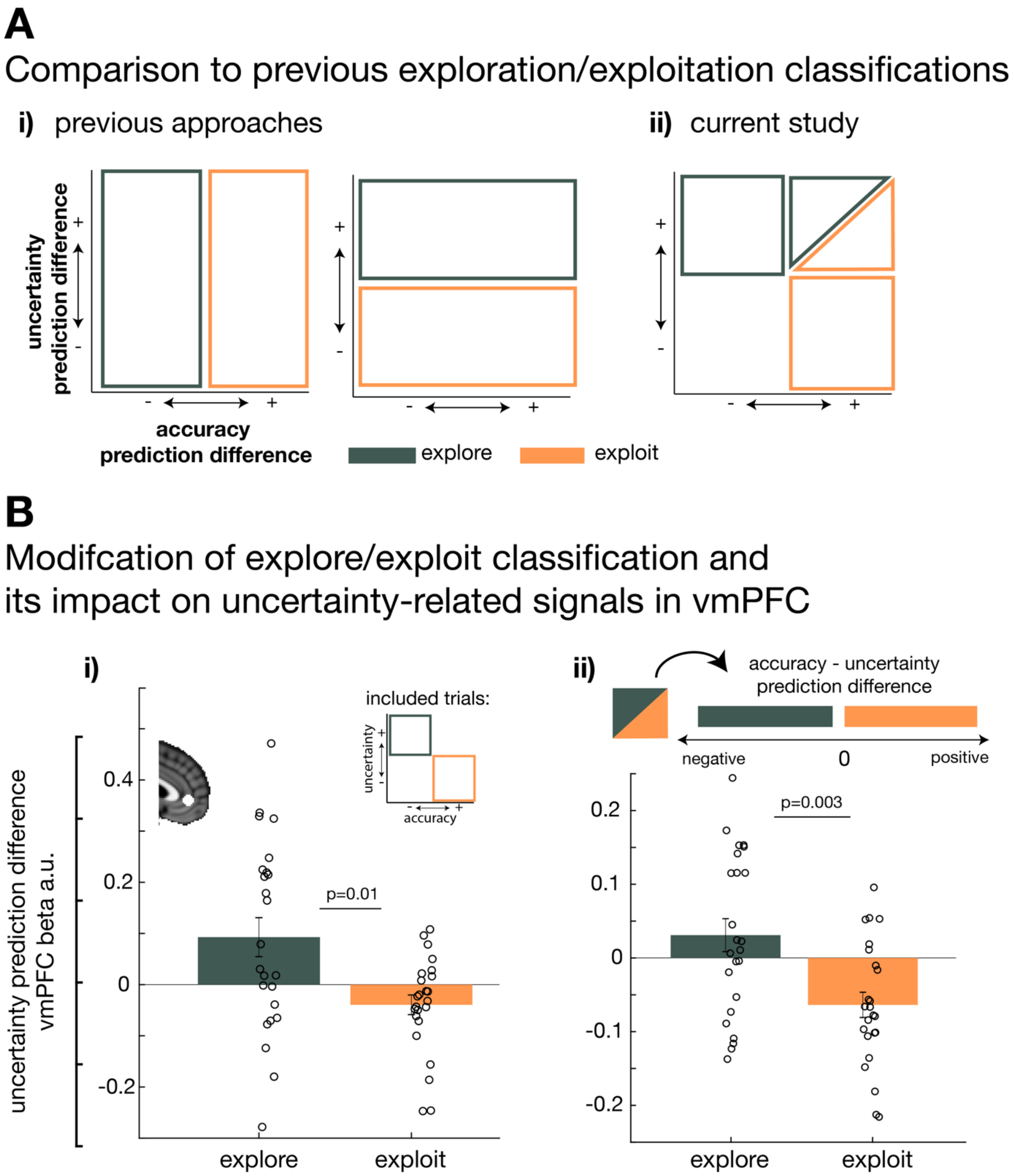

positive uncertainty prediction difference
negative accuracy prediction difference
x = -4

Supplementary Figure 6, related to Figure 5. Dorsal anterior cingulate cortex encodes accuracy and uncertainty predictor differences during exploration.
During exploration trials, in the decision phase, dorsal anterior cingulate cortex (dACC) exhibited a distinctive neural activity pattern by representing both features of the chosen predictor, as opposed to the unchosen predictor, that defined the exploration period (compare explore/exploit trial separation in Extended Data Figure 4): a positive uncertainty and negative accuracy prediction difference. (n=24; whole-brain effects family-wise error cluster corrected with $z > 2.3$ and $p < 0.05$.)



