



Polarity of uncertainty representation during exploration and exploitation in ventromedial prefrontal cortex

Nadescha Trudel, Jacqueline Scholl, Miriam C Klein-Flügge, Elsa Fouragnan, Lev Tankelevitch, Marco K Wittmann, Matthew F S Rushworth

► To cite this version:

Nadescha Trudel, Jacqueline Scholl, Miriam C Klein-Flügge, Elsa Fouragnan, Lev Tankelevitch, et al.. Polarity of uncertainty representation during exploration and exploitation in ventromedial prefrontal cortex. *Nature Human Behaviour*, 2021, 5 (1), pp.83-98. 10.1038/s41562-020-0929-3 . hal-04011414

HAL Id: hal-04011414

<https://hal.science/hal-04011414>

Submitted on 2 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Polarity of uncertainty representation during exploration and
2 exploitation in ventromedial prefrontal cortex

3

4

5 Nadescha Trudel^{1*}, Jacqueline Scholl¹, Miriam C Klein-Flügge¹, Elsa Fouragnan^{1,2}, Lev
6 Tankelevitch¹, Marco K Wittmann^{1a}, Matthew FS Rushworth^{1a}

7

8 1 Wellcome Integrative Neuroimaging (WIN), Department of Experimental Psychology,
9 University of Oxford, Tinsley Building, Mansfield Road, Oxford OX1 3TA, UK

10 2 School of Psychology, University of Plymouth, PL4 8AA, UK

11

12 ^a Authors contributed equally to the work.

13

14 *Corresponding author/ Lead contact: Nadescha Trudel (nadescha.trudel@psy.ox.ac.uk).
15

16 **Abstract**

17

18 Environments furnish multiple information sources for making predictions about future
19 events. Here we use behavioural modelling and fMRI to describe how humans select
20 predictors that might be most relevant. First, during early encounters with potential
21 predictors, participants' selections were explorative and directed towards subjectively
22 uncertain predictors (positive uncertainty effect). This was particularly the case when many
23 future opportunities remained to exploit knowledge gained. Then, preferences for accurate
24 predictors increased over time, while uncertain predictors were avoided (negative uncertainty
25 effect). The behavioural transition from positive to negative uncertainty- driven selections
26 was accompanied by changes in representations of belief uncertainty in ventromedial
27 prefrontal cortex (vmPFC). The polarity of uncertainty representations (positive or negative
28 encoding of uncertainty) changed between exploration and exploitation periods. Moreover,
29 the two periods were separated by a third transitional period in which beliefs about
30 predictors' accuracy predominated. VmPFC signals a multiplicity of decision variables, the
31 strength and polarity of which vary with behavioural context.
32

33

34 **Introduction**

35
36 Humans and other animals are often presented with multiple information sources in the
37 environment that can predict different outcomes such as reward. Selecting the right predictor
38 to guide behaviour towards a particular outcome requires determining the predictors'
39 relevance in forecasting that outcome^{1,2}. Biases in information seeking can lead to mistaken
40 beliefs about the relationships that prevail in the world^{3,4}. It has been argued that animals
41 should attend either to certain predictors⁵ or, on the contrary, to uncertain predictors⁶. Certain
42 predictors might be relevant as they deliver an outcome with known prediction accuracy,
43 while attending to uncertain predictors might turn out to be more beneficial in the long-term.
44

45 We propose that which type of predictor should be considered most relevant changes during
46 different phases of the learning process. When selecting between multiple predictors for the
47 first time, selections should maximize information about available predictors. Selections
48 should be "explorative" and directed towards "uncertain" predictors. The degree of
49 exploration should also be determined by the time horizon. The time horizon is the remaining
50 time in the current context (or block in the current experiment)^{7,8}: exploration is beneficial in
51 longer compared to shorter time horizons as the knowledge gained can be used in later
52 predictor selections. Once an estimate about a predictor's accuracy has formed, selections
53 should be "exploitative" and guided by the "accuracy" and "certainty" of predictors in line
54 with reward maximization. This perspective draws on both previously formulated hypotheses
55 in the field of learning theory^{5,6}. Predictors should be selected based on the learner's
56 uncertainty about predictors' accuracy during exploration and on the learner's certainty about
57 predictors' accuracy during exploitation. Our first aim in the current study was to examine
58 whether this was the case.
59

60 Evidence for uncertainty-guided exploration has, however, recently been questioned⁹. It has
61 been argued that behaviour may sometimes appear exploratory but on closer inspection the
62 decisions that people make can be understood as having been guided by noisy estimates of
63 the values of the choices that are formed during learning. In other words, when people appear
64 exploratory, they may in fact be attempting to make exploitative decisions, but their
65 exploitative decisions are informed by noisy estimates of choice values. Our second aim was
66 to ascertain whether people genuinely engage in exploratory behaviour. This can be tested by
67 comparing rates of exploratory behaviour when past experience is held constant, but the
68 length of the future time horizon is manipulated; a longer future time horizon should elicit
69 more exploration even when previous learning opportunities are the same. Moreover, the
70 appropriateness of computational models of exploratory behaviour can also be tested by
71 obtaining more direct empirical indices of participants' subjective uncertainty; we obtained
72 such measures in our experiment. In addition, the computational model can be used to
73 identify trials in which exploratory behaviour appears to be guided by information seeking in
74 order to reduce uncertainty and trials in which exploratory behaviour simply reflects
75 randomness in the response selection or learning process⁹.
76

77 Our third aim was to examine neural activity related to exploratory and exploitative modes of
78 decision making. Many previous studies have shown that vmPFC activity reflects
79 information relevant for making value-guided decisions between choices. When making a
80 decision between choice options, vmPFC activation covaries with the decision variable that
81 guides the decision – the difference in value between the choice taken as opposed to the
82 choice rejected^{10–18}. If, as has been argued, such vmPFC activity changes reflect allocation of
83 attention to a choice option^{19–21}, then it is possible that vmPFC activity also reflects selection

84 of a predictor to guide behaviour and the reason why it is being selected to guide behaviour:
85 either because of its predictive accuracy, because of the certainty of its prediction, or because
86 of the uncertainty of its prediction.

87
88 We use a combination of behavioural analysis, computational modelling, and functional
89 magnetic resonance imaging (fMRI) to investigate at both behavioural and neural levels
90 which predictors are classified as informative, uncertain or certain, as a function of time
91 horizon, and the current behavioural mode (exploration, exploitation, or the period of
92 transition from exploration to exploitation). We designed a novel task in which participants
93 selected between multiple predictors which gave partial information about the location of a
94 target that the participants were asked to find. During the course of multiple experimental
95 blocks, participants encountered a series of potential predictors while transitioning through
96 time horizons of different lengths, inducing explorative and exploitative selections. We used
97 a Bayesian model to extract trial-by-trial estimates of participants' beliefs about both the
98 accuracy of predictors and their subjective uncertainty in those beliefs. This allowed us to test
99 their independent and complementary impact on selection behaviour and their neural
100 representations.

101
102 We found predictor selections are made as a function of time in two important ways. They
103 change as a function of the time that has elapsed since learning began and they change as a
104 function of the remaining time horizon – the time period over which the learner expects the
105 current conditions to prevail. These changes occur in tandem with the evolution of predictor-
106 related activity patterns in vmPFC. Activity in vmPFC was sensitive to participants'
107 uncertainty in their beliefs about predictors but the polarity of uncertainty representations
108 (positive or negative encoding of uncertainty) changed with the behavioural mode: a positive
109 uncertainty decision signal was present in vmPFC during exploration, while activity in the
110 same region signalled negative uncertainty during exploitation. By contrast, other brain areas
111 such as anterior cingulate cortex (ACC) and other dorsomedial frontal cortical areas,
112 signalled uncertainty only during explorative phases. We also found that exploration and
113 exploitation modes were separated by a transitional period in which beliefs about predictors'
114 accuracy predominated in their impact on vmPFC activity. These results show that a
115 predictor's relevance for guiding behaviour is not defined by a single attribute (accuracy,
116 positive or negative uncertainty), but rather it is dynamically modulated by the behavioural
117 modes of exploration, exploitation, and their transition. We show that vmPFC carries similar
118 information, representing a multiplicity of predictor selection variables, the strength and
119 polarity of which vary according to their relevance for the current behavioural mode.

120

121

122 **Results**

123

124 On each trial of the experiment (Figure 1A), participants made two decisions. First, they
125 made a binary choice between two predictors to find a target's location on a circle (decision
126 phase). Participants knew that the target location changed constantly on every trial and could
127 not be predicted directly from previous observations of its location. The only way to infer the
128 target's location was through learning how well each predictor predicted the target location.
129 Participants learned how well a predictor predicted the target by observing the distance
130 between the location estimated by the selected predictor and the true target location (which
131 we refer to as "angular error"). Importantly, predictors differed in how well they estimated
132 the target location (see S1 for details on the cover story). Selecting a better predictor led to
133 more rewards at the time of a second decision in the trial. During the second decision, the

134 predictor's estimate of the target location was revealed, and participants expressed their
135 confidence in it (confidence phase). They did this by adjusting the size of an interval around
136 the predictor's estimate such that the true target location would fall within this interval. At
137 the end of a trial, the true target location and possible points were revealed (outcome phase).
138 Participants gained points when the target fell within the chosen interval and the amount of
139 points increased when the interval size was small. This payoff scheme incentivised selecting
140 predictors with smaller angular errors in the first place. In addition to being informed about
141 whether they had won or lost, the outcome phase enabled participants to update their beliefs
142 about how well the chosen predictor estimated the target by observing the angular error.
143 Participants took part in two versions of the task that differed in their framing aspect
144 (social/non-social framing). Here, we collapsed data across versions after finding that
145 versions did not differ in the results depicted here (see details on task versions in
146 Supplementary Information).

147

148 **<insert Figure 1 about here>**

149

150 The value of exploration lies in revealing more accurate predictors, but this is only useful if
151 the time horizon (the amount of trials remaining) offers sufficient opportunity to exploit the
152 newly discovered predictors⁷. To test this idea, participants transitioned through blocks of
153 different lengths (45, 30 and 15 trials) each with a unique set of four predictors (Figure 1B-i).
154 This made it possible to examine the balance between exploration and exploitation as a
155 function of time horizon. Time horizon and current progress were explicitly cued on each
156 trial. Each block comprised two good predictors with a relatively low average angular error
157 between predicted reference point and target and two bad predictors with a higher angular
158 error (Figure 1B-ii).

159

160 *Dissociable effects of uncertainty and accuracy on predictor selections and subjective*
161 *confidence judgments*

162

163 Exploration should not only be guided by one's belief in the predictor's accuracy, but also by
164 one's own uncertainty in that belief. For this reason, we used a Bayesian model to capture
165 participants' belief distribution over the angular error between the reference point and the
166 true target location (Figure 2A-i). The trial-by-trial angular errors were derived from a
167 normal distribution centred on the true target location. Predictors' normal distributions varied
168 in their standard deviations (referred to here as sigma), making some predictors better in
169 estimating the target location (lower sigma value) and other predictors worse (higher sigma
170 value). Hence, by tracking the angular errors of a predictor, participants could estimate the
171 sigma value associated with each predictor's distribution (see Figure 2A-ii). We used the
172 Bayesian model to capture participants' beliefs in the sigma value after observing the angular
173 error of the chosen predictor at each trial (Figure 2A-iii;2B). This belief distribution allowed
174 us to derive two independent model-based estimates that we hypothesized to influence choice
175 in parallel: first, an estimate in the "accuracy" of a predictor (a point-estimate derived by the
176 mode of the belief distribution, representing the sigma believed to be the most likely of that
177 of the chosen predictor):

$$178 \quad \text{accuracy} = \max [\text{belief distribution}] * (-1) \quad (1)$$

179 Note that a higher accuracy value denoted in Eq.(1) indicates bigger deviations of the target
180 from the reference point. To derive an accuracy estimate that can be interpreted intuitively,
181 the sign of Eq.(1) is reversed so that positive values can be interpreted as higher accuracy.

183 Second, an estimate of the “uncertainty” in that predictor (variability around the accuracy
184 estimate, representing the uncertainty) (Figure 2A-iv):
185

$$186 \text{ uncertainty} = \widehat{\sigma}_{\text{cumulative belief distribution} = 97.5\%} - \widehat{\sigma}_{\text{cumulative belief distribution} = 2.5\%}. \quad (2)$$

188 The terms “accuracy” and “uncertainty” will from now onwards refer to the model-derived
189 parameters defined in Eq. (1) and (2), respectively (Figure 2A-iv). We used a Bayesian model
190 that assumed uniform prior beliefs for all four predictors at each block start. However, we
191 compared this Bayesian model to two competing models: a Bayesian model using
192 informative priors (Extended Data Figure 1) and a reinforcement learning (RL) model
193 tracking payoff history (Extended Data Figure 2). The Bayesian model with uniform priors
194 provided a better model fit to choice behaviour compared to either of the other models (see
195 Method; Supplementary Information: alternative computational models; Extended Data
196 Figures 1 and 2).

197
198 **<insert Figure 2 about here>**
199

200 We measured the degree to which participants were exploiting accurate predictors and the
201 degree to which they were exploring uncertain predictors. We hypothesized, first, that
202 uncertainty drives exploration between choices at the beginning of a block and so choices
203 might be directed to uncertain predictors. Then, over the course of a block, participants
204 should become increasingly uncertain avoiding in other words, choices should be directed
205 towards certain predictors (negative uncertainty effect) (Figure 2C-i). Second, we
206 hypothesized that the initial choice pattern in a block should depend on how many more trials
207 were still to be encountered in the block (effect of time horizon). Longer blocks favour more
208 uncertainty-driven exploration and less accuracy-driven exploitation compared to shorter
209 blocks (Figure 2C-ii).

210 To test the first hypothesis, we applied a logistic general linear model (GLM, see GLM1 in
211 Methods) to participants’ selections during the decision phase and then averaged beta weights
212 across participants (Figure 3A, Supplementary Figure 1). Regressors of interest (accuracy and
213 uncertainty) were coded as the difference between left and right predictors to predict leftward
214 selections. As would be expected if participants were attempting to maximize payoff,
215 participants generally sought out accurate predictors (main effect of accuracy: $t(23)=7.5$,
216 $p<0.001$, $d=1.53$, 95% confidence interval=[0.82 1.45]). There was no credible evidence that
217 uncertainty impacted choice behaviour ($t(23)=-1.9$, $p= 0.07$, $d=-0.39$, 95% confidence
218 interval=[-0.51 0.018], Bayes factor₁₀=1.05, %error=1.1017e-4). Next, to examine the time-
219 dependent effect of uncertainty and accuracy on selection, we included the percentage of
220 trials remaining in a block (referred to as ‘block time’) into the GLM model and examined its
221 interaction with accuracy and uncertainty. Participants alternated between behavioural modes
222 of exploration and exploitation by integrating information about the remaining trials into their
223 predictor selections: a positive interaction term between uncertainty and block time
224 ($t(23)=5.8$, $p<0.001$, $d=1.18$, 95% confidence interval=[0.53 1.1]) showed that uncertain
225 sources were explored when many trials remained. By contrast, a negative interaction term
226 between accuracy and block time indicated that, as time passed, choices were increasingly
227 directed towards accurate predictors (accuracy x block time interaction: $t(23)=7.5$, $p<0.001$,
228 $d=-1.53$, 95% confidence interval=[-0.91 -0.52]; Figure 3A).

229
230 **<insert Figure 3 about here>**
231

233 In a follow-up analysis, we further examined the interaction effects. We binned trials into
234 those that occurred in the first and second halves of each time horizon (Figure 3B-i). A
235 logistic GLM with accuracy and uncertainty as regressors was fitted to both halves of each
236 block's trials. Once again we found that decisions were influenced by both factors but in
237 dynamically distinct ways (paired t-test between the differences of block halves for accuracy
238 and uncertainty: $t(23) = -8.1$, $p < 0.001$, $d = -1.7$, 95% confidence interval = [-2.27 -1.02]; Figure
239 3B-i). Uncertain predictors were more likely to be sought out early compared to late in a
240 block (paired t-test early vs late: uncertainty ($t(23) = -8.1$, $p < 0.001$, $d = 1.66$, 95% confidence
241 interval = [1.06 1.8]): while during the first half there was only anecdotal support for the
242 interpretation that participants sought out uncertain predictors (positive uncertainty effect in
243 half 1: $t(23) = 2$, $p = 0.057$, $d = 0.41$, 95% confidence interval = [-0.007 0.48], Bayes
244 factor₁₀ = 1.18, %error = 9.954e-5), during the second half of blocks, uncertain predictors were
245 avoided (negative uncertainty effect in half 2: $t(23) = -6.2$, $p < 0.001$, $d = -1.27$, 95% confidence
246 interval = [-1.59 -0.79]). Accurate predictors were preferred to inaccurate ones and this was
247 increasingly the case in the second half of the blocks (paired t-test early vs late time points
248 accuracy: $t(23) = -4.2$, $p < 0.001$, $d = -0.85$, 95% confidence interval = [-1.63 -0.55]). These results
249 replicated when regressors were normalised across or within blocks.
250

251 In response to the reviewers' comments, we considered the possibility that such a result
252 might have arisen because the overall model fit was better for either the first or second half of
253 the block. It is important to consider differences in model fit across sets of trials (or
254 participants) because a poor model fit might indicate that the model is not appropriate for the
255 behaviour under investigation in one part of the data. However, *a priori* such an argument
256 would predict that an effect, such as uncertainty, would be stronger in the part of the data that
257 was better fit by the model than in the part worse fit by the model; it cannot predict a polarity
258 change in the uncertainty prediction effects when moving from exploration (earlier trials) to
259 exploitation (later trials). We excluded trials on the basis of the trial wise choice residuals so
260 that both first and second block halves were no longer different in their residual variance
261 (Extended Data Figure 3). Even under such conditions, we were able to replicate evidence for
262 the same pattern of results (Extended Data Figure 3D). Moreover, below we show that
263 several brain regions only represent uncertainty prediction difference during exploration and
264 not exploitation (Supplementary Figure 7, in particular 7B) even though model fits were
265 better for later compared to earlier phases.
266

267 Next, we tested our second hypothesis that the degree of exploration during initial choices
268 should be stronger in longer time horizons, i.e. if subsequent encounters with the same
269 predictor are expected to be more frequent. We compared choices during the first 15 trials
270 across all time horizons by fitting a linear robust GLM to data from each time horizon. The
271 first 15 trials in all three horizons were identical in their order presentation and importantly,
272 their trial-by-trial target estimates were drawn from a Gaussian distribution with the same
273 parameters (sigma of either 50 or 70). As predicted, participants adjusted their behavioural
274 strategy in the initial trials according to the horizon type: participants explored more in longer
275 than shorter horizons and in a complementary manner, shorter horizons led to a rapid
276 convergence onto accurate predictors (3x2 repeated measures ANOVA with horizon (long,
277 medium, short) and variable (accuracy, uncertainty); horizon x variable interaction:
278 $F(2,46) = 36.7$, $p < 0.001$, $\eta^2 = 0.61$, assumption of sphericity is met with Mauchly's test:
279 $\chi^2(2) = 0.28$, $p = 0.87$; Figure 3B-ii). Uncertain predictors were particularly sought out during
280 initial trials within long and medium time horizons (long horizon: $t(23) = 4$, $p < 0.001$,
281 $d = 0.8$, 95% confidence interval = [0.053 0.164]; medium horizon: $t(23) = 2.8$, $p = 0.009$,
282 $d = 0.56$, 95% confidence interval = [0.02 0.13]).

283
284 So far we have shown that model-derived estimates of the accuracy and uncertainty
285 determined participants selections between predictors. Next, we examined whether
286 participants also relied on both of these estimates when making their subjective confidence
287 report during the second phase of each trial (the confidence phase in Figure 1A). Accuracy
288 reflects a point-estimate of the most likely angular error between target and the predictor's
289 estimate and should therefore have an impact on the interval size the participants use to
290 indicate their subjective confidence during the confidence phase. Indeed, participants
291 indicated higher confidence for predictors that were believed to be accurate ($t(23)=11.7$,
292 $p<0.001$, $d=2.4$, 95% confidence interval=[0.66 0.98]). The Bayesian model also suggests that
293 participants form a representation about other possible angular errors that might underlie a
294 predictor's distribution (i.e. the width of the belief distribution). If participants are very
295 uncertain in their point-estimate of the angular error (i.e. if the Bayesian belief distribution is
296 very wide), then they should report a larger interval size to guarantee that the target falls
297 within the interval. In tandem with above effect of accuracy, participants were less confident
298 and selected a larger interval size when they evaluated predictors they believed were
299 uncertain (uncertainty: $t(23)=-10.4$, $p<0.001$, $d=-2.12$, 95% confidence interval=[-1.1 -0.73];
300 Figure 3C).

301
302 In summary, accuracy, uncertainty, and a time modulation of both effects influenced
303 participants' predictor selections. Early selections were uncertainty-driven explorative
304 selections and occurred particularly when time horizons were longer. Later selections were of
305 exploitative selections, directed towards accurate and away from uncertain predictors. The
306 exploratory behaviour we identify cannot simply be the result of noise in the learning
307 process⁹; people are more exploratory when the future time horizon is longer even if learning
308 opportunities are identical. Moreover, we show that our model-derived estimates of
309 participants' beliefs about the accuracy of a predictor and uncertainty about those beliefs
310 correspond to features of their subjective confidence judgments.

311
312 *Polarity of uncertainty decision signal in vmPFC changes from exploration to exploitation*

313
314 Our behavioural analyses show that participants incorporated the uncertainty in their beliefs
315 when selecting between two predictors. We went on to examine the coding of uncertainty in
316 the brain during predictor selection (fMRI-GLM1, see Methods). Our variable of interest was
317 the difference in uncertainty (as captured by our model) between the chosen and unchosen
318 predictors, i.e. "uncertainty prediction difference". This is similar to studies of value-guided
319 decision-making, where the difference in value between the option chosen and the option
320 rejected is regressed against the BOLD signal. A value difference signal often prominently
321 implicates the vmPFC in decision making processes^{10-14,17}.