Supplementary Figure 7, related to Figure 5. Frontopolar cortex, dlPFC and dACC represent uncertainty prediction difference during exploration, but not exploitation and thereby show a distinct profile compared to vmPFC.

(A) Our analyses also identified regions that have been previously associated with exploration, namely frontopolar cortex and dorsolateral prefrontal cortex (dlPFC). Frontopolar cortex has been causally linked to the exploration of uncertain options[3](MNI x/y/z-peak coordinates: [35,50,15]). Similarly, dorsolateral prefrontal cortex has been associated with the mean uncertainty between two options[4](MNI x/y/z-peak coordinates: [46 14 28]). dACC is active during exploration[5] and choosing alternative options during foraging[6]. Fixation crosses show the peak of brain clusters reported in these previous studies. In our study, our contrast of uncertainty prediction difference during the exploration decision phase highlights similar regions in frontopolar cortex (A-i: MNI x/y/z-peak coordinates: [-40,50,14]), dlPFC (A-ii: MNI x/y/z-peak coordinates: [38 10 28]), and dACC[6] (A-iii: MNI x/y/z-peak coordinates: [0, 28,30])). **(B)** We illustrate the effects of uncertainty prediction difference during exploitation. We used an ROI approach to visualise the (absence of an) uncertainty prediction difference in frontopolar cortex and dlPFC. ROIs were extracted according to cluster-corrected regions that were close to previous reported areas and associated with the uncertainty prediction difference during exploration (see coordinates in previous brackets). There was no polarity change in the representation of uncertainty when comparing exploration to exploitation, because none of the brain areas represented a negative prediction difference during exploitation (no cluster survived cluster-correction with z>2.3; time courses are shown for illustration). These areas, as well as vmPFC are involved in representing uncertainty during exploration, however their profile diverges during exploitation as a polarity change is only observed in vmPFC. (n=24, whole-brain effects were family-wise error cluster corrected with $z > 2.3$ and $p < 0.05$.)