322
323 When testing for an uncertainty prediction difference signal across all trials, we found a
324 negative uncertainty prediction difference in vmPFC (whole brain cluster-corrected; Figure
325 4A-i, Supplementary Table 1). This neural effect was in line with the negative effect
326 uncertainty exerted on choice behaviour towards the end of a block when participants
327 avoided uncertain predictors or in other words, sought out certain predictors. In addition, we
328 also found an "accuracy prediction difference" in a similar anatomical location in vmPFC
329 (Figure 4A-ii, Supplementary Table 1). Again, this accords with participants' general
330 preference for selecting accurate predictors to help them find the target location. To
331 additionally show that both accuracy and uncertainty prediction differences were encoded in
332 a similar anatomical region, we derived a domain general prediction difference by first,

333 calculating the mean across both absolute contrasts “((chosen uncertainty – unchosen
334 uncertainty) + (chosen accuracy – unchosen accuracy))” and second, by deriving a
335 conjunction between both absolute contrasts (Supplementary Figure 3A, 3B, respectively,
336 and Supplementary Table 3). A domain general prediction difference peaked within vmPFC.
337 Accuracy and uncertainty prediction differences are independent variables sharing across all
338 trials, on average, 0.01% of their variance (0.137% and 0.09 % of their variance is shared
339 when exploration and exploitation trials are each considered separately; Figure 4D)
340 suggesting both variables have independent effects on activity but within the same part of
341 vmPFC (for more details on regressor correlations, see Supplementary Figures 1,2). These
342 findings underline the role of vmPFC in guiding predictor selection as a function of both the
343 differences in accuracy and uncertainty of the predictors.
344

345 Having identified vmPFC as representing a negative uncertainty prediction difference across
346 all trials, we then went on to test whether this signal was modulated by distinct behavioural
347 modes of exploration and exploitation. We have shown that uncertainty tended to drive
348 exploration of predictors at the beginnings of blocks; at that time, selections were directed to
349 uncertain predictors (i.e. there was a positive effect of uncertainty during the first 15 trials in
350 medium and long horizons, Figure 3B-ii). Then, over the course of the block, participants
351 became increasingly uncertainty avoiding shown by a negative effect of uncertainty on
352 choice behaviour. We refer to this pattern of change as an “uncertainty polarity change”. We
353 investigated whether there was a brain region with similar characteristics: transitioning from
354 encoding a positive to negative uncertainty-based prediction difference as participants
355 switched from exploration to exploitation (Figure 4B). To test this hypothesis, we made use
356 of the fact that our computational model allowed us to classify individual trials into
357 exploration or exploitation according to the selection made on each trial: an exploitative
358 selection was defined as one in which the more accurate and less uncertain predictor was
359 selected while a directed uncertainty-guided explorative selection was defined as the
360 opposite: a trial in which the more inaccurate and uncertain predictor was chosen (Extended
361 Data Figure 4). Importantly this is distinct to other types of decision that might initially
362 appear exploratory, because the less accurate predictor was chosen, but which may simply be
363 due to noise in the learning or decision process^{9,22}. On such trials, selection is not just of the
364 less accurate predictor but are also made with certainty (Supplementary Figure 4A).
365

366 **<insert Figure 4 about here>**
367

368 To test for a neural polarity change of uncertainty prediction difference, we extracted time
369 courses from an independent region of interest (ROI) associated with the accuracy prediction
370 difference effect across all trials. This ensured that we did not bias the analysis towards
371 finding an effect in an area that was previously associated with the uncertainty prediction
372 difference. First, we used a time course analysis to extract both components of the
373 uncertainty prediction difference signal (variance in activity related to the chosen predictor
374 and variance in activity related to the unchosen predictor) during exploration and
375 exploitation. Activation in vmPFC covaried with a decision signal that changed its polarity
376 depending on the current behavioural mode: during exploitation, vmPFC carried a decision
377 signal that reflected a negative uncertainty prediction difference (negatively encoding the
378 uncertainty of the chosen predictor as opposed to the unchosen predictor; Figure 4C-ii); in
379 exploration, when behaviour was guided by uncertainty, vmPFC activity carried a positive
380 uncertainty prediction difference (positively encoding the uncertainty of the chosen predictor
381 as opposed to the unchosen predictor; Figure 4C-i). Given that the same variable is reflected
382 in both increase and decrease in activity at different task stages suggests an important change

383 in the nature of the representation. In response to reviewers' comments, we verified the
384 robustness of these results when the precise criteria for drawing boundaries between
385 exploration/ exploitation categories were modified (Supplementary Figure 8). It might be
386 argued that the vmPFC activity pattern simply reflects absolute uncertainty differences
387 between the presented predictors irrespective of behavioural mode (exploration versus
388 exploitation). We repeated the analysis and included the absolute uncertainty prediction
389 difference as an additional regressor. Nevertheless, we replicated the uncertainty polarity
390 change across modes in vmPFC (Supplementary Figure 5).

391

392 The trials we define as uncertainty-guided exploration trials are comparable to trials that have
393 previously been described as directed explorative choices⁷. They are, however, hypothesized
394 to be distinct to apparently random choice selections that may result simply from noise in the
395 decision process²² or the learning process⁹. In the current experiment, random exploration
396 trials were defined as ones on which participants selected predictors that they believed to be
397 inaccurate with certainty (i.e. negative uncertainty) (Supplementary Figure 4A). While it is
398 not possible to be sure that all uncertainty-guided exploration and all noise-based exploration
399 trials are classified correctly, on average the classification should capture a potential
400 difference in exploration type that may be associated with different neural mechanisms. To
401 test this possibility we therefore, in addition examined vmPFC activity on random
402 exploration trials. We extracted a time course from vmPFC associated with the previous
403 cluster of accuracy prediction difference and tested for an uncertainty prediction difference
404 during random exploratory trials. We tested beta weights extracted from the time course with
405 a leave-one-out procedure and found that unlike on uncertainty-guided exploratory trials,
406 there was no credible evidence that vmPFC represented uncertainty prediction difference
407 during these random exploratory selections (Supplementary Figure 4B).

408

409 **<insert Figure 5 about here>**

410

411 We have shown that behavioural modes were associated with different polarities of
412 uncertainty representation in vmPFC. Next, we were interested in whether the different
413 behavioural modes were associated with any distinct neural networks. We performed a
414 whole-brain GLM of exploration and exploitation trials and focused again on the uncertainty
415 prediction difference during the decision phase (fMRI-GLM2). During exploitation, we
416 observed activity centred on vmPFC related to a negative uncertainty prediction difference
417 (Figure 5A; Supplementary Table 2), confirming our previous findings. During exploration, a
418 positive uncertainty prediction difference signal was represented in vmPFC, but also across
419 an extensive network of brain regions, including dorsomedial frontal areas (Figure 5B). A
420 direct contrast of activation patterns in exploration and exploitation trials confirmed these
421 differences between behavioural modes (compare panels A, B, and C of Figure 5). Dorsal
422 ACC (dACC) in particular has been associated with exploratory²² and foraging behaviour²³.
423 We show that dACC represents uncertainty prediction differences during directed exploration
424 (Figure 5B, Supplementary Figures 6, 7A-iii), but there was no credible evidence for such a
425 representation during random exploration (Supplementary Figure 4B) or, unlike vmPFC,
426 exploitation (Supplementary Figure 7B-iii). We also observed an uncertainty prediction
427 difference in frontopolar cortex and dorsolateral prefrontal cortex (dlPFC), replicating results
428 of previous exploration studies^{24,25} (Supplementary Figure 7A). However, like dACC and
429 other dorsomedial frontal areas, both dlPFC and frontopolar cortex have distinct profiles
430 compared to vmPFC, as there was no credible evidence for a representation of uncertainty
431 prediction difference during exploitation and hence unlike vmPFC did not show an
432 uncertainty polarity change across behavioural modes (Supplementary Figure 7B).

433
434 In summary, we have shown a polarity change in the influence that uncertainty in one's belief
435 exerts not just on behaviour but also on vmPFC activity. During exploitative modes, when
436 differences in predictor certainty are the key decision variable, vmPFC reflects negative
437 uncertainty prediction difference, but when participants are in an explorative mode, vmPFC
438 activity reflects positive uncertainty prediction differences. During exploration, vmPFC is co-
439 active with an extensive network of regions carrying a similar uncertainty-related signal.
440

441 *Uncertainty-related signals in subcortical structures during exploration and exploitation*
442

443 We used a region-of-interest approach to test for an uncertainty prediction difference in
444 subcortical structures during both behavioural modes. We focused on amygdala and ventral
445 striatum as they have been previously associated with modes of exploration and
446 exploitation²⁶. We also focused on ventral tegmental area (VTA) which exhibited cluster-
447 corrected positive and negative uncertainty prediction difference during exploration and
448 exploitation respectively (Figure 5). All three subcortical regions represented uncertainty
449 prediction difference during at least one behavioural mode – either exploration or
450 exploitation – but with a different pattern of activation in each case: amygdala predominantly
451 represented uncertainty prediction difference during exploration (Extended Data Figure 5A),
452 while VS (Extended Data Figure 5B) activation was most apparent during exploitative phases
453 when it reflected a negative uncertainty prediction difference. VTA activity suggested a
454 representation of uncertainty prediction difference during both, exploration and exploitation
455 in the decision phase (Extended Data Figure 5C). These patterns show that a network of areas
456 including multiple cortical and subcortical areas represent uncertainty-related information
457 during both exploration and exploitation. While it was not identical, the pattern in the VTA
458 most closely resembled that seen in the vmPFC; it carried uncertainty signals that reversed in
459 polarity between exploration and exploitation but there was no credible evidence for an
460 accuracy-related signal during the transition phase between exploration and exploitation (see
461 paragraph on transition between exploration and exploitation; $t(23) = -0.97$, $p=0.35$, $d=-$
462 0.197 , 95% confidence interval=[-0.07 0.026], Bayes factor₁₀=0.325, %error=0.037). These
463 analyses were conducted in response to the reviewers' comments.

464
465 *Uncertainty representation in vmPFC scales with predictor repetition*
466

467 We have shown a polarity change in the effect of uncertainty on guiding behaviour and
468 influencing vmPFC activity when comparing exploratory and exploitative behavioural
469 modes. One possible way to interpret the negative uncertainty representation during
470 exploitation is that vmPFC encodes a default choice^{21,23,27}. In the context of the current task,
471 an effective default choice is repetition of previously made choices particularly when there
472 has been certainty about the predictor's accuracy. We therefore asked whether there was
473 evidence of a higher frequency of choice repetition on exploitation as opposed to exploration
474 trials; this was indeed the case (paired t-test explore vs exploit: $t(23)=-16.2$, $p <0.001$, $d= -$
475 3.3 , 95% confidence interval = [-0.36 -0.28]; Figure 6A). Moreover, activity in the same
476 location in vmPFC reflected whether, on each trial, participants would repeat a choice they
477 had made the last time it was offered. There was more activity in vmPFC when participants
478 were repeating a choice made previously (repetition: $t(23) = 4$, $p <0.001$, $d= 0.8$, 95%
479 confidence interval=[0.017 0.06]; Figure 6B, grey time course). In addition, the effect was

480 greater when there was a stronger negative uncertainty signal (repetition x chosen
481 uncertainty: $t(23) = -3.4216$, $p = 0.002$, $d = -0.7, 95\% \text{ confidence interval} = [-0.07 -0.02]$, Figure
482 6B, red time course): in other words, the repetition signal was greater when there was more
483 certainty about the selected predictor during repetitive trials compared to non-repetitive trials
484 during which they switched to a new choice that had not been made on a previous trial, a
485 behaviour more likely to occur during exploration (Figure 6A), then vmPFC had the opposite
486 polarity (positively related to uncertainty; Figure 6B, right panel).

487

488 **<insert Figure 6 about here>**
489

490 *The transition from positive to negative uncertainty representations is accompanied by the
491 processing of accuracy between predictors*
492

493 So far we have shown that the transition from exploration to exploitation and choice
494 repetition behaviour is accompanied by a change in the polarity of uncertainty signals and
495 emergence of choice repetition signals in vmPFC. However, it remains unclear how the
496 transition between directing behaviour towards uncertain and then certain predictors occurs
497 as the behavioural mode shifts from exploration to exploitation. It is possible that, after initial
498 exploration but before repetitively choosing certain predictors there might be a phase in
499 which participants focus on how well – how accurately – each predictor estimates the target’s
500 location (Figure 7A, see illustration). Such a period might naturally precede a period when
501 the most accurate predictors are identified and continuously chosen. During such a transition
502 period, one would expect neural activity correlating with an accuracy prediction difference,
503 the difference between the accuracy estimates associated with the chosen and unchosen
504 predictors. Moreover, because participants are transitioning from positive to negative
505 uncertainty-guided behaviour, the accuracy estimates held by participants for the chosen and
506 unchosen predictors should be close in value. This would suggest that participants have no
507 strict preference between predictors yet, as they are still learning about them. We identified a
508 new subset of trials by selecting trials with accuracy prediction differences close in value
509 (Supplementary Figure 9A). We hypothesized that vmPFC computes decision variables that
510 are most relevant for guiding choice behaviour in the current context, therefore when the
511 accuracy difference is small in value, participants need to carefully compare accuracy
512 estimates between predictors to make their choice. First, we tested whether these trials
513 occurred in time between the exploration and exploitation periods that we had previously
514 identified. Indeed, these transition trials occurred later in time compared to previously
515 defined explorative choices (paired t-test, explore vs. transition: $t(23)=6$, $p<0.001$, $d=1.2$,
516 $95\% \text{ confidence interval} = [0.056 0.12]$) and earlier in time compared to exploitative choices
517 (paired t-test, exploit vs. transition: $t(23)=-2.8$, $p=0.01$, $d=-0.57$, $95\% \text{ confidence interval} = [-$
518 $0.04 -0.006]$) (Figure 7A).

519

520 **<insert Figure 7 about here>**
521

522 We then examined whether vmPFC activity reflected the accuracy prediction difference
523 during this transitional period. To test for this effect, we chose an independent ROI in vmPFC
524 extracted from the cluster-corrected uncertainty prediction difference effect across all trials
525 (Figure 4A). As predicted, activation in vmPFC correlated with an accuracy prediction
526 difference during this transitional phase ($t(23) = 3.5$, $p= 0.002$, $d=0.71, 95\% \text{ confidence}$
527 $95\% \text{ confidence interval} = [0.03 0.1]$; Figure 7B). In further support of the suggestion that accuracy processing
528 is especially prominent during this transition phase (in which chosen and unchosen predictors
529 have similar accuracy values), we found no credible evidence of an accuracy prediction

530 difference signal in vmPFC when very inaccurate predictors (accuracy prediction difference:
531 $t(23) = -0.84$, $p = 0.41$, $d = -0.17$, 95% confidence interval=[-0.13 0.055], Bayes
532 factor₁₀=0.296,%error=0.037; Supplementary Figure 9B-i) or very accurate predictors were
533 selected (accuracy prediction difference: $t(23) = -1.3$, $p = 0.21$, $d = -0.27$, 95% confidence
534 interval=[-0.06 0.02], Bayes factor₁₀=0.447,%error=1.178e-4; Supplementary Figure 9B-ii).
535 This pattern of results suggests that the periods in which vmPFC activity reflects first positive
536 and then negative uncertainty prediction difference are separated by a transition period in
537 which vmPFC reflects the accuracy estimate of the predictor chosen to guide behaviour.
538

539 We tested whether activation during the transition phase was related to behavioural changes
540 across time – from positive to negative uncertainty-driven behaviour – when selecting
541 between predictors. As the transition phase bridges exploration (positive uncertainty) to
542 exploitation (negative uncertainty), we tested whether accuracy-related vmPFC activation
543 during the transition period was related to a behavioural effect of uncertainty that changes
544 across time, i.e. the interaction term uncertainty x block time (see behavioural choice GLM,
545 Figure 3A). We used a partial correlation analysis to examine the relationship between each
546 individual's accuracy-related vmPFC activity extracted from the vmPFC cluster (accuracy
547 prediction difference effect across all trials) and the behavioural transition from positive to
548 negative uncertainty-driven predictor selection. In the same analysis, we controlled for all
549 other behavioural variables included in the previous GLM1 (Figure 3A). We found that
550 accuracy prediction difference-related activity in vmPFC during the transition period was
551 positively correlated with uncertainty x block time ($r = 0.58$, $p = 0.007$, 95% confidence
552 interval= [0.23 0.8]; Figure 7C-i). That is, the larger the vmPFC signature encoding accuracy
553 prediction difference during the transition period, the stronger the behavioural transition from
554 positive to negative uncertainty-driven behaviour over the course of a block (Figure 7C-ii).
555 Notably, these results were not confounded by variation across participants' in the number of
556 transition trials that were identified; a partial correlation that controlled additionally for the
557 number of transition trials remained significant ($r=0.57$, $p=0.01$, 95% confidence interval=
558 [0.22 0.79]).
559

560 This result suggests that a transition phase during which the accuracy between predictors is
561 represented in vmPFC may facilitate a neural polarity change from first representing positive
562 uncertainty when selections are exploratory to later, representing negative uncertainty when
563 repeatedly selecting the same certain predictor during exploitation (Figure 8). Participants
564 exhibiting stronger predictor accuracy signals in vmPFC during the transition period
565 exhibited a more drastic change from positive to negative uncertainty-driven behaviour.
566

567 <insert Figure 8 about here>
568
569

570 **Discussion**

571

572 Humans select between multiple information sources that can predict outcomes in an
573 adaptive manner that enables them efficiently both to gather information about the predictors
574 and to use that information to make choices. Using Bayesian modelling, we derived estimates
575 of two kinds of beliefs that simultaneously influenced choice and neural activity. To select
576 between predictors, participants integrated beliefs about how accurately a predictor predicted
577 the target ("accuracy") and beliefs about the uncertainty in that estimate ("uncertainty"). How
578 much these beliefs influenced predictor selection depended on how many opportunities
579 participants had had to learn about the predictors already⁷. Behaviourally, participants

580 initially gathered information about available predictors by selecting more uncertain
581 predictors, while over time they converged towards accurate and certain (i.e. the negative
582 uncertainty effect) predictors. However, importantly the influence of accuracy beliefs and
583 their uncertainty depended on the future time horizon; participants explored uncertain
584 predictors more during initial phases of a block when they knew that they had a longer time
585 horizon remaining to exploit the knowledge gained. Behaviour that initially appears
586 uncertainty-directed and exploratory in nature may simply reflect noise in the decision
587 process²² or the learning process⁹ but in the present study behaviour is uncertainty seeking
588 and exploratory in nature because it manifests to a greater degree when the future time
589 horizon is longer even when the decision context and past learning opportunities remain the
590 same.

591
592 Similar flexibility was also observed on a neural level. VmPFC activity reflected different
593 decision variables at different times in a manner that reflected their relevance for the current
594 context of exploration or exploitation. Behaviour and neural activity in vmPFC were not
595 determined by only exploration or exploitation, but rather it reflected several different
596 variables but only when they were relevant to the current mode.

597
598 Our findings are related to studies of attention during the learning of cue-outcome
599 relationships. Here, two influential models have made opposite predictions: one model
600 suggests that selective attention is driven by cues that are most predictive of reward^{2,5},
601 reminiscent of the accuracy-driven, repetition-driven, and certainty-driven predictor
602 selections in the present study. The second model assumes that the uncertainty of a predictor
603 is crucial for selective attention⁶. By using a Bayesian model to dissociate participants'
604 beliefs about accuracy and uncertainty, we were able to show that in fact, both are important
605 to determine whether a predictor will be selected to guide behaviour. Importantly, the
606 magnitude of their influence on predictor selection depends on their relevance to the current
607 context which varies across time.

608
609 In accordance with the behavioural results, we found that neural activity reflected predictor
610 differences. Activity in vmPFC reflected the difference between the selected and rejected
611 predictor, in terms of the key feature that was currently of relevance for guiding behaviour:
612 first positive uncertainty, then accuracy, and then negative uncertainty. Previous studies have
613 often focused on the manner in which activation in vmPFC is correlated with differences in
614 the reward values of chosen and rejected choices^{11,28,29}. In such studies, differences in the
615 reward values associated with the choices constitute the evidence in favour of taking one
616 choice rather than the other. Although we focus on vmPFC's role in representing
617 information-based belief estimates of accuracy and uncertainty, on sub-threshold vmPFC also
618 represented the difference in expected value between predictors (Extended Data Figure 2).
619 Here, we show when selecting between predictors to guide behaviour, multiple types of
620 information, rather than just a single one, can be of importance. This can be linked to the idea
621 that vmPFC integrates a diverse set of variables that are currently choice-relevant³⁰ and to
622 recent evidence that exploitation and exploration are not simply behaviours that are
623 controlled by completely separate neural circuits but rather they are, at least in part,
624 controlled by changes in mode within neural structures²⁶. An alternative interpretation could
625 be that vmPFC's signal represents variables that are relevant for long-term reward
626 expectation: early uncertainty-driven exploration is beneficial for reward maximization
627 during later exploitative phases. Although we do not differentiate between immediate or
628 long-term representations, other studies have shown that dACC in particular represents value
629 expectations modulated by different timescales^{8,31-33}.

630
631 Our results also suggest that vmPFC does not guide behaviour in isolation, but that there are
632 additional broader differences in the recruitment of choice-relevant brain networks between
633 exploration and exploitation. Although activation associated with negative uncertainty
634 prediction difference during exploitation was mainly present in vmPFC, positive uncertainty
635 prediction difference during exploration was associated with a wider network including areas
636 such as dACC, dlPFC, and frontopolar cortex that have previously been associated with
637 exploration^{24,25}. Activation in dACC has often been related to behavioural adaptation and the
638 search for better alternatives, for example during foraging^{8,22,23,33–40} and to the update of
639 internal models during environmental changes^{41–43}. Our results may therefore suggest that in
640 some cases during exploration wider updates in decision networks occur that encompass both
641 vmPFC, dACC and prefrontal areas in a similar fashion. Nevertheless, it is important to
642 remember that the pattern of activity in vmPFC, when considered across both behavioural
643 modes, is different from that seen in dACC, dlPFC, and frontopolar cortex where activity
644 only reflects uncertainty during exploration while the change in the polarity of positive to
645 negative uncertainty-related activation, between exploration and exploitation, only occurs in
646 vmPFC. Additionally, vmPFC did not carry a clear uncertainty signal during random
647 exploration as opposed to uncertainty-guided exploration. An important new finding is that
648 effective exploratory behaviour may simply emerge from noise in the learning process⁹ and
649 this may impact on activity in brain areas such as dACC that reflect choice value learning at
650 multiple time scales^{31,33,44}. However, the current findings suggest that an uncertainty signal is
651 also carried in these areas when it is relevant for behaviour.
652

653 A related line of research supports the notion that vmPFC not only represents the value
654 difference between choice options, but also a second-order representation, that is one's own
655 confidence in a choice^{13,45}. These results are compatible with our finding that both accuracy
656 and uncertainty are represented in vmPFC. However, we show in addition that the polarity of
657 the uncertainty representation (which is a second-order representation similar to confidence)
658 in vmPFC changes depending on the behavioural mode. This suggests that in some cases
659 second-order representations in vmPFC are themselves choice-guiding and highly context
660 sensitive. The change in signal in vmPFC from signalling positive to negative uncertainty
661 prediction differences, i.e. uncertainty polarity change in vmPFC, might be related to the
662 presence of a learning phase during which predictors' accuracies are compared. We identified
663 a transition phase between exploration/exploitation periods, when no clear preference had yet
664 been formed for predictors. At that point, we observed that vmPFC most prominently
665 reflected participants' accuracy estimates for the predictors. Notably, the accuracy effect in
666 vmPFC during the transition phase was related to the degree of change from positive to
667 negative uncertainty-driven behaviour exhibited by each participant: participants exhibiting
668 stronger accuracy-related vmPFC activation during the transition period also showed more
669 drastic behavioural changes.
670

671 Although predictor selections were accuracy-guided throughout the task, we did not observe
672 an accuracy prediction difference in vmPFC during the final exploitation stages of predictor
673 selection. This is similar to the way in which vmPFC activity related to value comparison
674 during choice selection has been shown to be stronger during earlier compared to later phases
675 of a task²⁸. A predictor accuracy representation was present in vmPFC during the transition
676 phase between exploration and exploitation when accuracy estimates between predictors
677 were close in value, meaning that a careful comparison between predictors was required to
678 guide predictor selections successfully. In comparison, during exploitative trials participants
679 established which predictors were accurate resulting in repeated selections of the same

680 predictors. At that point vmPFC activity reflected this repetitive mode of decision making
681 and it did so in a manner that interacted with the representation of certainty (i.e. negative
682 uncertainty) about the predictor.

683

684 *Summary*

685

686 In summary, the combination of computational modelling and fMRI made it possible to show
687 that beliefs concerning the accuracy of predictors and the uncertainty of those beliefs inform
688 predictor selection to guide behaviour. Their influence on both behaviour and activity in
689 vmPFC changed and transitioned in tandem. The vmPFC carried information about a
690 multiplicity of decision variables (uncertainty, accuracy and repetition), the strength and
691 polarity of which varied according to their relevance for the current context.

692

693

694

695 **Methods**

696

697 **Participants**

698

699 Thirty participants took part in the experiment. Participants were excluded because they fell
700 asleep repeatedly during the scan (N=2), exhibited excessive motion during the scan (N=1),
701 or because of premature termination of an experimental session (N=3) (final sample: 24
702 participants; 14 female, age range: 19-35, mean age: 25.6, standard deviation: 4). No statistical
703 methods were used to pre-determine sample sizes but our sample sizes are larger to those
704 reported in previous publications^{31,33}. Moreover, participants took part in two versions of the
705 task which were averaged within participant and thereby statistical power was increased. The
706 study was approved by the Central Research Ethics Committee (MSD-IDREC-C1-2013-13)
707 at the University of Oxford. All participants gave informed consent.

708

709 **Experimental Procedure**

710

711 Participants took part in two magnetic resonance imaging (MRI) sessions on separate days
712 (Supplementary Information, details on task versions). We collapsed participants data across
713 two versions of the task (social/non-social) as the presented results did not show differences
714 between versions. The order of task version was counterbalanced across participants. Stimuli
715 used in each version were randomized across participants. Data collection and analysis were
716 not performed blind to the conditions of the experiments. Each session lasted approximately
717 two hours, including one hour of scanning. Participants received £15 per hour and a bonus
718 based on task performance (per session: £5 - £7).

719

720 Before each scanning session, participants were instructed about the task and performed
721 seven practice trials outside the scanner. After completion of both sessions, participants filled
722 in a questionnaire that assessed their understanding of the study.

723

724 **Experimental design**

725

726 On every trial, participants made decisions to maximise rewards over the course of the
727 experiment. The experiment was subdivided into six blocks. Each block included four new
728 predictors associated with four new stimuli. Although each predictor was unique, every block
729 comprised two good and two bad predictors. After selecting between a pair of predictors, the
730 chosen predictors suggested the location of a target. The true target location varied from trial
731 to trial and could not be predicted directly. The only way to estimate the target location was
732 to learn about the distance, in terms of the angular error, between true target location and the
733 predictor-suggested target location. The goal was to identify and exploit the good predictors
734 in each block. On every trial, at the first stage, participants made a binary choice between the
735 two presented predictors pseudo-randomly drawn from the four-predictor set (decision
736 phase). Choosing better predictors at this first stage of each trial led potentially to more
737 rewards through a decision that was made in the second stage of each trial (confidence
738 phase). The predictors' estimates varied around a true target location according to a normal
739 distribution with a given standard deviation. Better options were characterised by a smaller
740 standard deviation of the normal distribution. At the second stage, participants expressed
741 their confidence by changing the size of an interval (symmetric interval around the predicted
742 target location) and were rewarded if the target fell within the selected area (Figure 1A). The
743 payoff scheme was such that participants earned most if they indicated a small angular error
744 and the target appeared within the selected area in the subsequent outcome phase. Therefore,

745 choosing a better predictor in the decision phase allowed participants to earn more rewards in
746 the long run.

747

748 Overall, each MRI session comprised 180 trials, subdivided into 6 blocks, and lasted
749 approximately one hour. The length of a block (time horizon) was either short (15 trials),
750 medium (30 trials), or long (45 trials) (Figure 1B-i). Each time horizon was presented twice
751 and their order was pseudo-randomised with the constraint that blocks of the same horizon
752 did not succeed each other directly. Note that there was only one temporal order of predictor
753 presentation: the order for short and medium horizons were extracted from the long horizon
754 such that the first 15 trials were identical across horizons. The order of predictors was
755 carefully constructed such that variables of interest, model-derived estimates of accuracy and
756 uncertainty, were decorrelated statistically and across time. As shown in the Figure 4D, the
757 critical correlations between accuracy and uncertainty prediction differences are $r = 0.1$ (95%
758 confidence interval=[-0.32 0.48]) across all trials, $r = 0.37$ (95% confidence interval=[-0.04
759 0.67]) within exploration and $r = 0.30$ (95% confidence interval=[-0.12 0.63]) within
760 exploitation, on average across participants. This means that the maximum shared variance in
761 these conditions is 0.14 (in exploration). For more information on how experimental design
762 features helped to further decorrelate accuracy and uncertainty estimates across time, see
763 Supplementary Figure 2. In response to the reviewers' comments, we simulated a scenario
764 during which accuracy and uncertainty are correlated across time and show that this scenario
765 does not exist in the current study because of multiple precautions that were taken when
766 designing the experiment. One of the main precautions was the order of predictors across
767 time. We created the sequence of predictors in each block such that all possible binary
768 combinations of high/low uncertainty and high/low accuracy predictors were likely to occur
769 irrespective of the particular choice pattern of the participant. To achieve this, we introduced
770 two of the four predictors at slightly later times in each block, making them more uncertain
771 compared to the earlier presented predictors. We determined the precise order of predictors in
772 behavioural pilot experiments.

773

774 How good a predictor was, was determined by how well it estimated the target in the
775 confidence phase. Estimations followed a Gaussian distribution centred on the true target
776 location (Figure 1B-ii). Values, x , for each predictor were drawn from a Gaussian
777 distribution and represented the difference between the true target location and the predictor's
778 estimate:

$$x \sim N(\mu, \sigma) \text{ with } -180 < x < 180 \quad (1)$$

779 where at a given trial, value x was derived from a normal distribution with mean of $\mu = 0$ and
780 sigma of either $\sigma = 50$ for good predictors or $\sigma = 70$ for bad predictors. Note that sigma
781 determined the distance (i.e. the angular error) between the true target location and the target
782 position indicated by the predictor. Averaging across all observations of the angular error
783 allowed participants to estimate the sigma associated with each predictor (see Figure 2A for
784 detailed mapping between task space and belief estimates). As participants learned about the
785 predictor's performance through observing the angular error, they learned about the sigma of
786 each predictor's distribution.

787

788 Participants maximized their points by decreasing the interval size during the confidence
789 phase (Figure 1A). Participants changed the interval size with a precision of up to 20 steps on
790 each side of the reference location, that is a maximum of 40 steps as the interval was set
791 symmetrically. A step size was derived by dividing the circle size (6.3 radians) by the
792

794 maximum number of possible steps, resulting in a step size of approximately 0.16 radians.
795 The interval size was determined like follows:
796

$$797 \text{Interval size} = (\text{number of steps} \times 2) / 40 \quad (2)$$

799 When the target fell within the interval set by the participant, the magnitude of the payoff was
800 determined by subtracting the interval size from 1. However, if the target fell outside the
801 confidence interval, it resulted into a null payoff. This meant that the payoff per trial ranged
802 between 0 and 1.

803

$$804 \text{Payoff} = \begin{cases} (1 - \text{interval size}) & \text{if target is included} \\ 0 & \text{if target is excluded} \end{cases} \quad (3)$$

805 Trial structure

806

807 Each trial included a decision, confidence, and outcome phase (Figure 1A). Trials started
808 with the presentation of two options, a time bar, and question mark (1.5 sec on screen,
809 decision phase). The time bar indicated the amount of trials left in the current block; it
810 decreased after each trial until the end of a block. At the start of a new block, the type of
811 horizon was identifiable by inspecting the time bar. After the question mark disappeared,
812 participants chose between two predictors to receive information about the location of the
813 target on the circle. The chosen predictor was marked with a red box (0.5 sec). In the
814 confidence phase, the chosen predictor was shown in the centre of a circle and an interval
815 was depicted around a reference point (i.e., predictor's suggested target location) which was
816 indicated by a dot. The interval covered a portion of the circle symmetrically around the
817 reference point. The interval size was randomly initiated on each trial between a minimum of
818 one and a maximum of 20 steps (one step corresponds to one button press) away from the
819 predictor's estimated target location. After participants made a choice how to set the interval
820 size, a black frame appeared around the chosen predictor to indicate their response (0.5 sec).
821 The duration of the confidence phase was determined by the participant's reaction time.
822 Finally, a second marker appeared on the circle representing the true target location and the
823 number of points (between zero and one) below the predictor (3 sec, outcome phase).

824

825 To decorrelate variables of interest between trial phases, short intervals were included
826 between trials (inter-trial-intervals) and randomly, but equally allocated to either the
827 transition between decision- and confidence phase or confidence- and outcome phase. The
828 duration of an interval was drawn from a Poisson distribution with the range of 4s to 10s and
829 a mean of 4.5s. During these intervals, a fixation cross was shown on the screen.

830

831 Bayesian Model

832

833 We used a Bayesian model to estimate the beliefs participants might optimally hold about the
834 sigma (σ) characterising the normal distribution of each predictor. Sigma (σ) refers to the
835 standard deviation of the normal distribution from which observations of the angular error
836 were drawn, i.e. distance between target and reference location at each trial. Participants learn
837 about how well a predictor predicts the target location across time and by doing so, they
838 implicitly estimate the sigma value (σ) of the distribution (see Figure 2 for detailed mapping
839 between task parameters and subjective estimates). Using a Bayesian model, we derived
840 subjects' beliefs about the sigma value (σ) of each predictor's distribution, resulting in
841

842

843 sigma-hat ($\hat{\sigma}$) that denotes participants' estimated sigma. Before a belief can be formed,
844 participants selected a predictor and then made an observation x of how good the predictor
845 was on a given trial, defined by the angular error between the true target location and the
846 predictor-estimated location (reference location):
847

$$848 \quad x \text{ (angular error)} = \text{reference location} - \text{true target location}, \quad (4)$$

849 where the reference location indicated the predictor's prediction of the target location. Key
850 features of beliefs can be captured by a probability density function (pdf) over sigma (Figure
851 2A-iii,iv; 2B). The parameter space comprised possible sigma values that could be estimated
852 by the participant. The parameter space of sigma was bound between 1 and 140 degrees to
853 allow a broad range of sigma values considering the circle shape.
854

855 Following Bayes' rule, a belief is updated by multiplication of a prior belief and a likelihood
856 distribution resulting in a posterior belief, i.e. belief update (Figure 2B). Before the very first
857 observation, participants' belief in sigma, $\hat{\sigma}$, was assumed to be uniformly distributed across
858 parameter space, i.e. possible sigma values in parameter space were predicted to occur with
859 equal probability:
860

$$861 \quad p(\hat{\sigma}) = U(1, 140). \quad (5)$$

862 A likelihood function was then calculated that described the probability of the observation x
863 given each possible sigma value:
864

$$865 \quad p(x | \hat{\sigma}) = N(x | \mu=0, \hat{\sigma}). \quad (6)$$

866 With Bayes rule, we derived a trial-by-trial posterior distribution that was proportional to the
867 multiplication of a prior distribution and likelihood:
868

$$869 \quad p(\hat{\sigma} | x) \propto p(x | \hat{\sigma}) p(\hat{\sigma}) \quad (7)$$

870 where,

- 871 a. $p(\hat{\sigma})$ is the prior distribution.
- 872 b. $p(x | \hat{\sigma})$ is the likelihood function.
- 873 c. $p(\hat{\sigma} | x)$, is the posterior pdf across parameter space. The posterior pdf is the updated
874 belief across sigma space and is used as prior for the next trial of the same predictor.

875 Each posterior was normalised to ensure that probabilities across all sigma values added up to
876 one:
877

$$878 \quad p(\hat{\sigma} | x) = \frac{p(\hat{\sigma} | x)}{\sum p(\hat{\sigma} | x)} \quad (8)$$

881 *Model parameters*

882 We used features of an option's prior distribution on every trial to approximate participants'
883 estimates of the accuracy of the predictor and their uncertainty in those accuracy estimates.
884 The mode (peak of distribution) of the prior pdf was used to define "accuracy", while a 95%
885 interval around the mode was used to define "uncertainty". Note that both variables depended
886 on choices made by participants, because feedback was only provided for the chosen
887

889 predictor and hence only beliefs for the chosen predictor could be updated. On trial i,
890 variables of interest were defined as follows (Figure 2A-iv):
891

$$892 \quad \text{accuracy} = \max [p(\hat{\sigma})] * (-1) \quad (10)$$

893

894 Note that a higher $\max[p(\hat{\sigma})]$ of the pdf indicated bigger deviations of the target from the
895 reference point. To derive an accuracy estimate that can be interpreted intuitively, the sign of
896 $\max[p(\hat{\sigma})]$ is reversed (multiplication with (-1)) so that positive values can be interpreted as
897 higher accuracy. The accuracy estimate represents a point-estimate of a subject's belief
898 distribution in sigma-hat ($\hat{\sigma}$). This means it represents the subject's belief in the sigma value
899 associated with the predictors' distribution.
900

901 To derive a trial-wise uncertainty estimate from the distribution, we identified a percentage
902 (2.5%) of the lower and upper tail of the prior pdf, representing the distribution around the
903 believed sigma value ($\hat{\sigma}$). We extracted the estimated sigma value $\hat{\sigma}_{\text{high}}$ and $\hat{\sigma}_{\text{low}}$ at each of the
904 two positions. The difference of both sigma values constituted the estimated "uncertainty"
905 variable:
906

$$907 \quad \hat{\sigma}_{\text{high}} \leftarrow \text{cumulative}(p(\hat{\sigma})) = 97.5\% \\ 908 \quad \hat{\sigma}_{\text{low}} \leftarrow \text{cumulative}(p(\hat{\sigma})) = 2.5\% \\ 909 \quad \text{uncertainty} = \hat{\sigma}_{\text{high}} - \hat{\sigma}_{\text{low}} \quad (11)$$

910

911 From now onwards, the terms of accuracy and uncertainty refer to the model-derived
912 estimates defined in equations (10) and (11) respectively.
913

914 Alternative computational models

915

916 We used a Bayesian model with uniform priors at the start of each block for all four
917 predictors, assuming participants do not have prior knowledge about the underlying
918 distributions associated with predictors. We refer to this model as 'the original model'
919 because it is the model used elsewhere in this study. We compared the original model to two
920 alternative computational models: a Bayesian model with informative priors (Extended Data
921 Figure 1) and a reinforcement learning (RL) model which tracks the payoff history of each
922 predictor (Extended Data Figure 2). We explain in detail the rationale behind each
923 computational model, their construction and the results in the Supplementary Information
924 (Section 2: Alternative computational models). The results demonstrate that a Bayesian
925 model using uniform priors had a better model fit compared to a Bayesian model with
926 informative priors or an RL model. However, a combination of the original Bayesian model
927 with uniform priors and value-based variables derived from an RL model showed the best
928 model fit to choice behaviour. In conclusion, RL value terms complement the Bayesian
929 model but do not substitute for the Bayesian model terms as an explanation of behaviour;
930 participants' beliefs in the accuracy and uncertainty of a predictor explained additional
931 variance in choice behaviour above and beyond that explained by their choice value
932 estimates. These analyses were conducted in response to the reviewers' comments.
933

934 Behavioural Analyses 935

936
937 We applied a set of general linear models (GLM) to the behavioural data. All GLM analyses
938 were applied to both versions (social and non-social) of the experiment separately. The
939 resulting beta weights for each subject were first averaged across versions and then across
940 participants. We used two-tailed statistical tests for all analyses. Additionally, we report
941 effect size as Cohen's d (d) for t-tests and eta squared (η^2) for ANOVAs, a 95% confidence
942 interval and Bayes factors for non-significant results.

943
944 *Decision Phase*
945

946 We analysed the trial-wise impact of Bayesian-derived estimates of accuracy, uncertainty,
947 and their modulations across time in a block on choice behaviour. Our first analysis aimed to
948 show that the belief in the accuracy of a predictor ("accuracy") and the uncertainty in that
949 belief ("uncertainty") influenced choice behaviour. Moreover, we focused on how these
950 effects changed with the percentage of remaining trials in a block (referred to as block time),
951 suggesting a transition between exploration and exploitation as time within a block pass. We
952 used a logistic general linear model (GLM) to investigate these effects across all trials on
953 choice behaviour (Choice GLM1). For all GLM analyses, regressors were normalised across
954 all trials (mean of 0 and standard deviation of 1). The first GLM comprised the following
955 regressors.

956
957 Choice GLM1 (Figure 3A)
958 accuracy difference (left – right),
959 uncertainty difference (left – right),
960 block time,
961 accuracy difference (left – right) x block time,
962 uncertainty difference (left – right) x block time.
963

964 The dependent variable was whether or not participants made a leftward choice on the current
965 trial. Accordingly, for each regressor (except block time), we calculated the difference in the
966 variable for the left and right option. To calculate the interaction term, we multiplied the
967 normalised uncertainty and accuracy variables with the normalised block time variable and
968 then normalised this interaction term again. Note that we use the accuracy and uncertainty
969 regressors as defined in the "Bayesian model" section.

970
971 To further examine how the influence of uncertainty and accuracy on choice changed over
972 time in a block, we binned trials within a given time horizon into first and second halves
973 (Figure 3B-i). We fitted a logistic GLM on each half with uncertainty and accuracy as
974 regressors, irrespective of the overall time horizon length of the block. Although we
975 normalise regressors here within blocks, results replicate when regressors are normalised
976 across blocks.

977
978 Time GLM 1 (Figure 3B-i):
979 accuracy difference (left – right)
980 uncertainty difference (left – right)

981 Next, we predicted an effect of time horizon (Figure 3B-ii) on the first trials of a block. We
982 fitted a robust linear GLM on the first 15 trials (a multiple of all horizons, which were 15, 30
983 and 45) with accuracy and uncertainty as regressors to investigate whether a variable's effect
984 covaried with the amount of remaining trials.