A

Comparison to previous exploration/exploitation classifications

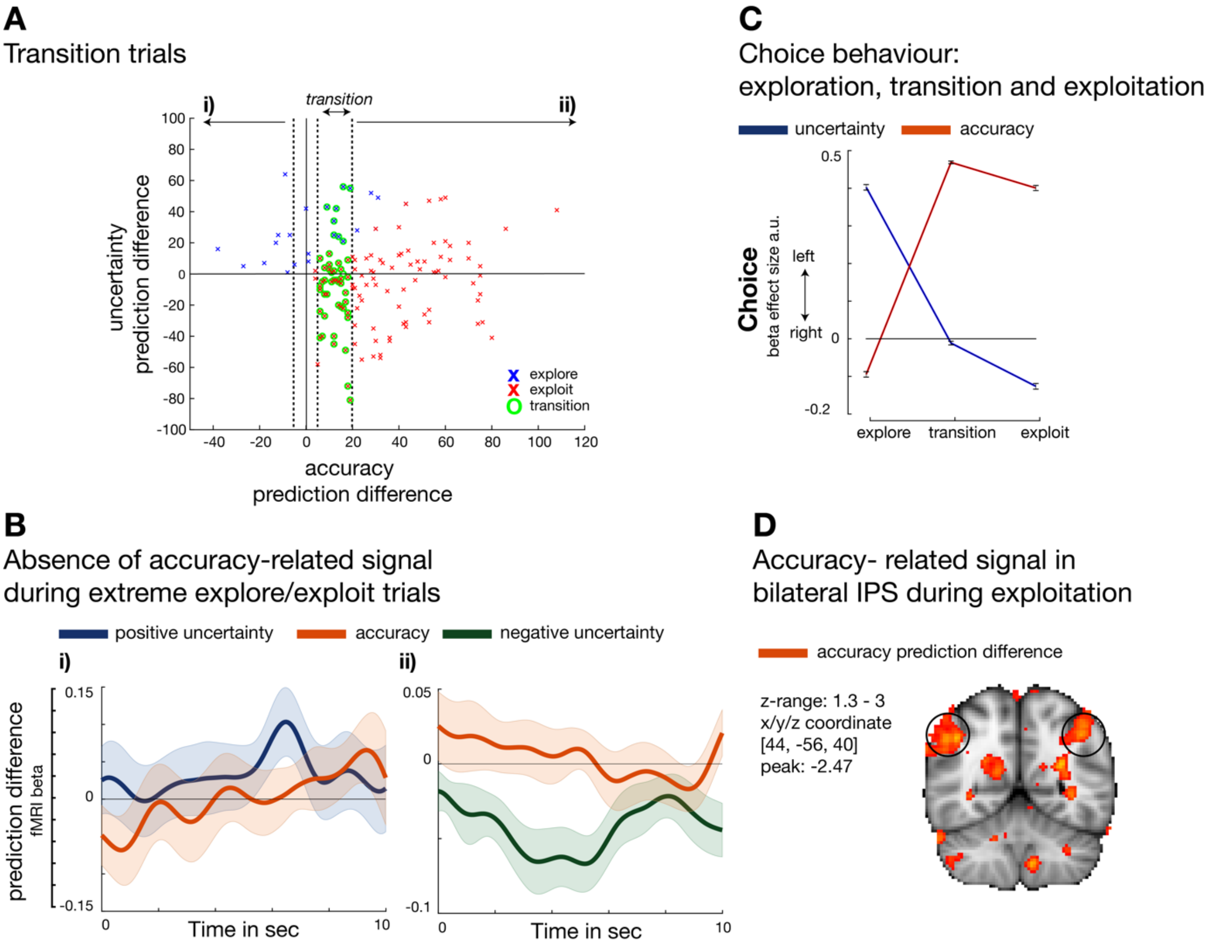
i) previous approaches

ii) current study

B

Modifcation of explore/exploit classification and its impact on uncertainty-related signals in vmPFC

Supplementary Figure 8. Exploration/ Exploitation classification and the robustness of neural results when modifying their boundaries. (A-i) Here, we elaborate on the logic of exploration/exploitation classification used in the study. For detailed information on how trials were classified, see Extended Data Figure 4. (1) Exploration/ exploitation has often been described as a dichotomy or a trade-off, representing two distinct behavioural modes[4,5]. (2) Previous studies have defined exploitation and exploration, respectively as choosing options with higher or lower expected values, respectively[5] (A-i, left panel). (3) Other studies, however, have defined exploration as being directed towards options with higher relative uncertainty[4] (A-i, right panel). However, we merged all three principles for distinguishing between explore-exploit trials and define exploration and exploitation in reference to both uncertainty and accuracy/value as used in the current study **(A-ii)**. **(B-i)** We tested the robustness of the polarity change in vmPFC's representation of uncertainty during exploration and exploitation by examining the impact of modifying the classification boundary between exploration and exploitation. More precisely, trials with both a positive accuracy and a positive uncertainty prediction difference could, in fact, be allocated to either exploration or exploitation. We used a rule to classify these trials (see Extended Data Figure 4), however here we show that neural results are robust when changing the classification boundaries. We removed the entire quadrant of trials (positive accuracy, positive uncertainty prediction differences) from our analyses and were still able to replicate the polarity switch of uncertainty prediction difference in vmPFC when comparing exploitation and exploration (paired t-test, t(23) = 2.8, p=0.01, d=0.56, 95% confidence interval=[0.03 0.23]). **(B-ii)** Another robustness check can be applied to the classification of trials that had a small difference between accuracy and uncertainty prediction differences (below an arbitrary value of 5). These trials were previously classified into both exploration and exploitation categories (Extended Data Figure 4). When excluding these trials from the analysis, a polarity change of uncertainty prediction difference in vmPFC could be replicated (paired t-test, t(23) = 3.5, p=0.003, d=0.67, 95% confidence interval=[0.035 0.15]). (n = 24; error bars are SEM across participants).