986
987 Time GLM 2, for the first 15 trials within horizons (Figure 3B-ii):
988 accuracy difference (left – right)
989 uncertainty difference (left – right)
990
991 We used a linear robust regression to better estimate effects given the small amount of trials
992 included in the analysis. The first 15 trials were identical across horizons in terms of their
993 predictor order and statistical properties (apart from the specific choice sequence taken by
994 participants). All significant results reported in Figure 3B-ii remained significant when
995 basing the statistical tests on the t-stats of the effect sizes obtained from a logistic regression
996 (reported interaction effect: 3x2 repeated measures ANOVA with horizon (long, medium,
997 short) and variable (accuracy, uncertainty); horizon x variable interaction: $F(2,46)=27.6$,
998 $p<0.001$, $\eta^2=0.965$, 95% confidence interval [0.052 1.13], assumption of sphericity is met
999 with Mauchly's test: $x^2(2)=0.26$, $p=0.88$; reported main effects: positive uncertainty during
1000 long horizon: $t(23)=4.7$, $p<0.001$, $d=0.96$, 95%confidence interval=[0.51 1.3]; medium
1001 horizon: $t(23)=2.6$, $p=0.017$, $d=0.5$,95%confidence interval=[0.1 1]).
1002
1003 *Confidence phase*
1004
1005 We analysed the effect of accuracy and uncertainty on confidence judgments reported at the
1006 second phase of a trial (Figure 1A). Confidence judgments were indicated by modifying the
1007 interval size around the chosen predictor with a smaller interval representing higher
1008 confidence. To make this measure intuitive, we sign-reversed their relationship such that a
1009 higher confidence index represents greater confidence in the chosen predictor. We analysed
1010 the trial-by-trial confidence judgements by applying the following linear GLM:
1011
1012 Confidence GLM1 (Figure 3C):
1013 chosen accuracy,
1014 chosen uncertainty.
1015
1016 *Exploration, exploitation and transitional trials*
1017
1018 We subdivided trials into exploration and exploitation trials to compare neural signals
1019 between both behavioural modes. For each subject, we categorized trials based on the
1020 predictor selections during the decision phase (Extended Data Figure 3). On each trial, we
1021 calculated the difference between chosen and unchosen accuracy and chosen and unchosen
1022 uncertainty. Exploitative trials were defined by a positive “accuracy prediction difference”
1023 (chosen predictor had higher accuracy than unchosen ones) and negative “uncertainty
1024 prediction differences” (the chosen predictor was the predictor participants were more certain
1025 about). Vice versa, exploration trials were defined by a negative accuracy prediction
1026 difference and positive uncertainty prediction differences (the more uncertain predictor is
1027 picked even though it has yielded less accurate results in the past). Trials with both positive
1028 accuracy prediction difference and uncertainty prediction difference (i.e. that were both
1029 accuracy and uncertainty guided) were allocated to either the exploitative or the exploratory
1030 bin depending on the relative predominance of the accuracy prediction difference or the
1031 uncertainty prediction difference. For example, if the chosen predictor and the unchosen
1032 predictor differed more in the uncertainty of their predictions as opposed to the accuracy of
1033 their predictions (the chosen predictor was more uncertain than the unchosen predictor and
1034 the chosen predictor was, to a smaller degree, more accurate in its predictions than the
1035 unchosen predictor) then the predictor selection on that trial was labelled as exploratory.

Finally, trials with differences between both accuracy and uncertainty close to zero (absolute difference of 5) were assigned to both categories. We elaborate on the robustness of the current classification and compare it to those used in previous studies in the Supplementary information (Supplementary Figure 8). Furthermore, we defined a new subset of trials to understand the transition from positive uncertainty prediction difference signals (exploration) to a negative uncertainty prediction difference signal (exploitation) in vmPFC. Because predictor selections are not driven by uncertainty alone, we tested whether accuracy prediction difference signals were particularly prominent in a transitional phase between exploration and exploitation in vmPFC. We defined a threshold in a range of accuracy prediction difference values between [5 20] that classified trials into the transition period. We chose this subset such that it would comprise trials that are close in accuracy values for both options and at the same time predictor selection would still be guided rationally by accuracy. Moreover, this window resulted in a sufficiently large sample for analysis (approximately 20% of the trials in the range of positive accuracy prediction difference). The threshold is arbitrary and slightly smaller or greater ranges (compromising positive values) did not alter the results. To show that the transition period was characterized by learning about predictors, and that periods outside this transition were defined by the processing of either positive uncertainty or negative uncertainty, we defined two separate subsets of trials (Supplementary Figure 9A). One subset included extreme positive accuracy-driven trials [accuracy values > 20] (Supplementary Figure 9A-ii), while a second subset contained extreme negative accuracy-driven [accuracy values < -5] trials (Supplementary Figure 9A-i).

FMRI data acquisition and data processing

Imaging data were acquired with a Siemens Prisma 3T MRI using a multiband T2*-weighted echo planar imaging sequence with acceleration factor of two and a 32-channel head-coil. Slices were acquired with an oblique angle of 30° to the PC-AC line to reduce signal dropout in frontal pole. Other acquisition parameters included 2.4x2.4x2.4 mm voxel size, TE = 20 ms, TR = 1030 ms, 60° flip angle, a 240 mm field of view and 60 slices per volume. For each session, a fieldmap (2.4x2.4x2.4mm) was acquired to reduce spatial distortions. Bias correction was applied directly to the scan. A structural scan was obtained with slice thickness = 1 mm; TR = 1900 ms, TE = 3.97 ms and 1x1x1 mm voxel size.

Imaging data was analysed using FMRIB's Software Library (FSL)⁴⁶. Preprocessing stages included motion correction, correction for spatial distortion by applying the fieldmap, brain extraction, high-pass filtering and spatial smoothing using full-width half maximum of 5mm. Images were co-registered to an individuals' high-resolution structural image and then non-linearly registered to the MNI template using 12 degrees of freedom⁴⁷.

FMRI Data analysis

MRI whole-brain analyses

We used FSL FEAT for first-level analysis⁴⁶. First, data was pre-whitened with FSL FILM to account for temporal autocorrelations. Temporal derivatives were included into the model. We used two GLMs to analyse fMRI data across the whole brain. FMRI-GLM1 was applied to all trials and FMRI-GLM2 was fitted separately to trials that had been identified as exploration and exploitation trials. Results were calculated using FSL's FLAME 1 with a

1084 cluster-correction threshold of $z>2.3$ and $p<0.05$, two-tailed. To analyse BOLD changes
1085 associated with the processing of uncertainty and accuracy across participants, a second-level
1086 analysis was applied in a two-step approach: data was first average across both versions
1087 within subject (fixed-effect analysis) and then sessions were analysed across participants
1088 (FLAME1). We included all three phases of a trial (decision, confidence and outcome) into
1089 the fMRI-GLM. Each phase included a constant regressor, which was the onset of each phase
1090 as well as parametric regressors that were modelled as stick functions (i.e. duration of zero)
1091 time-locked to the relevant phase onset.

1092
1093 The decision phase began at the time the predictor appeared and lasted until a selection was
1094 made by the participant (Figure 1A). The decision phase was modelled as a constant and was
1095 accompanied by the following parametric regressors:

1096
1097 fMRI-GLM 1, decision phase:

1098 chosen uncertainty,
1099 unchosen uncertainty,
1100 chosen accuracy,
1101 unchosen accuracy,

1102
1103 All regressors were normalised before inclusion into the analysis. We calculated the
1104 difference between chosen and unchosen predictors for both accuracy and uncertainty to
1105 derive prediction differences. To derive a “domain general prediction difference”, we
1106 calculated the mean across absolute uncertainty and accuracy prediction differences: $((\text{chosen} - \text{unchosen uncertainty}) + (\text{chosen} - \text{unchosen accuracy}))$ (Supplementary Figure 3A) and
1107 calculated a conjunction between both cluster-corrected maps of accuracy and uncertainty
1108 prediction differences with a cluster-correction of $z>2.3$ and $p<0.05$ (Supplementary Figure
1109 3B). For the conjunction analysis, we used the provided FSL script ‘easythresh_conj’ with
1110 $z>2.3$ and $p<0.05$.
1111

1112
1113 The confidence phase was defined from the onset of circle and interval presentation (Figure
1114 1A) until a decision about the interval size was made. It included a constant and the following
1115 parametric regressors:
1116

1117 fMRI-GLM 1, confidence phase:

1118 chosen uncertainty,
1119 chosen accuracy,
1120 block time,
1121 chosen uncertainty x block time,
1122 chosen accuracy x block time.

1123
1124 All regressors were normalized before, and, where relevant, after building the interaction
1125 term (chosen accuracy/ uncertainty x block time). We only included the chosen predictor, as
1126 participants evaluated their uncertainty and accuracy estimates according to the predictor they
1127 selected during the decision phase.
1128

1129 The outcome phase was defined by the onset of the target and payoff presentation and lasted
1130 for a fixed duration of three seconds. In addition to the constant regressor, we included the
1131 following parametric regressors:
1132

1133 fMRI-GLM 1, outcome phase:

1134 chosen accuracy,
1135 chosen uncertainty,
1136 payoff (as defined in equation 3).
1137

1138 In the second fMRI-GLM2, trials were binned into exploratory and exploitative trials as
1139 described above. For this purpose, we included decision, confidence and outcome phases for
1140 exploratory and exploitative trials separately. This meant that, in total, there were six phases
1141 within the fMRI-GLM2. We included the same set of regressors in the exploratory and
1142 exploitative phases. The constants for each phase was modelled as in the previous GLM, but
1143 we used separate constants for exploration and exploitation phases.
1144

1145 fMRI-GLM 2, decision phase (for explore and exploit separately):
1146 uncertainty prediction difference (i.e., chosen – unchosen)
1147 accuracy prediction difference (i.e., chosen – unchosen)
1148 fMRI-GLM 2, confidence phase (for explore and exploit separately):
1149 chosen accuracy
1150 chosen uncertainty
1151 fMRI-GLM 2, outcome phase (for explore and exploit separately):
1152 chosen accuracy
1153 chosen uncertainty
1154 payoff.
1155

1156 To test whether the uncertainty prediction difference significantly differed between
1157 exploration and exploitation, we built a contrast comparing uncertainty prediction differences
1158 between exploration and exploitation (Figure 5C).
1159

1160 In addition, fMRI-GLM1 contained one regressor time-locked to all button presses, modelled
1161 as a stick function. For fMRI-GLM2, two regressors were time-locked to the button presses:
1162 one relating to the exploration phase and the other related to the exploitation phase.
1163

1164 *Region of Interest (ROI) analyses*

1165

1166 We calculated ROIs with a radius of three voxels that were centred on the peak voxel of
1167 significant clusters derived from whole brain fMRI-GLM1 and fMRI-GLM2. The selected
1168 ROI was transformed from MNI space to subject space and the pre-processed BOLD time
1169 courses were extracted for each participant's session. Time courses were averaged across
1170 volumes, then normalized and oversampled by a factor of 20 for visualisation. Time courses
1171 were time-locked to the onsets of each phase consistent with timings used in whole-brain
1172 fMRI-GLMs (decision, confidence or outcome). Then, a GLM was applied to each timepoint
1173 to derive beta weights per time point for each regressor. For analyses across versions, we
1174 used the same principle as applied to the whole-brain fMRI-GLMs and our behavioural
1175 analyses: first, we averaged the time course within a subject across both social and non-social
1176 versions, then we averaged across participants. For all ROI analyses, regressors were
1177 normalized (mean of zero and standard deviation of one).
1178

1179 To illustrate positive and negative uncertainty in exploration and exploitation phases,
1180 respectively, we included the following regressors:
1181

1182 ROI-GLM 1, decision phase (for explore and exploit separately), Figure 4C:
1183 chosen uncertainty,

1184 unchosen uncertainty,
1185 chosen accuracy,
1186 unchosen accuracy.
1187

1188 Effects of ROI-GLM1 were extracted from the whole-brain cluster corrected accuracy
1189 prediction difference effect in vmPFC to allow for an unbiased test.
1190

1191 Next, we tested whether the uncertainty effect changed when repeating the same predictor as
1192 on the last encounter. We used a ROI analysis to test for a main effect of repetition and
1193 interaction effect between repetition and chosen uncertainty. We used ROI-GLM1 and
1194 additionally included the following regressors:
1195

1196 ROI-GLM 2, decision phase (across all trials), Figure 6:
1197 additional regressors to ROI-GLM1:
1198 repetition (1= repetition of the same predictor as during last encounter with same predictor;
1199 0=no repetition of the same predictor)
1200 repetition x chosen uncertainty,
1201 repetition x chosen accuracy.
1202

1203 Then, we split trials into repetition and no-repetition categories to investigate the simple
1204 effect of chosen uncertainty per category (ROI-GLM3). We used ROI-GLM1, but now
1205 applied separately to repetition and no-repetition trials (Figure 6). For both ROI-GLM2 and
1206 3, we used an unbiased ROI extracted from the whole-brain cluster corrected accuracy
1207 prediction difference effect across all trials in vmPFC.
1208

1209 Next, we applied a ROI analysis to show activation for accuracy prediction difference during
1210 the transitional phase (Figure 7) in vmPFC, using fMRI-GLM2. We were interested whether
1211 the accuracy prediction difference effect occurred in the transition between the previously
1212 observed positive and then negative uncertainty prediction differences. Because we
1213 hypothesized that the accuracy prediction difference effect would occur in the same ROI as
1214 the uncertainty prediction difference effects, we used an independent ROI based on the
1215 cluster-corrected accuracy prediction difference effect across all trials (fMRI-GLM 1). The
1216 same ROI and GLM was used to test extreme positive and negative accuracy-driven trials
1217 (Supplementary Figure 9B).
1218

1219 ROI-GLM 4, decision phase (transition trials and extreme positive and negative accuracy
1220 trials), Figure 7B; Supplementary Figure 9B:
1221 see fMRI-GLM2.

1222

1223 *Leave-one-out procedure*

1224

1225 A leave-one-out procedure was used to test the unbiased significance of the time courses
1226 extracted from ROI-GLM2,3. For every participant ($n = 24$), we extracted the average time
1227 course based on the 23 remaining participants. We identified the peak of the group time
1228 course in a time window between 4-8 seconds and then extracted the beta value for the
1229 excluded subject at the time of the group peak. This procedure was repeated for all
1230 participants which resulted in individual peak values that were independent from the subject
1231 to be analysed. The extracted peak values were tested with a one-sample t-test against zero.
1232

1233 *Correlations between neural and behavioural beta weights*

1234
1235 To calculate the correlation between the time course of neural activations and behavioural
1236 beta values, we used neural beta weights extracted from the group peak. We calculated a
1237 partial correlation between the vmPFC accuracy prediction difference effect during the
1238 transition phase and the behavioural interaction term of uncertainty x block time (Figure 7C),
1239 controlling for all other behavioural variables (main effects of accuracy, uncertainty, block
1240 time (in percentage) and the interaction between block time and accuracy, see behavioural
1241 GLM1). A second partial correlation additionally included the number of individual
1242 transition trials.
1243
1244
1245

1246 Data availability

1247

1248 We have deposited all choice raw data used for the analyses in an OSF repository. The
1249 accession code is: https://osf.io/d5qzw/?view_only=037ea3b875914623a06999cef97ac57f.

1250 We have deposited unthresholded fMRI maps of all contrasts depicted in the manuscript on
1251 Neurovolt. The accession code is: <https://identifiers.org/neurovault.collection:8073>.

1252 The source data underlying Figure 3,6,7 and Extended Data Figure 1,2,3,5 are provided as a
1253 Source Data file.

1254

1255

1256 Code availability

1257

1258 The above OSF repository includes the full Bayesian modelling pipeline. Relevant
1259 behavioural and neural regressors were derived from this pipeline. We also provide the code
1260 for behavioural GLMs shown in Figure 3. Please follow the README file inside the
1261 repository for details of its use:

1262 https://osf.io/d5qzw/?view_only=037ea3b875914623a06999cef97ac57f.

1263

1264 References

1265

12661. Akaishi, R., Kolling, N., Brown, J. W. & Rushworth, M. Neural Mechanisms of Credit
1267 Assignment in a Multicue Environment. *J. Neurosci.* **36**, 1096–1112 (2016).
12682. Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V. & Niv, Y. Dynamic Interaction
1269 between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*
1270 **93**, 451–463 (2017).
12713. Garrett, N., González-Garzón, A. M., Foulkes, L., Levita, L. & Sharot, T. Updating Beliefs
1272 under Perceived Threat. *J. Neurosci.* **38**, 7901–7911 (2018).
12734. Charpentier, C. J., Bromberg-Martin, E. S. & Sharot, T. Valuation of knowledge and
1274 ignorance in mesolimbic reward circuitry. *Proc Natl Acad Sci USA* **115**, E7255–E7264
1275 (2018).
12765. Mackintosh, N. J. A theory of attention: Variations in the associability of stimuli with
1277 reinforcement. *Psychological Review* **82**, 276–298 (1975).
12786. Pearce, J. M. & Hall, G. A model for Pavlovian learning: Variations in the effectiveness of
1279 conditioned but not of unconditioned stimuli. *Psychological Review* **87**, 532–552 (1980).
12807. Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A. & Cohen, J. D. Humans Use Directed
1281 and Random Exploration to Solve the Explore–Exploit Dilemma. *J Exp Psychol Gen* **143**,
1282 2074–2081 (2014).
12838. Kolling, N., Scholl, J., Chekroud, A., Trier, H. A. & Rushworth, M. F. S. ProspectioN,
1284 Perseverance, and Insight in Sequential Behavior. *Neuron* **99**, 1069–1082.e7 (2018).
12859. Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S. & Wyart, V. Computational noise
1286 in reward-guided learning drives behavioral variability in volatile environments. *Nature* **59**.
128710. Basten, U., Biele, G., Heekeren, H. R. & Fiebach, C. J. How the brain integrates costs and
1288 benefits during decision making. *PNAS* **107**, 21767–21772 (2010).
128911. Boorman, E. D., Behrens, T. E. J., Woolrich, M. W. & Rushworth, M. F. S. How Green Is the
1290 Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative
1291 Courses of Action. *Neuron* **62**, 733–743 (2009).
129212. Chau, B. K. H., Kolling, N., Hunt, L. T., Walton, M. E. & Rushworth, M. F. S. A neural
1293 mechanism underlying failure of optimal choice with multiple alternatives. *Nature
1294 Neuroscience* **17**, 463 (2014).
129513. De Martino, B., Fleming, S. M., Garrett, N. & Dolan, R. J. Confidence in value-based choice.
1296 *Nature Neuroscience* **16**, 105 (2012).
129714. FitzGerald, T. H. B., Seymour, B. & Dolan, R. J. The Role of Human Orbitofrontal Cortex in
1298 Value Comparison for Incommensurable Objects. *J. Neurosci.* **29**, 8388–8395 (2009).
129915. Fouragnan, E. F. *et al.* The macaque anterior cingulate cortex translates counterfactual choice
1300 value into actual behavioral change. *Nature Neuroscience* **22**, 797–808 (2019).

130116. Papageorgiou, G. K. *et al.* Inverted activity patterns in ventromedial prefrontal cortex during
1302 value-guided decision-making in a less-is-more task. *Nature Communications* **8**, 1886 (2017).
130317. Philiastides, M. G., Biele, G. & Heekeren, H. R. A mechanistic account of value computation
1304 in the human brain. *PNAS* **107**, 9430–9435 (2010).
130518. Wunderlich, K., Dayan, P. & Dolan, R. J. Mapping value based planning and extensively
1306 trained choice in the human brain. *Nature Neuroscience* **15**, 786 (2012).
130719. Hunt, L. T. *et al.* Triple dissociation of attention and decision computations across prefrontal
1308 cortex. *Nature Neuroscience* **21**, 1471–1481 (2018).
130920. Lim, S.-L., O'Doherty, J. P. & Rangel, A. The Decision Value Computations in the vmPFC
1310 and Striatum Use a Relative Value Code That is Guided by Visual Attention. *J. Neurosci.* **31**,
1311 13214–13223 (2011).
131221. Lopez-Perse, A., Domenech, P. & Pessiglione, M. How prior preferences determine
1313 decision-making frames and biases in the human brain. *eLife* **5**, (2016).
131422. Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for
1315 exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
131623. Kolling, N., Behrens, T. E. J., Mars, R. B. & Rushworth, M. F. S. Neural Mechanisms of
1317 Foraging. *Science* **336**, 95–98 (2012).
131824. Zajkowski, W. K., Kossut, M. & Wilson, R. C. A causal role for right frontopolar cortex in
1319 directed, but not random, exploration. *eLife* <https://elifesciences.org/articles/27430> (2017)
1320 doi:10.7554/eLife.27430.
132125. Badre, D., Doll, B. B., Long, N. M. & Frank, M. J. Rostrolateral Prefrontal Cortex and
1322 Individual Differences in Uncertainty-Driven Exploration. *Neuron* **73**, 595–607 (2012).
132326. Costa, V. D., Mitz, A. R. & Averbeck, B. B. Subcortical Substrates of Explore-Exploit
1324 Decisions in Primates. *Neuron* **103**, 533-545.e5 (2019).
132527. Noonan, M. P., Kolling, N., Walton, M. E. & Rushworth, M. F. S. Re-evaluating the role of
1326 the orbitofrontal cortex in reward and reinforcement: Re-evaluating the OFC. *European
1327 Journal of Neuroscience* **35**, 997–1010 (2012).
132828. Hunt, L. T. *et al.* Mechanisms underlying cortical activity during value-guided choice. *Nat
1329 Neurosci* **15**, 470-S3 (2012).
133029. Rushworth, M. F. S., Noonan, M. P., Boorman, E. D., Walton, M. E. & Behrens, T. E.
1331 Frontal Cortex and Reward-Guided Learning and Decision-Making. *Neuron* **70**, 1054–1069
1332 (2011).
133330. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal Cortex as a
1334 Cognitive Map of Task Space. *Neuron* **81**, 267–279 (2014).
133531. Meder, D. *et al.* Simultaneous representation of a spectrum of dynamically changing value
1336 estimates during decision making. *Nat Commun* **8**, 1942 (2017).