Supplementary Figure 9, related to Figure 7. (The absence of) accuracy-related signals during exploration, exploitation, and the transition between exploration and exploitation. These results relate to Figure 7 and support the contention that vmPFC processes accuracy prediction differences during the transition period between exploration and exploitation. (**A**) Example of one subject showing explore, exploit and transitional choices (one data point is one choice). Exploratory (blue crosses) and exploitative (red crosses) choices were categorised as described in Extended Data Figure 4A. Transitional trials (green circles) were identified independently of the criteria used to identify explore/exploit trials; they were defined by a positive but small accuracy prediction difference in the range of [5 20], which resulted into approximately 20% of the trials (see methods for details). **(B)** To further show the specificity of accuracy prediction difference in vmPFC during transitional trials, we show that vmPFC lacks an accuracy prediction difference signal during strong inaccuracy-driven (B-i; trials on the left-hand side of the transition trials) and strong accuracy-driven (B-ii; trials on the right-hand side of the transition trials) selections. These selections are likely to occur during very exploratory (A-i) or exploitative (A-ii) periods. Time courses of positive uncertainty (blue) and negative uncertainty (green) or accuracy (orange) prediction differences are plotted in the previously defined vmPFC region that showed a cluster-corrected effect of, respectively, accuracy or uncertainty prediction differences across all trials (see Figure 4A; we used the respective other contrast for ROI selection to guarantee independence of ROI analysis from ROI selection). Accuracy-processing is specific to the transitional period (Figure 7B) and does not occur during the period before (B-i: t(23)=-0.84, p=0.41, d=-0.17, 95% confidence interval=[-0.13 0.055], Bayes factor_{10}=0.296,%error=0.037) of after (B-ii: t(23)=-1.3, p=0.21, d=-0.27, 95% confidence interval=[-0.06 0.02], Bayes factor_{10}=0.447,%error=1.178e-4). During extreme inaccuracy-driven choices that are prominent during exploration, uncertainty prediction difference had a marginally significant effect on vmPFC activity (t(23) = 2, p= 0.059, d=0.4, 95% confidence interval= [-0.004 0.2], Bayes factor_{10}=1.14,%error=1.002e-4) while during extreme accuracy-driven trials, which are prominent during exploitation, vmPFC activation represents a negative uncertainty prediction difference (t(23) = -3.8, p<0.001, d=-0.8, 95% confidence interval = [-0.12 -0.03]. Note that we used here an even stricter cut-off to define exploration/exploitation periods (compared to Extended Data Figure 4), and it was still possible to show support for a change in the polarity of vmPFC signals, from positive to negative uncertainty, during exploration and exploitation, respectively (compare to Figure 4C). **(C)** Here, we show how choice behaviour during each learning period (exploration, exploitation and their transition) relate to neural findings in vmPFC. We applied a GLM including accuracy and uncertainty differences between left and right predictors to predict leftwards choices in each block period (exploration, transition, and exploitation). We observe an analogous result in choice behaviour during each learning period as was observed neurally in vmPFC: we observe a polarity change in the direction of influence of uncertainty on behaviour as well as neural activity when transitioning from exploration to the next block period. The effect of accuracy increases across learning phases, peaking within the transition period, before decreasing during exploitation. A fuller summary of the behavioural results shows that, choices are initially directed towards uncertain options, then accurate options, but the influence of accuracy diminishes towards the final exploitation period. Note that this GLM serves as a manipulation check only, as trials were separated into each category according to the difference between accuracy and uncertainty prediction differences (therefore standard errors are small). **(D)** In the present task, during exploitation, activity in parietal cortex reflected accuracy prediction difference (MNI x/y/z-peak coordinates: [44,-56,40]; z-value = -2.47). This is in line with previous studies, which have reported that activity in frontal regions such as vmPFC that is related to value difference is most important during difficult decisions[7] that occur at an earlier rather than later stage in task experience[8] (n = 24 for analyses shown here).

Supplementary Tables

5: Peak- coordinates of cluster-corrected whole brain effects

Supplementary table 1, related to Figure 4A. These cluster-corrected coordinates relate to results depicted in Figure 4A of the main manuscript.