133732. Kolling, N., Wittmann, M. & Rushworth, M. F. S. Multiple Neural Mechanisms of Decision
1338 Making and Their Competition under Changing Risk Pressure. *Neuron* **81**, 1190–1202
1339 (2014).
134033. Wittmann, M. K. *et al.* Predictive decision making driven by multiple time-linked reward
1341 representations in the anterior cingulate cortex. *Nature Communications* **7**, 12327 (2016).
134234. Boorman, E. D., Behrens, T. E. & Rushworth, M. F. Counterfactual Choice and Learning in a
1343 Neural Network Centered on Human Lateral Frontopolar Cortex. *PLOS Biology* **9**, e1001093
1344 (2011).
134535. Boorman, E. D., Rushworth, M. F. & Behrens, T. E. Ventromedial Prefrontal and Anterior
1346 Cingulate Cortex Adopt Choice and Default Reference Frames during Sequential Multi-
1347 Alternative Choice. *J. Neurosci.* **33**, 2242–2253 (2013).
134836. Kolling, N., Behrens, T., Wittmann, M. & Rushworth, M. Multiple signals in anterior
1349 cingulate cortex. *Current Opinion in Neurobiology* **37**, 36–43 (2016).
135037. Kolling, N. *et al.* Value, search, persistence and model updating in anterior cingulate cortex.
1351 *Nature Neuroscience* **19**, 1280 (2016).
135238. Hayden, B. Y., Pearson, J. M. & Platt, M. L. Neuronal basis of sequential foraging decisions
1353 in a patchy environment. *Nature Neuroscience* **14**, 933–939 (2011).
135439. Quilodran, R., Rothé, M. & Procyk, E. Behavioral Shifts and Action Valuation in the
1355 Anterior Cingulate Cortex. *Neuron* **57**, 314–325 (2008).
135640. Stoll, F. M., Fontanier, V. & Procyk, E. Specific frontal neural dynamics contribute to
1357 decisions to check. *Nature Communications* **7**, 11990 (2016).
135841. Karlsson, M. P., Tervo, D. G. R. & Karpova, A. Y. Network Resets in Medial Prefrontal
1359 Cortex Mark the Onset of Behavioral Uncertainty. *Science* **338**, 135–139 (2012).
136042. O'Reilly, J. X. *et al.* Dissociable effects of surprise and model update in parietal and anterior
1361 cingulate cortex. *PNAS* **110**, E3660–E3669 (2013).
136243. Tervo, D. G. R. *et al.* Behavioral Variability through Stochastic Choice and Its Gating by
1363 Anterior Cingulate Cortex. *Cell* **159**, 21–32 (2014).
136444. Bernacchia, A., Seo, H., Lee, D. & Wang, X.-J. A reservoir of time constants for memory
1365 traces in cortical neurons. *Nat Neurosci* **14**, 366–372 (2011).
136645. Lebreton, M., Abitbol, R., Daunizeau, J. & Pessiglione, M. Automatic integration of
1367 confidence in the brain valuation signal. *Nature Neuroscience* **18**, 1159 (2015).
136846. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and
1369 implementation as FSL. *Neuroimage* **23 Suppl 1**, S208-219 (2004).
137047. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of
1371 brain images. *Medical Image Analysis* **5**, 143–156 (2001).

1372

1373 Acknowledgements

1374

1375 NT was funded by a DTC ESRC studentship (ES/J500112/1), JS was supported by a MRC
1376 Skills Development Fellowship (MR/NO14448/1), MCKF by a Sir Henry Wellcome
1377 Fellowship (103184/Z/13/Z), MFSR was funded by a Wellcome Senior Investigator Award
1378 (WT100973AIA). We would like to thank all members of the Rushworth lab for great
1379 discussions on this project.

1380

1381

1382 Author contributions

1383

1384 NT, MKW and MFSR conceived and designed the experiment, NT, JS and MKW
1385 constructed the Bayesian model, NT conducted the experiment, NT, EF, LT, MKW and
1386 MFSR conceived behavioural analyses, NT, MCKF, MKW and MFSR conceived neural
1387 analyses, NT conducted data analyses, NT, MKW and MFSR wrote the manuscript, all
1388 authors provided expertise and feedback on the write-up, MKW and MFSR supervised the
1389 research project.

1390

1391

1392 Competing interest

1393

1394 The authors declare no financial or non-financial competing interests.

1395

1396

1397 **Figure legends**
1398

1399 **Figure 1. Experimental Task and Design.**

1400 (A) Trial timeline. In each trial, participants made two choices. First, a binary choice between
1401 two predictors (coloured boxes; decision phase) to receive information about a target's
1402 location on a circle. The goal was to choose predictors that accurately predicted the target
1403 location. The length of a black bar at the bottom of the screen informed participants about the
1404 number of remaining trials in the current block. Second, participants indicated their belief in
1405 the accuracy of the chosen predictor by modifying the size (dotted lines) of an interval
1406 symmetrical around the reference point (confidence phase). In the outcome phase, the target
1407 location (star) and any points earned were indicated. Two possible example outcomes are
1408 illustrated. In the above case, the participant's prediction was incorrect as the target fell
1409 outside the interval, resulting in a null payoff. In the bottom case, the target fell within the
1410 interval, resulting in a positive payoff. Positive payoffs increase with narrower intervals as
1411 long as the target falls within the interval. (B) Design. (B-i) Participants transitioned through
1412 blocks of different numbers of trials (time horizons). (B-ii) Each time horizon introduced four
1413 new predictors (illustrated as boxes) that were categorised into two good (green and yellow
1414 boxes) and two bad predictors (orange and blue boxes) according to how well they predicted
1415 the target. The quality of predictions was determined by the angular error between target and
1416 reference location with a smaller angular error representing better target predictions.
1417

1418 **Figure 2. Task statistics, Bayesian model, and choice hypotheses.**

1419 (A) Panels depict the mapping between observations during the task (i), their statistical
1420 properties (ii), and subjective beliefs about these properties derived with Bayes' rule (iii;iv).
1421 (A-i) A predictor's performance can be evaluated by the angular error at each trial (left
1422 panel), and by comparing angular errors between predictors across observations (right panel).
1423 Better predictors have on average smaller angular errors (green is better than orange). (A-ii)
1424 Predictors' angular errors were derived from normal distributions centred on the true target
1425 location. Critically, the normal distributions for good and bad predictors differed in their
1426 standard deviation (σ): smaller σ 's reflected smaller angular errors, i.e. more
1427 accurate predictions of the true target location. Learning about a predictor's angular error
1428 across time corresponded to forming beliefs about a predictor's σ value. (A-iii) To
1429 capture this learning process, we used Bayesian modelling and derived trial-wise belief
1430 distributions over σ for each predictor. In other words, we estimated a probability density
1431 function that expressed the belief strength in each possible σ over a large range of
1432 σ s, and that was updated with each new observation via Bayes' rule. The coloured
1433 vertical lines indicate the true underlying σ s of the predictors and the black distributions
1434 reflect the Bayesian approximation after extensive training. (A-iv) We captured two
1435 separable estimates about participants' beliefs concerning predictors: an estimate of the
1436 accuracy of a predictor (the mode of the distribution indicated by the position of the vertical
1437 line on the abscissa) and the uncertainty in that belief (width of the belief distribution). (B) In
1438 all panels, light to dark orange represents earlier and later trials, respectively, in a block. Left:
1439 Prior beliefs are updated after observing the angular error in the trial's outcome phase,
1440 resulting in a posterior belief. The posterior belief forms the prior for the next encounter with
1441 the same predictor. Right: Belief distribution when selecting the same predictor multiple
1442 times. Across time, the belief distribution will converge towards the true value of σ
1443 (here, true σ is 50). (C) Experimental hypotheses. Note that panels depict an illustration
1444 of hypothesized effect sizes of accuracy and uncertainty on choice akin to logistic GLM
1445 analyses of choice. (C-i) Participants' patterns of explore/exploit choices should
1446 systematically change over the course of the blocks. At the beginning of a block (light orange

area), participants should pursue the more uncertain predictor, that is choices should be driven by a positive uncertainty effect, but this tendency should reverse over time. Accurate predictors should be sought out throughout (positive accuracy effect), but particularly towards the end of the block (dark orange area) when the value of exploration diminishes. **(C-ii)** At the time of initial choices (indicated by black boxes in inset), the value of exploration should be modulated by the time horizon and choices towards uncertain predictors should systematically increase if there are more trials remaining in which to exploit the knowledge gained, i.e. in longer horizons (vice versa for accuracy-driven choices).

1456

1457 **Figure 3. Dissociable effects of accuracy and uncertainty on predictor selections and**
1458 **subjective confidence judgments.**

1459 **(A) Decision phase.** By using logistic GLM analyses we predict leftward predictor selection
1460 as a function of several variables (coded as left minus right). In general, participants preferred
1461 accurate predictors (accuracy: $t(23)=7.5$, $p<0.001$, $d=1.52$, 95% confidence interval=[0.8
1462 1.45]). There was no credible evidence for an uncertainty effect on behaviour ($t(23)=-1.9$, $p=$
1463 0.07, $d=-0.39$, 95% confidence interval=[-0.51 0.018], Bayes factor₁₀=1.05, %error=1.1017e-
1464 4). However, uncertainty and accuracy exerted different effects depending on when choices
1465 were made: uncertain predictors were explored when many trials remained (positive
1466 interaction term with percentage of remaining trials, i.e. block time; $t(23)=5.8$, $p<0.001$,
1467 $d=1.18$, 95% confidence interval=[0.53 1.1]), whereas decisions were accuracy-driven as the
1468 end of a block approached (negative interaction effect with block time; $t(23)=7.5$, $p<0.001$,
1469 $d=-1.53$, 95% confidence interval=[-0.91 -0.52]). **(B) Decision phase. (B-i)** Trials were
1470 binned into first and second halves of each block (independent of time horizon length) to
1471 examine the interaction effects shown in panel A. Earlier choices (i.e. first half) were more
1472 uncertainty-driven compared to later (i.e. second half) choices when uncertainty was avoided
1473 (paired-t-test early vs late: $t(23) = -8.1$, $p<0.001$, $d=1.66$, 95%confidence interval=[1.06 1.8]).
1474 In contrast, accuracy determined choices throughout both early and late block halves, but
1475 increasingly so in the second half (paired t-test early vs late: $t(23) = -4.2$, $p<0.001$, $d=-$
1476 0.85, 95%confidence interval=[-1.63 -0.55]). Both accuracy and uncertainty changed
1477 differently across block halves (paired t-test between differences of block halves for accuracy
1478 and uncertainty: $t(23) = -8.1$, $p<0.001$, $d=-1.7$, 95% confidence interval =[-2.27 -1.02]). **(B-ii)**
1479 Accuracy and uncertainty effects on choice also varied as a function of how many trials still
1480 remained within a block: differences in the initial choice patterns (first 15 trials; see inset)
1481 across horizons showed that the exploration of uncertain predictors was more pronounced
1482 when horizons were longer while shorter horizons demanded more rapid exploitation of
1483 predictors estimated as most accurate (3x2 ANOVA: $F(2,46)=36.7$, $p<0.001$, $\eta^2=0.62$). **(C)**
1484 **Confidence phase.** Trial-by-trial confidence judgments increased (i.e. the confidence interval
1485 size decreased) when selecting predictors that were believed to be accurate ($t(23)=11.7$,
1486 $p<0.001$, $d=2.4$, 95%confidence interval=[0.66 0.98]) but decreased when predictors were
1487 believed to be uncertain according to the Bayesian model ($t(23)=-10.4$, $p<0.001$, $d=-$
1488 2.12, 95% confidence interval=[-1.1 -0.73]). Note that we used the inverse of the confidence
1489 interval such that a greater confidence index also represents higher confidence. ($n = 24$; error
1490 bars are SEM across participants).

1491

1492

1493

1494

1495 **Figure 4. Modulation of uncertainty prediction difference in vmPFC according to**
1496 **behavioural mode.**

1497 (A) Across all trials, a negative uncertainty (i) and positive accuracy (ii) prediction
 1498 differences covaried with activation in vmPFC. (B) We found a polarity change in the impact
 1499 uncertainty exerted on predictor selection at a behavioural level; initial trials in longer
 1500 horizons were more likely to be explorative and directed towards more uncertain predictors
 1501 while behaviour in later trials was more exploitative and directed away from uncertain
 1502 predictors, in other words they selected certain predictors (see labels on y-axis). We tested for
 1503 a neural uncertainty polarity change in vmPFC comparing behavioural modes of exploration
 1504 and exploitation, respectively, representing a positive and then negative uncertainty
 1505 prediction difference. (C) Time courses extracted from vmPFC for both chosen and unchosen
 1506 components of an uncertainty prediction difference signal during exploration (i) and
 1507 exploitation (ii). VmPFC BOLD activity changed in accordance with the behavioural results;
 1508 it transitioned from activity positively related to uncertainty prediction difference (positively
 1509 encoding the uncertainty of the chosen predictor as opposed to the unchosen predictor) during
 1510 initial choices to activity negatively related to uncertainty prediction difference (negatively
 1511 encoding the uncertainty of the chosen predictor as opposed to the unchosen predictor) in
 1512 later trials. All effects were time-locked to the decision phase. ($n = 24$; error bars are SEM
 1513 across participants; whole-brain effects family-wise error cluster corrected with $z > 2.3$ and $p < 0.05$). (D)
 1514 The relationship between accuracy and uncertainty prediction differences used
 1515 for all neural analyses across all trials (left) exploration trials (centre), and exploitation trials
 1516 (right). Average correlations between accuracy and uncertainty prediction differences across
 1517 all participants are reported at the bottom of each panel, while panels show variables across
 1518 time taken from a representative participant for each analysis. Accuracy and uncertainty
 1519 prediction differences are similarly decorrelated in all other analyses (for details on
 1520 correlation, see Supplementary Figure 1, 2).

1521

1522 **Figure 5. Whole brain maps for uncertainty prediction difference during exploration**
1523 **and exploitation.**

1524 Illustrations above whole-brain images clarify the polarity (positive or negative) of the
 1525 uncertainty prediction difference signal represented in vmPFC (indicated by the black circle)
 1526 during exploitation, exploration and their contrast. (A) During exploitation, activity related to
 1527 an uncertainty prediction difference was restricted to a region centred on vmPFC and was
 1528 represented with a negative polarity (see inset). (B) However, during exploration uncertainty
 1529 prediction difference was represented with a positive polarity and associated with an
 1530 extended network including vmPFC but also dorsomedial frontal areas peaking in dorsal
 1531 anterior cingulate cortex (dACC) (see also Supplementary Figure 6). (C) Difference in
 1532 uncertainty prediction difference between exploration and exploitation. Contrasting
 1533 activations between the behavioural modes of exploration and exploitation confirmed the
 1534 presence of mode-specific (e.g. dACC) and mode-general (e.g. vmPFC) activations. Note that
 1535 the sign of activation patterns resulting from a contrast between exploration and exploitation
 1536 need to be interpreted with reference to the levels of activity found in the exploration and
 1537 exploitation phases with respect to baseline (see illustration above each whole-brain map) ($n = 24$;
 1538 whole-brain effects family-wise error cluster corrected with $z > 2.3$ and $p < 0.05$).

1539

1540

1541
1542
1543

1544 **Figure 6. Interaction of repetition and uncertainty representation in vmPFC.** (A) The
1545 percentage of choice repetitions during exploitation was significantly higher than during
1546 exploration (paired t-test explore vs exploit: $t(23)=-16.2$, $p < 0.001$, $d = -3.3$, 95% confidence
1547 interval = [-0.36 -0.28]). Also note that within the two phases, this indicates a relative
1548 predominance of repetitions versus no repetitions in exploitation, but a relative predominance
1549 of no repetition choices versus repetitions in exploration. (B) VmPFC activity increased when
1550 participants repeated the same predictor selection as they had made on the last encounter with
1551 the predictor (grey time course; repetition is coded as “repeat – no repeat”; $t(23) = 4$, p
1552 < 0.001 , $d = 0.8$, 95% confidence interval=[0.017 0.06]). Moreover, we found a significant
1553 interaction effect of repetition x chosen uncertainty (red time course; $t(23) = -3.4$, $p = 0.002$,
1554 $d = -0.7$, 95% confidence interval=[-0.07 -0.02]). The interaction effect is illustrated in the
1555 right panel by decomposing it into the binned effects of chosen uncertainty during
1556 “repetition” and “no repetition” trials at the time of the interaction effect time course peak.
1557 This indicates that the increase in BOLD response accompanying choice repetition was even
1558 stronger if participants were very certain about their choice (i.e. negative uncertainty during
1559 repetition; green bar in right panel); whereas in case of switching choices, the BOLD signal
1560 increased as a function of chosen uncertainty (i.e. positive uncertainty; blue bar in right
1561 panel). Note that the statistical test comparing the blue and green bars was performed in the
1562 leftward panel of B by testing the interaction effect against zero ($n = 24$; error bars are SEM
1563 across participants).

1564

1565 **Figure 7. Accuracy processing mediates uncertainty polarity change from exploration to**
1566 **exploitation.**

1567 (A) Transition trials (Supplementary Figure 9A) occurred later than exploratory selections and
1568 earlier than exploitative selections (left panel) (explore vs transition: $t(23)=6$, $p < 0.001$, $d=1.2$,
1569 95%confidence interval=[0.056 0.12]; transition vs exploit: $t(23)=-2.8$, $p=0.01$, $d=-0.57$,
1570 95%confidence interval=[-0.04 -0.006]). We hypothesized activation in vmPFC to be
1571 correlated with positive uncertainty, accuracy and negative uncertainty prediction differences
1572 between predictors, but at different times during the experiment (see illustration, right panel).
1573 (B) During transition trials, activation in vmPFC covaried with the difference in the accuracy
1574 between the chosen and unchosen predictor, i.e. accuracy prediction difference ($t(23) = 3.5$,
1575 $p = 0.002$, $d=0.71$, 95% confidence interval=[0.03 0.1]). (C-i) Participants who showed a
1576 stronger vmPFC accuracy prediction difference during the transition period (variability
1577 around time course peak from panel b), also integrated more drastically the uncertainty
1578 between predictors across time into their choice behaviour (uncertainty x block time from
1579 Figure 3A; $r = 0.58$, $p = 0.007$, 95% confidence interval=[0.23 0.8]). (ii) For illustration, this
1580 means that participants with stronger accuracy-related vmPFC activation had a stronger
1581 change in integrating uncertainty across time, i.e. a stronger slope in the uncertainty x block
1582 time effect. The illustration depicts two example participants, dark orange indicates a subject
1583 with both a strong vmPFC accuracy activation and pronounced behavioural change in how
1584 uncertainty was used to drive choice behaviour. By contrast, the participant indicated in light
1585 orange shows a weak vmPFC BOLD accuracy effect and only a small change in how
1586 uncertainty was used over time. These findings support the idea that the transition between
1587 positive uncertainty-driven exploration to negative uncertainty-driven exploitation is
1588 mediated by representing the accuracy between predictors. ($n = 24$; error bars are SEM across
1589 participants).

1590

1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608

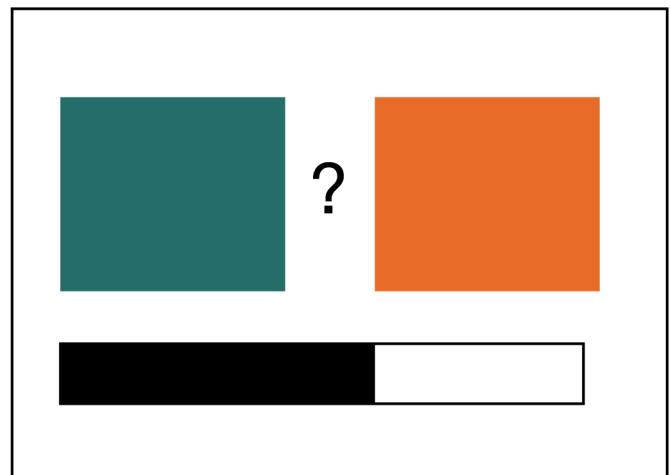
Figure 8. Summary. From exploration to exploitation: polarity of subjective uncertainty in vmPFC changes with behavioural mode.

At the beginning of a block, choices are exploratory and directed towards uncertain predictors (like a shuffle mode when playing music, left panel). VmPFC and an extended network centred in dACC represent the difference in uncertainty between the predictors that might be selected. With time passing, participants learn about the predictors' accuracy through observing how well they predict an outcome. A participant's belief in the accuracy of the predictors exerts the predominant influence on vmPFC activity during this transition phase (middle panel). Towards the end of a block, vmPFC activity represents the difference in negative uncertainty, in other words the certainty between predictors. In this exploitative period, choices are repeatedly directed towards certain predictors (like a repeat mode, right panel). We show that vmPFC carries information about a multiplicity of decision variables, the strength and polarity of which vary according to their relevance for the current context of exploration, exploitation or their transition.

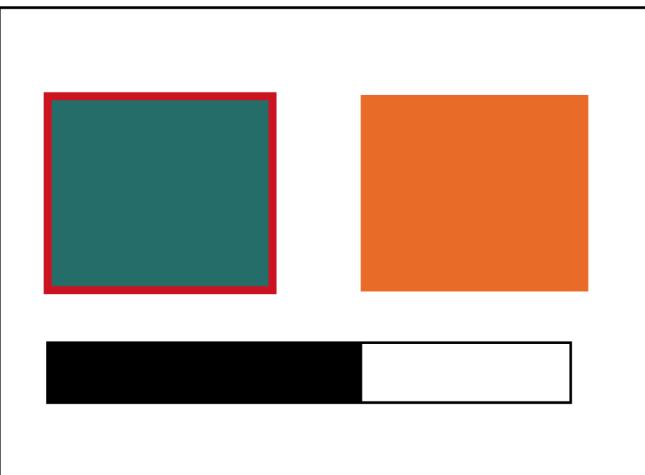
A Trial timeline

Decision

1.5 sec

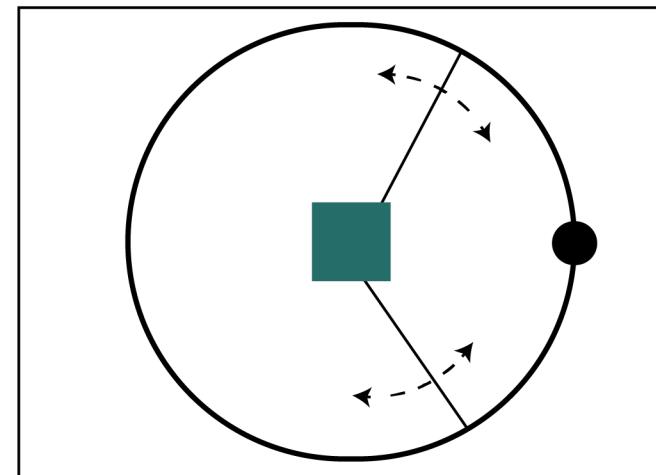


0.5 sec + RT

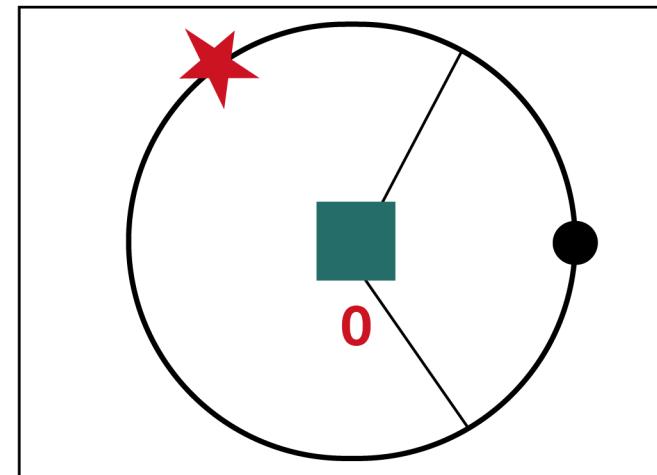


Confidence

RT



3 sec



ITI

B Design

i) time horizons:

45 trials

30 trials

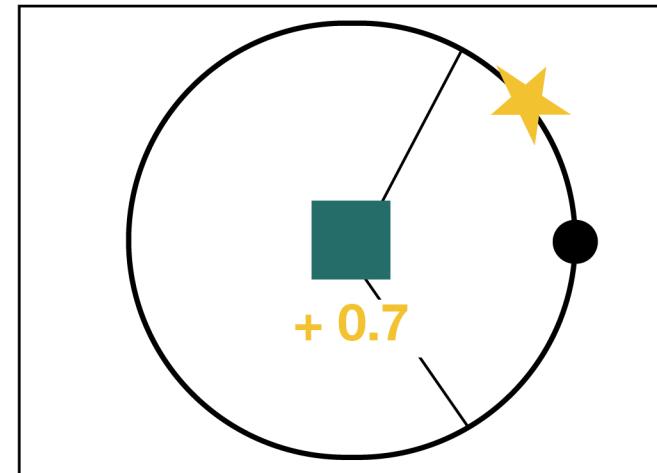
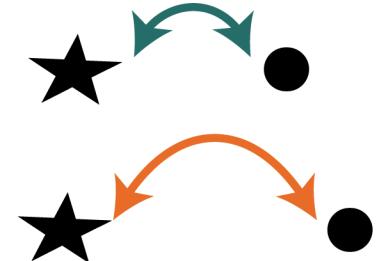
15 trials

ii) per block:

two good
predictors

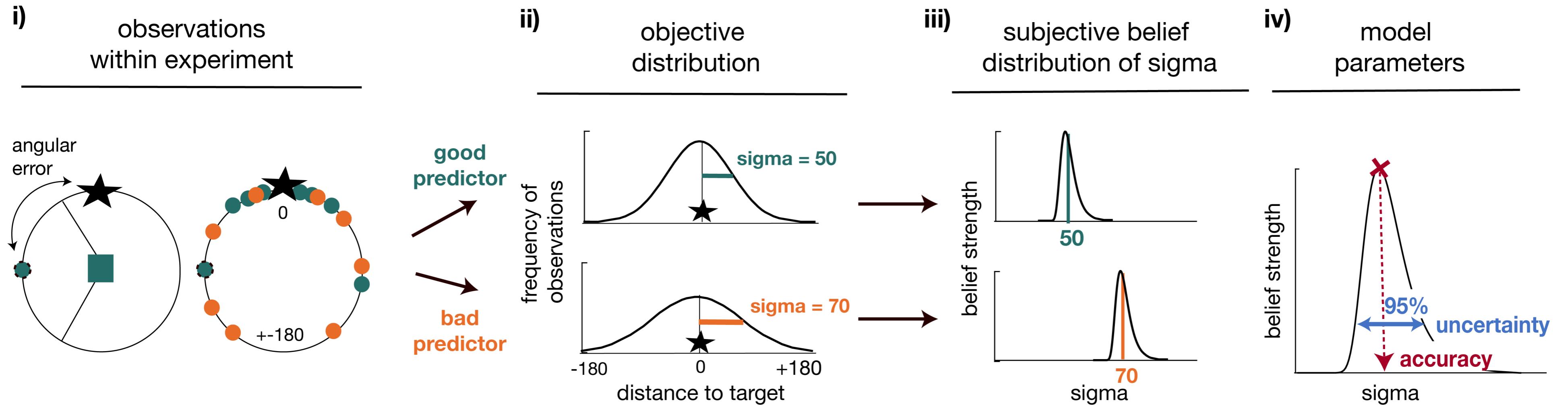
two bad
predictors

angular error:

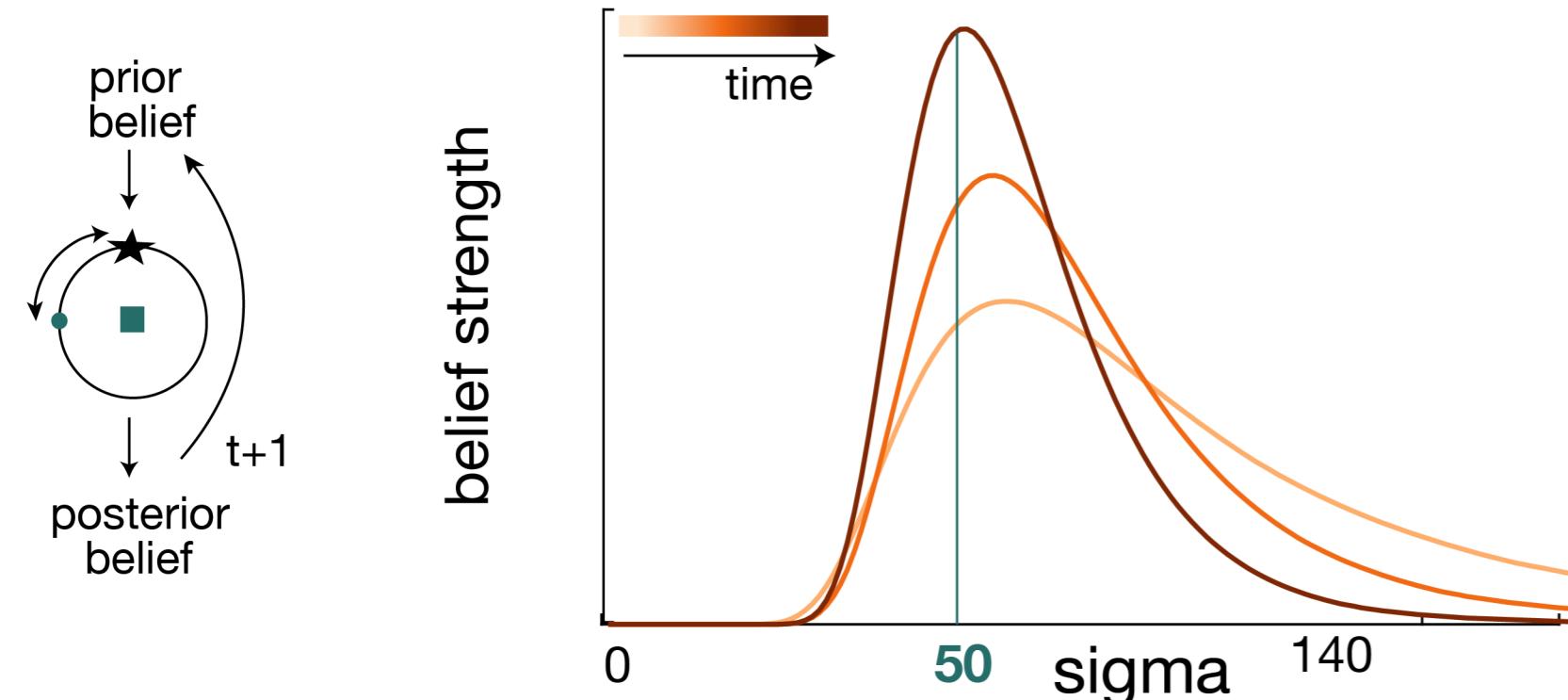


A

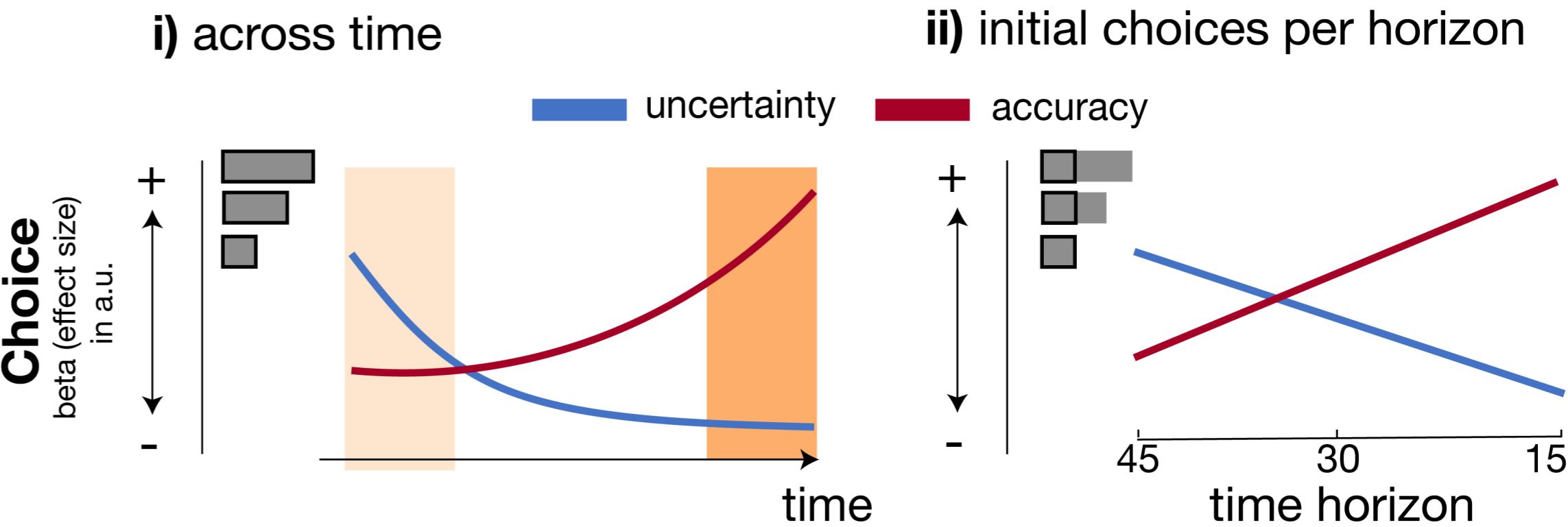
Relationship between task parameters and Bayesian belief formation

**B**

Bayesian update across time

**C**

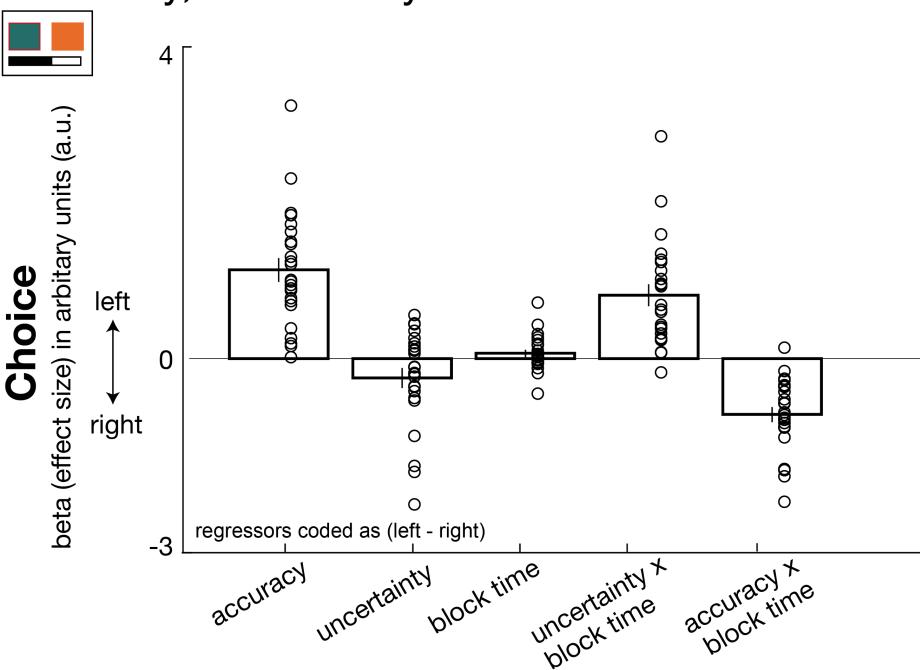
Hypotheses: accuracy and uncertainty effects on choice behaviour



DECISION PHASE

A

Accuracy, uncertainty and time

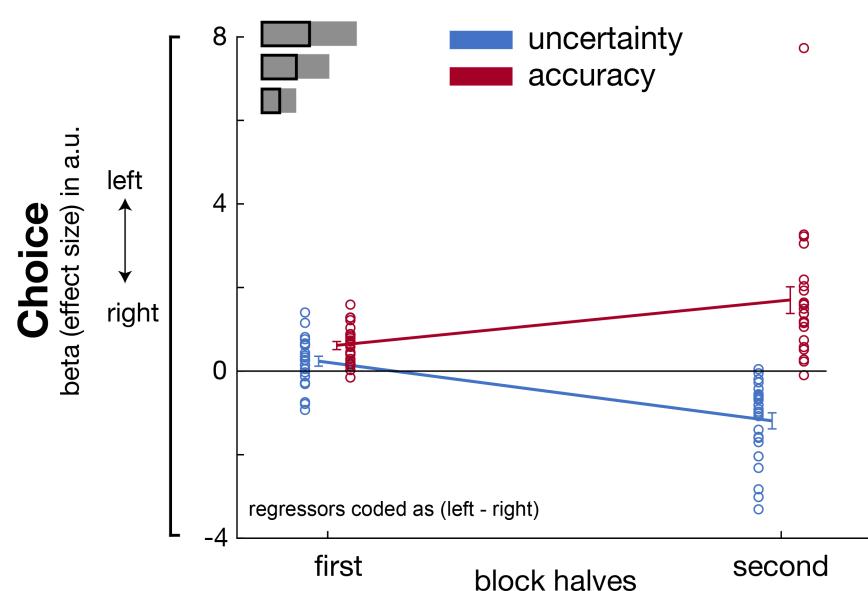


B

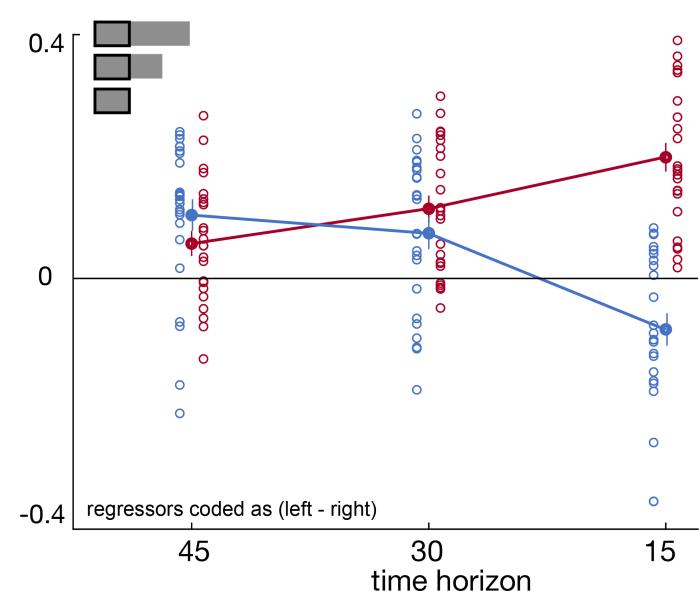
Time modulations of uncertainty and accuracy



i) Parameters across time



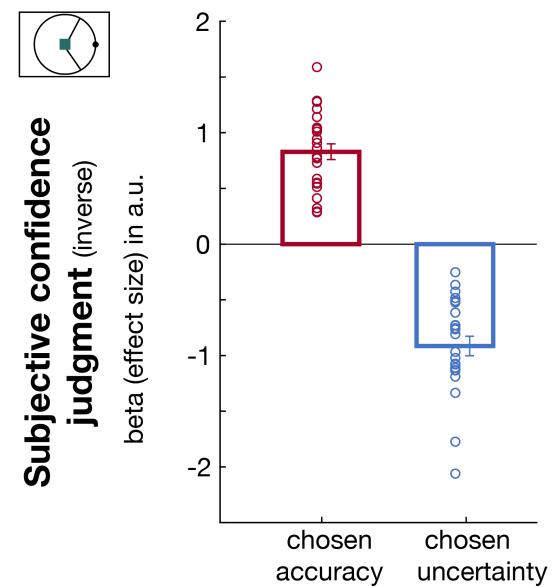
ii) Initial choices per horizon



CONFIDENCE PHASE

C

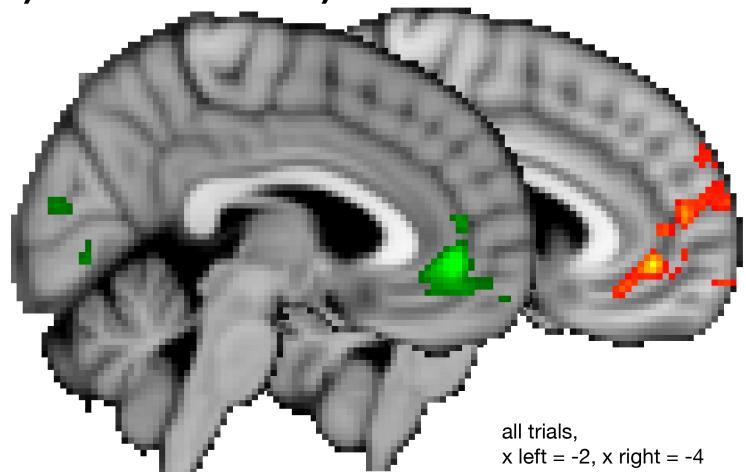
Subjective confidence judgments



A

Prediction differences across all trials

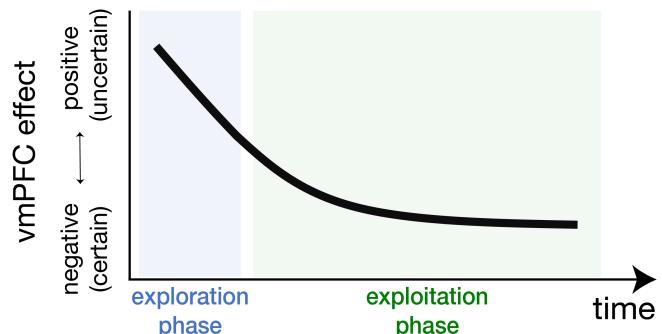
█ negative uncertainty █ accuracy

i)**ii)****B**

Polarity change:
from positive to negative uncertainty
representation across time

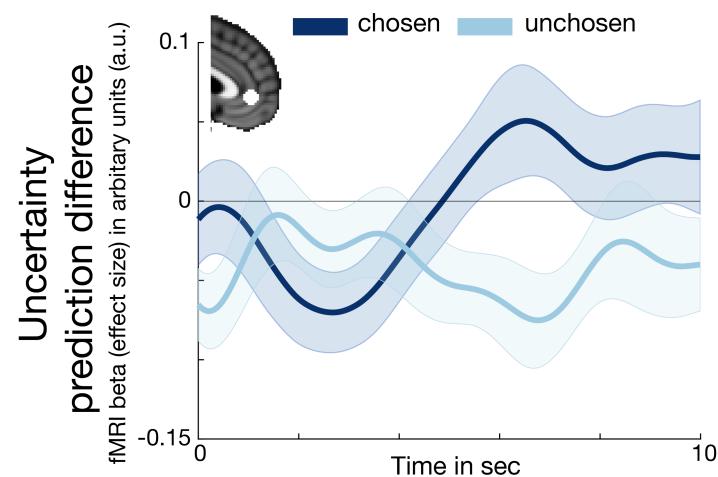
Hypothesis:

█ positive uncertainty █ negative uncertainty

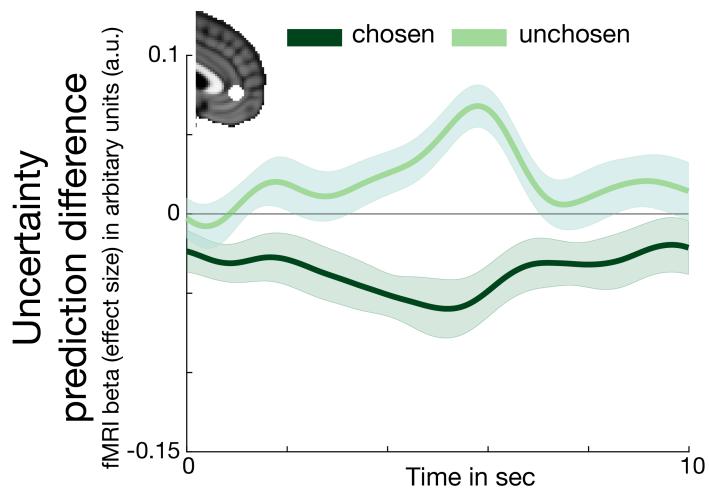
**C**

Polarity change in vmPFC covaries with behavioural mode

i) Exploration: positive uncertainty prediction difference



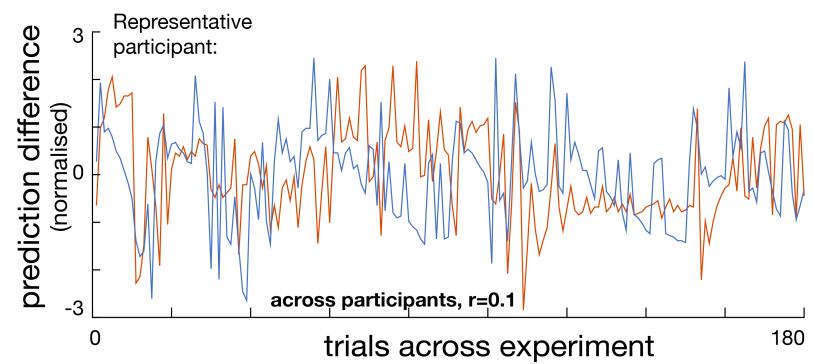
ii) Exploitation: negative uncertainty prediction difference