Contrast	Region	Peak Coordinates x/y/z (in mm MNI Space)	Z Value
Uncertainty prediction difference: chosen - unchosen	Ventromedial prefrontal cortex	4 44 -2	-4.44
	Occipital cortex	-8 -70 -4	-3.59
Accuracy prediction difference: chosen - unchosen	Ventromedial prefrontal cortex	6 44 -4	3.98
	Superior temporal gyrus	-52 -2 -12	3.38

Family-wise error cluster corrected, $z > 2.3$, $p < 0.05$

Supplementary table 2, related to Figure 5. These cluster-corrected coordinates relate to results depicted in Figure 5 of the main manuscript.

Contrast	Region	Peak Coordinates x/y/z (in mm MNI Space)	Z Value
Exploration: uncertainty prediction difference	Ventromedial prefrontal cortex	0 46 -4	3
	Dorsal Anterior Cingulate Cortex	-4 30 34	5.44
	Dorsal lateral prefrontal cortex (right)	38 10 28	5.75
	Frontopolar cortex (left)	-40 50 14	3
	Superior temporal gyrus	48 -10 -14	4.38
	Angular gyrus	30 -52 42	3.5
	Fusiform gyrus	30 -76 -12	7.1
Exploitation: uncertainty prediction difference	Ventromedial prefrontal cortex	-4 46 -2	-4.26
	Anterior insular (right)	34 18 -10	-4.57
	Ventral tegmental area	0 -22 -22	-4.11
	Occipital cortex	-24 -72 -12	-3.79
	Occipital cortex	26 -80 -12	-3.99
Exploration – Exploitation: uncertainty prediction difference	Ventromedial prefrontal cortex	-2 46 -4	5.1
	Dorsal Anterior Cingulate Cortex	-2 28 38	5.9
	Fusiform gyrus	30 -76 -12	9
	Cerebellum	-2 -54 -36	6.7
	Superior temporal gyrus	48 -8 -14	4.9

Family-wise error cluster corrected, $z > 2.3$, $p < 0.05$

Supplementary table 3, related to Figure 4A. These cluster-corrected coordinates relate to results depicted in Supplementary Figure 3 of the supplementary material.

Contrast	Region	Peak Coordinates x/y/z (in mm MNI Space)	Z Value
Domain general: (accuracy prediction difference – uncertainty prediction difference); absolute of each individual contrast	Ventromedial prefrontal cortex	6 44 -4	5.77
	Posterior cingulate cortex	-4 -38 38	4.1
	Dorsolateral prefrontal cortex	50 36 -4	4.05
	Superior temporal gyrus	-60 -4 -10	3.84
	Superior temporal junction	-64 -44 12	4.07
	Brainstem	8 -48 -42	4.19
Conjunction analysis: (accuracy prediction difference – uncertainty prediction difference); absolute of each individual contrast	Ventromedial prefrontal cortex	6 44 -4	3.9
Family-wise error cluster corrected, $z > 2.3$, $p < 0.05$			

Supplementary References

1. Costa, V. D., Mitz, A. R. & Averbeck, B. B. Subcortical Substrates of Explore-Exploit Decisions in Primates. *Neuron* **103**, 533-545.e5 (2019).
2. Klein-Flügge, M. C., Hunt, L. T., Bach, D. R., Dolan, R. J. & Behrens, T. E. J. Dissociable Reward and Timing Signals in Human Midbrain and Ventral Striatum. *Neuron* **72**, 654–664 (2011).
3. Zajkowski, W. K., Kossut, M. & Wilson, R. C. A causal role for right frontopolar cortex in directed, but not random, exploration. *eLife* https://elifesciences.org/articles/27430 (2017) doi:10.7554/eLife.27430.
4. Badre, D., Doll, B. B., Long, N. M. & Frank, M. J. Rostrolateral Prefrontal Cortex and Individual Differences in Uncertainty-Driven Exploration. *Neuron* **73**, 595–607 (2012).
5. Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
6. Kolling, N., Behrens, T. E. J., Mars, R. B. & Rushworth, M. F. S. Neural Mechanisms of Foraging. *Science* **336**, 95–98 (2012).
7. Noonan, M. P., Kolling, N., Walton, M. E. & Rushworth, M. F. S. Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement: Re-evaluating the OFC. *European Journal of Neuroscience* **35**, 997–1010 (2012).
8. Hunt, L. T. *et al.* Mechanisms underlying cortical activity during value-guided choice. *Nat Neurosci* **15**, 470–476 (2012).