**D**

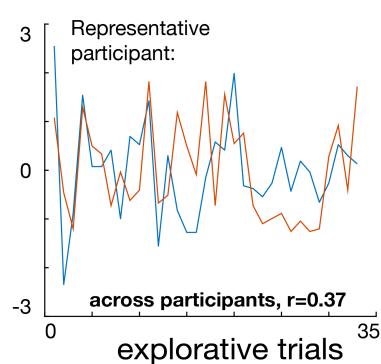
Relationship between accuracy and uncertainty prediction differences

All trials:

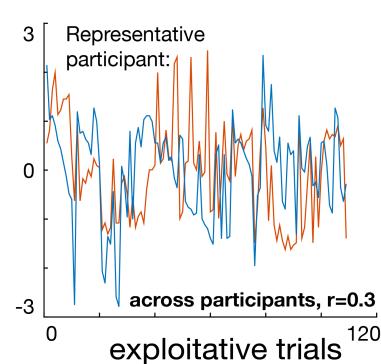
█ uncertainty █ accuracy



Exploration:

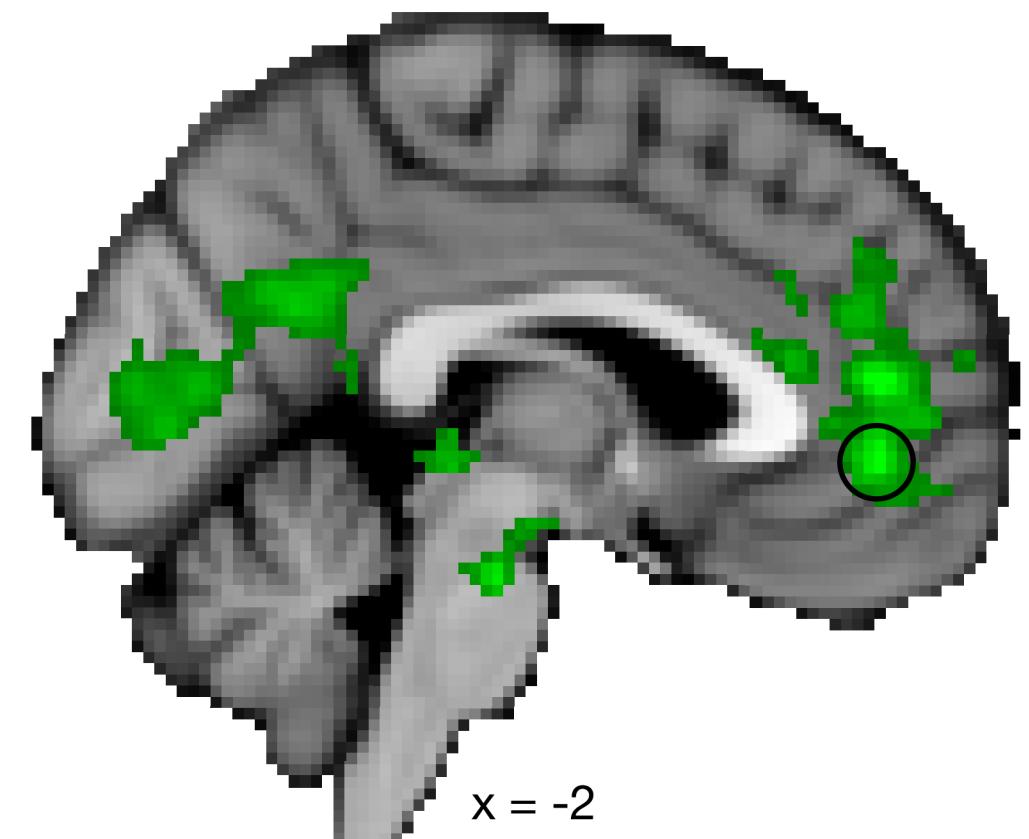
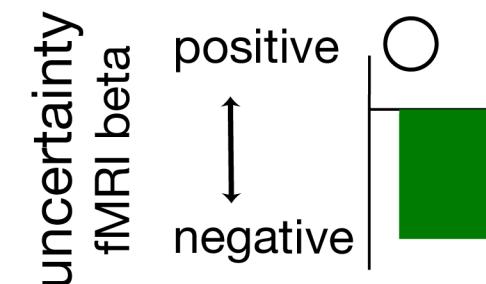


Exploitation:

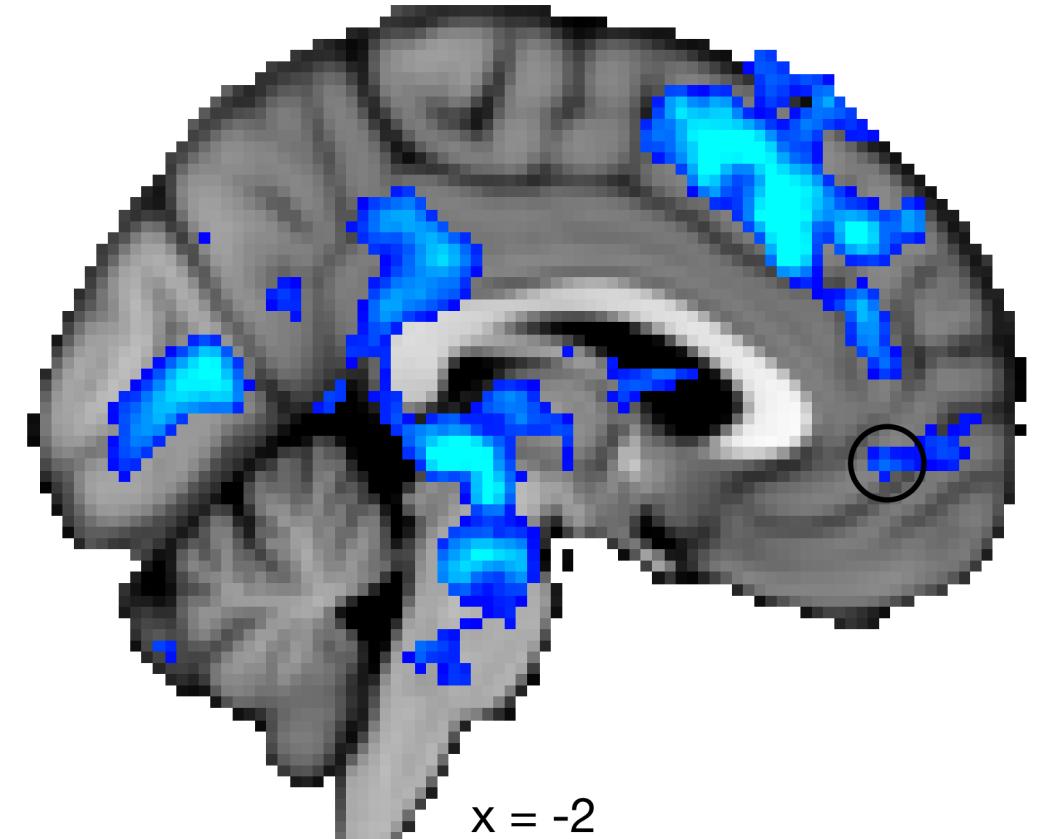
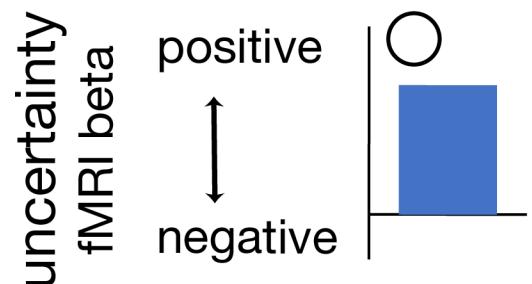


A

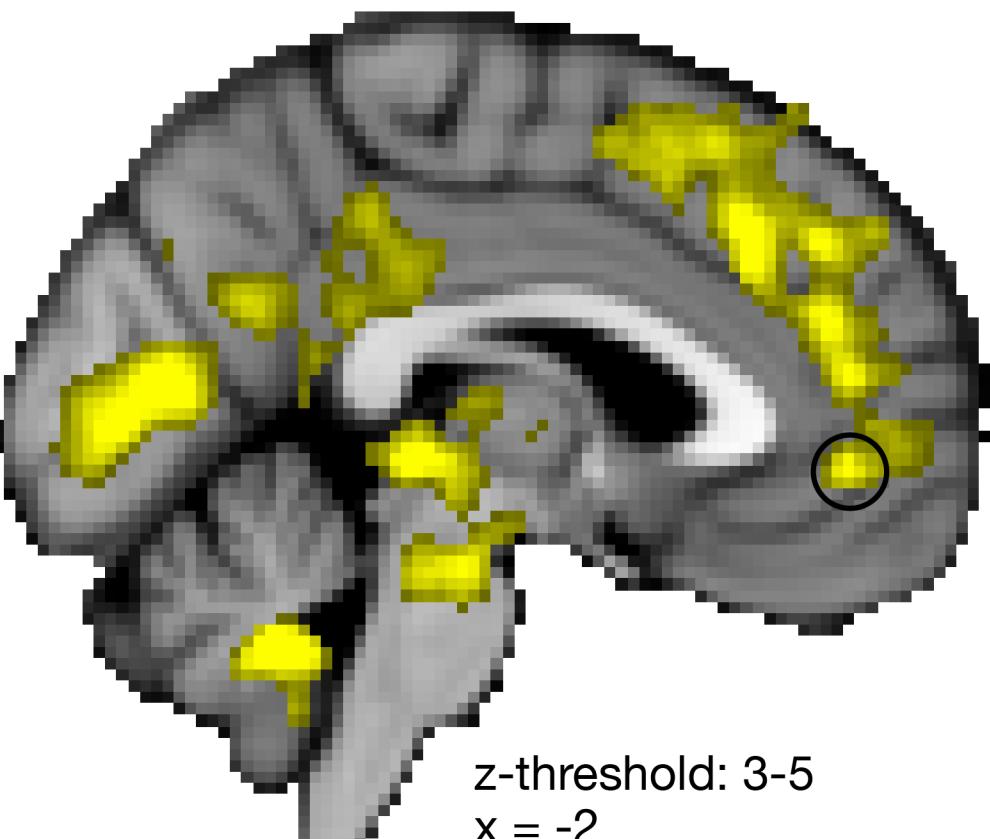
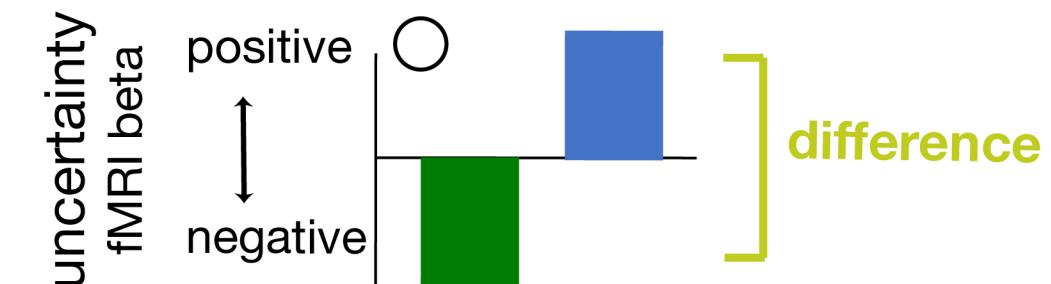
Exploitation:
negative uncertainty
prediction difference

**B**

Exploration:
positive uncertainty
prediction difference

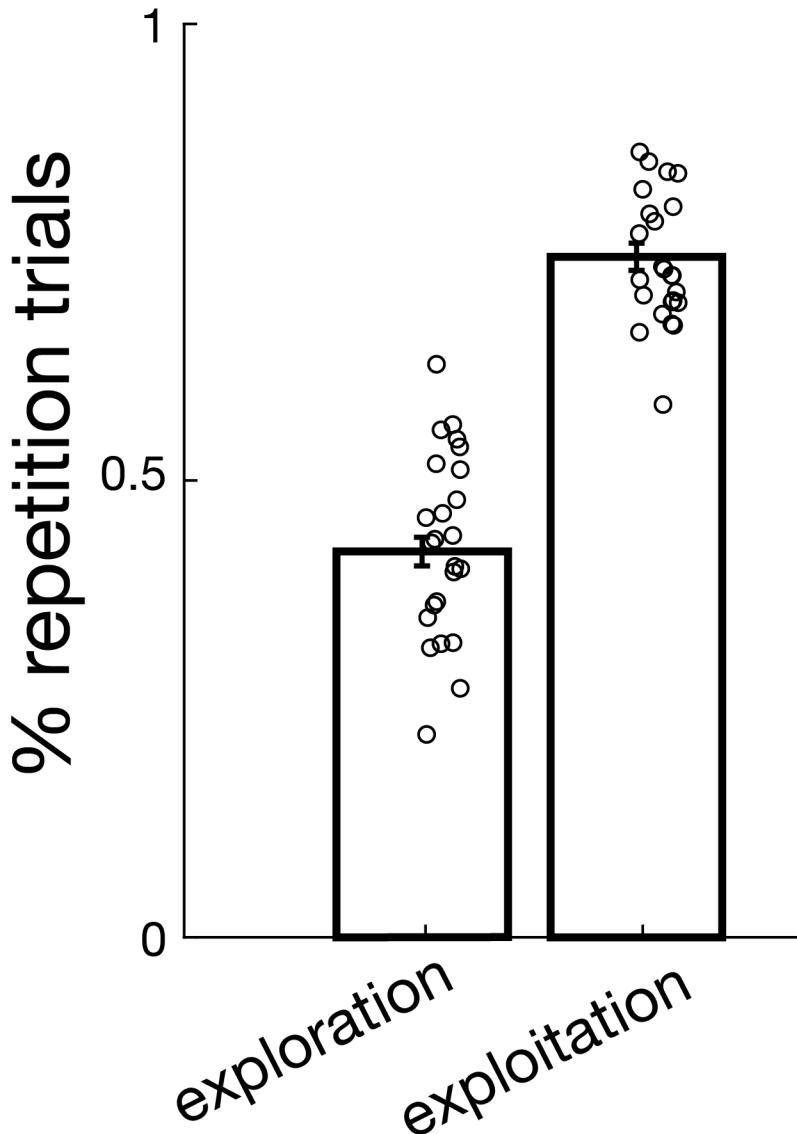
**C**

Exploration - Exploitation:

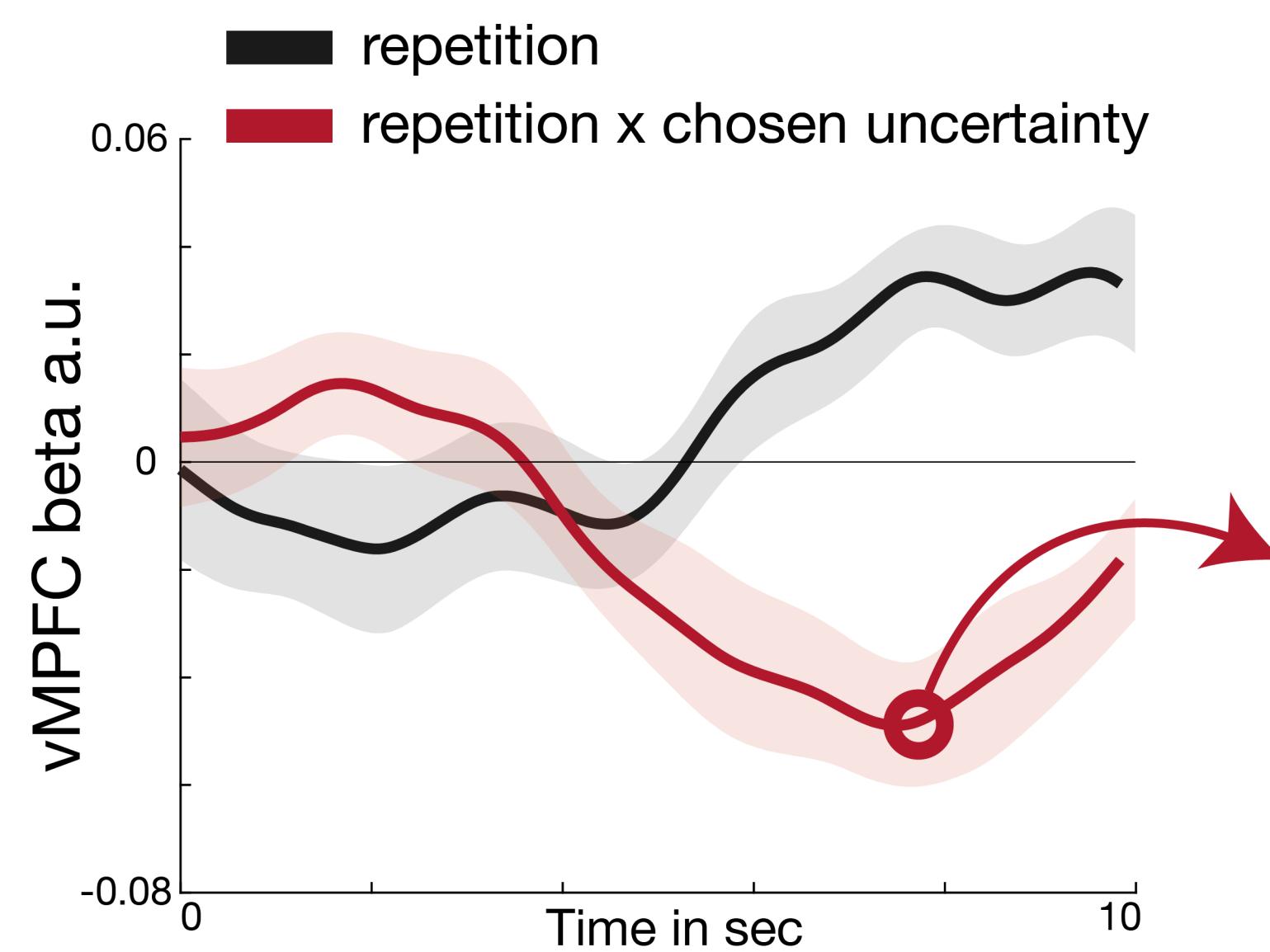


A

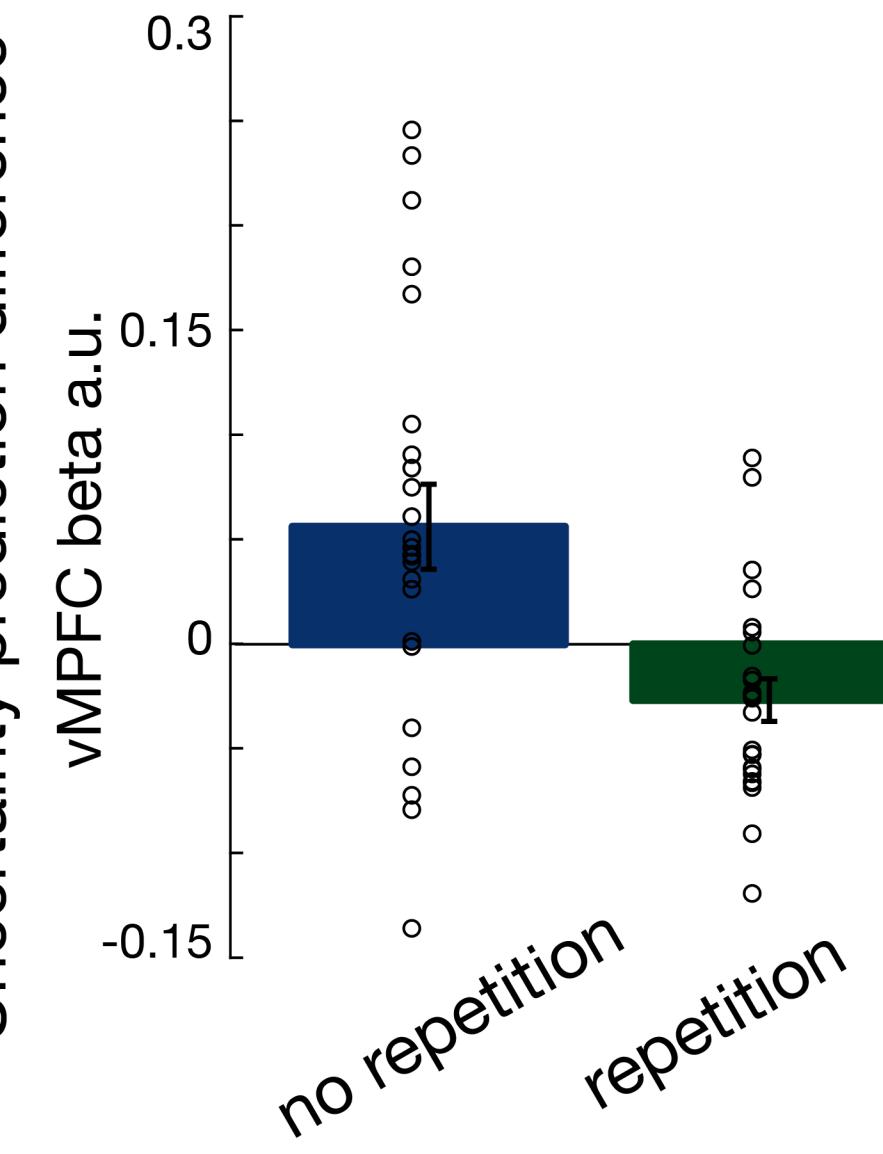
Repetition trials are mainly present during exploitation

**B**

Chosen uncertainty in vmPFC interacts with predictor repetition

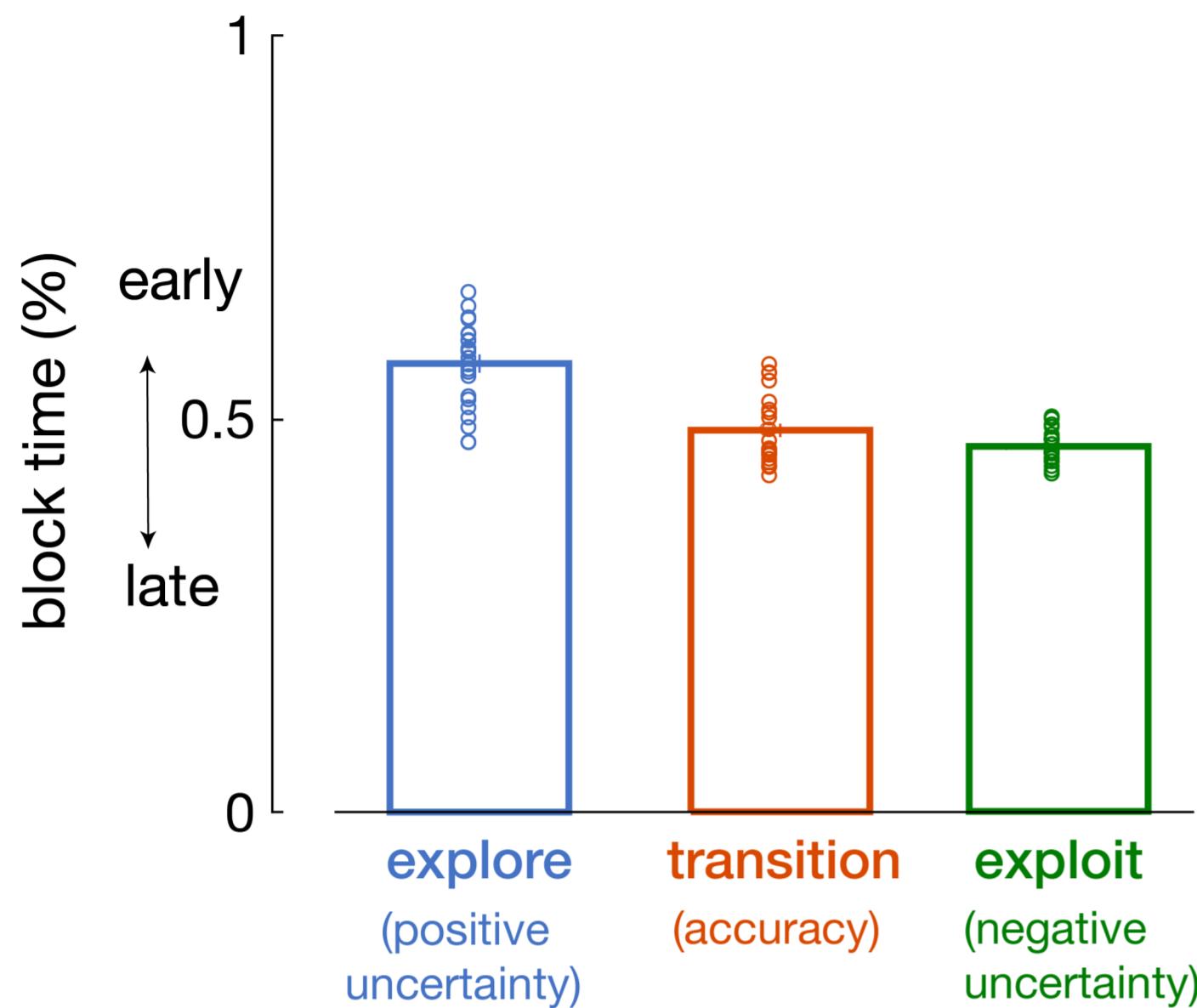


Uncertainty prediction difference

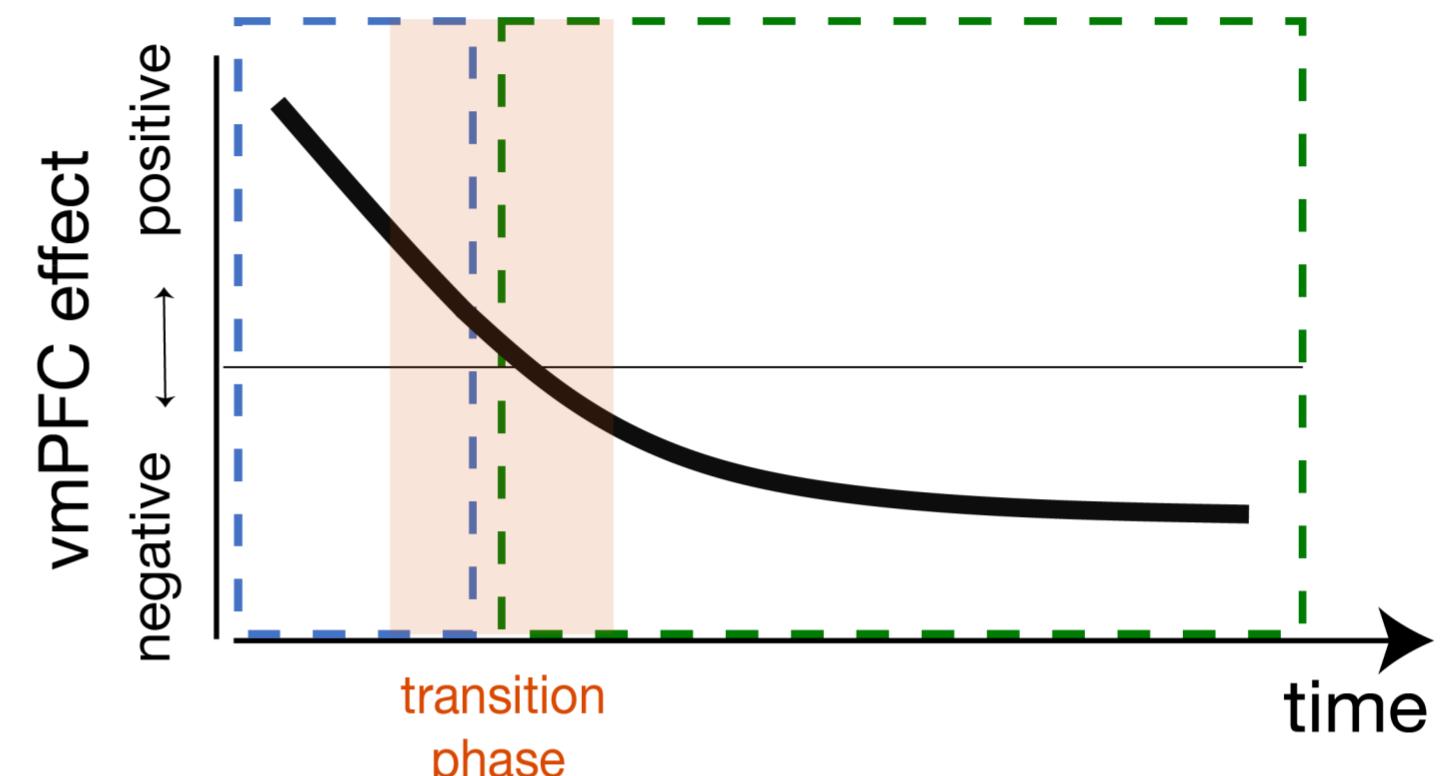


A

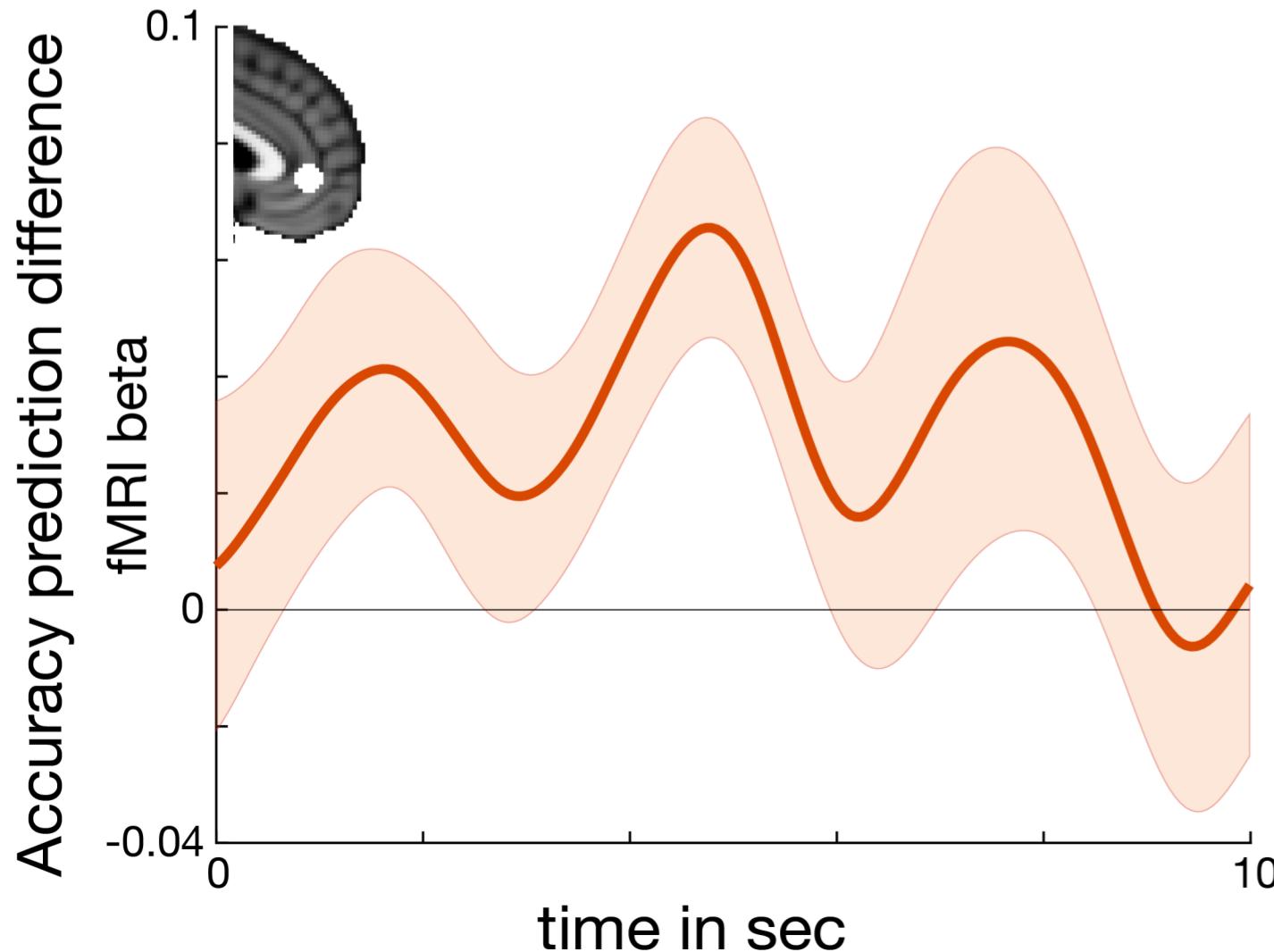
Timing of transition trials



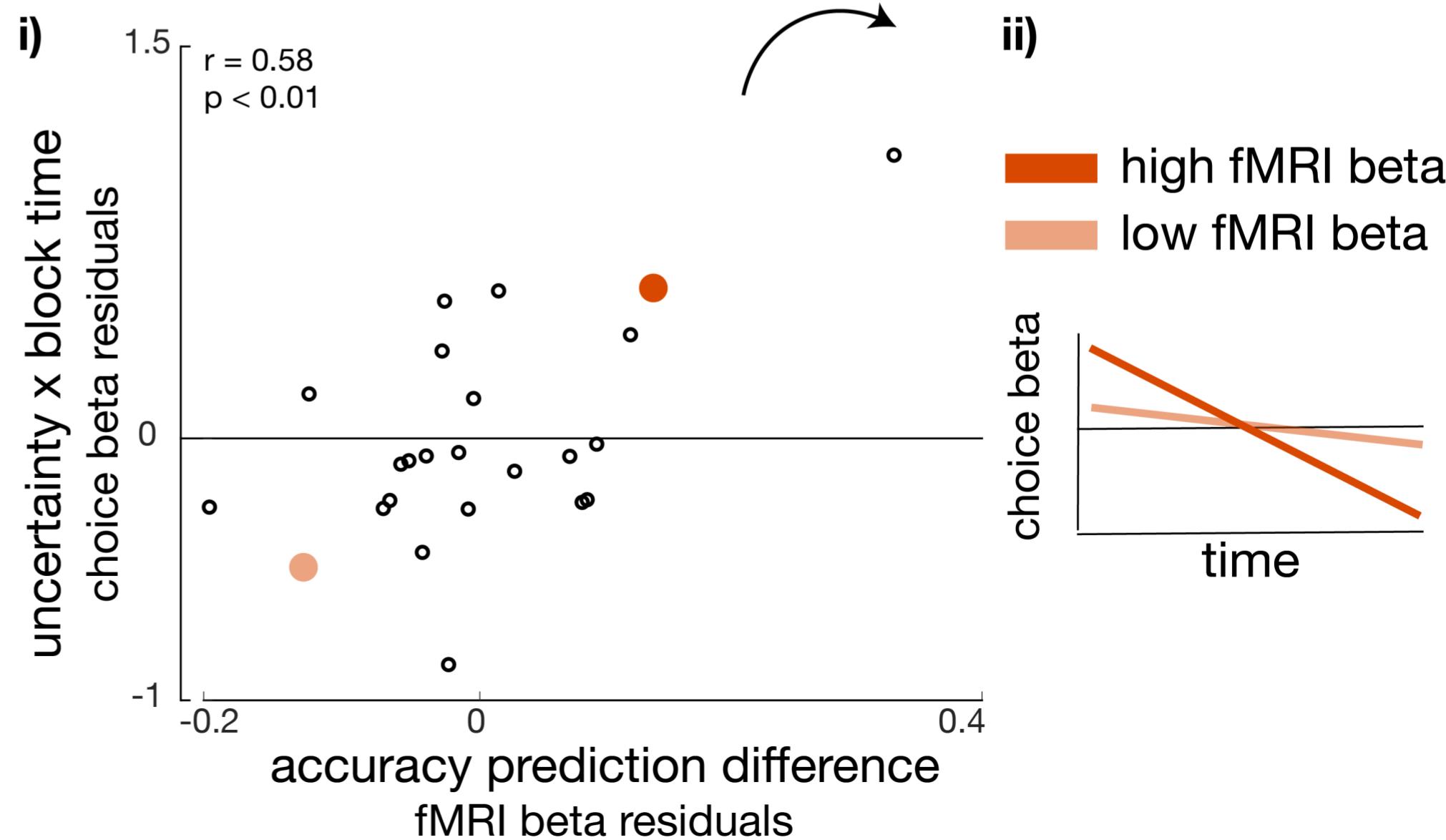
Hypothesis:

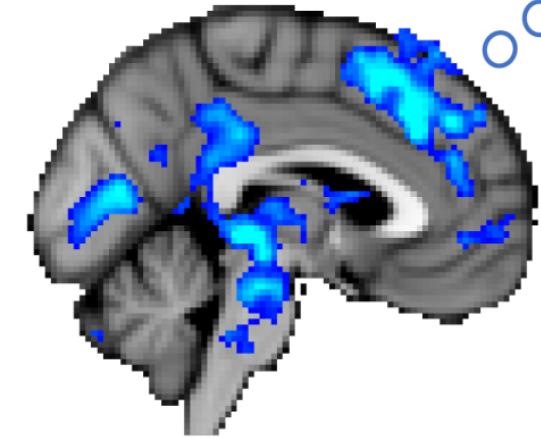
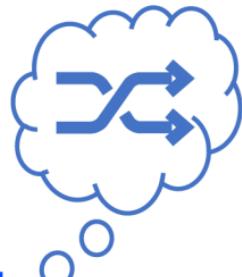
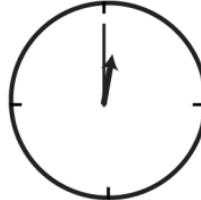
**B**

VmPFC activity covaries with accuracy prediction difference during transition

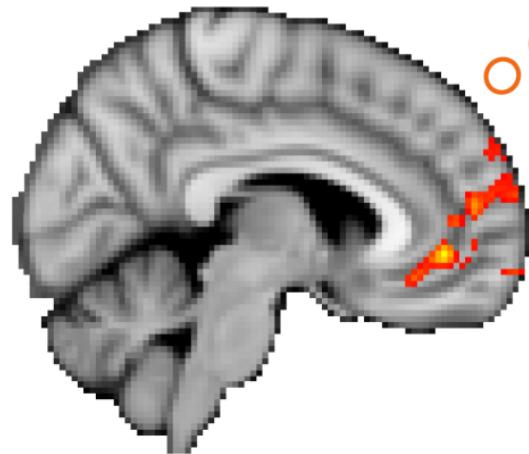
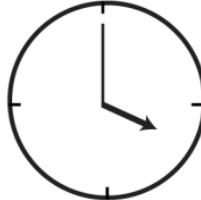
**C**

VmPFC activity during transition correlates with behavioural change across time

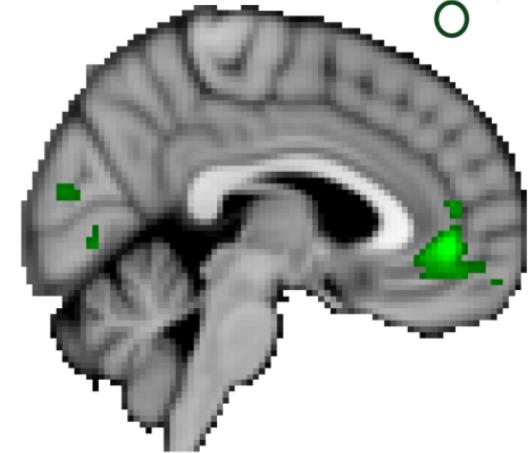
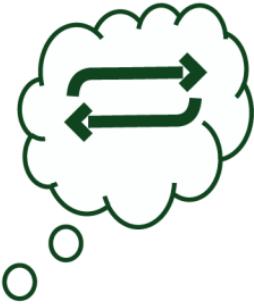
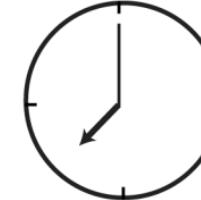




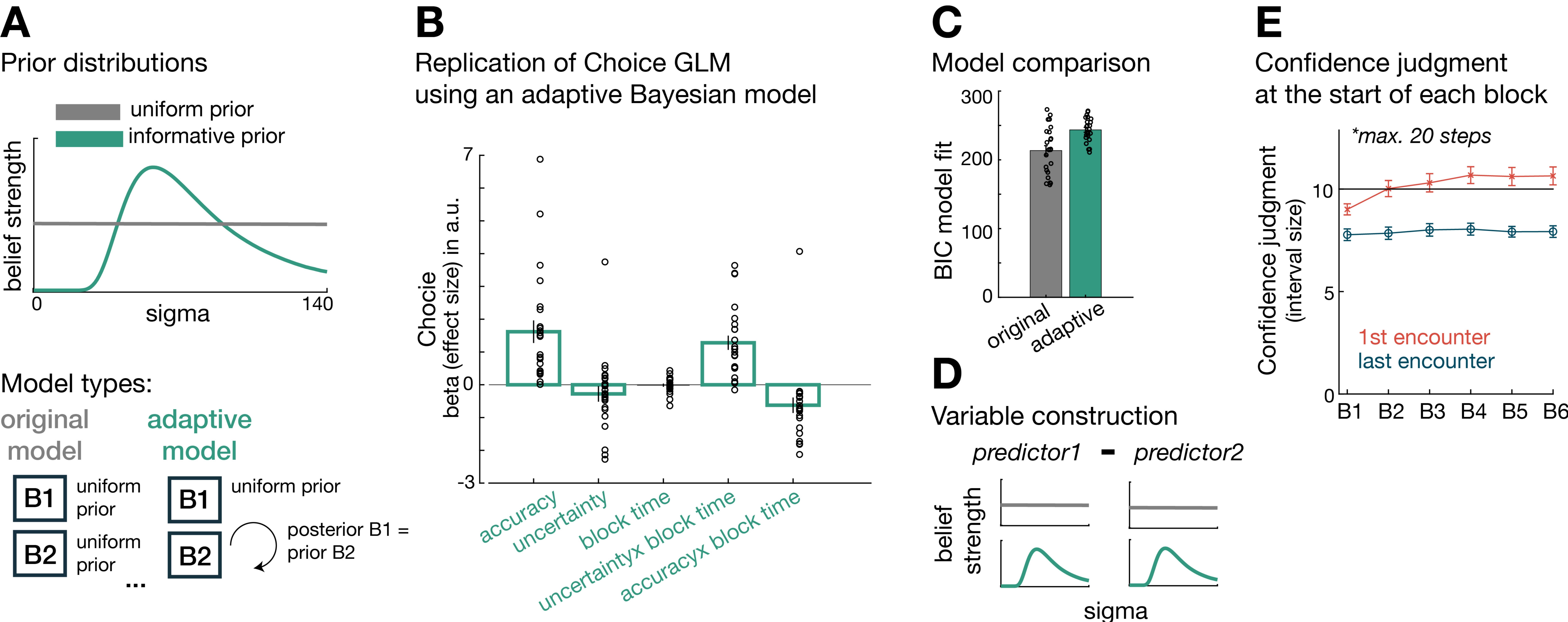
positive uncertainty
representation during
exploration



learning about predictors'
accuracies mediates
exploration-exploitation transition

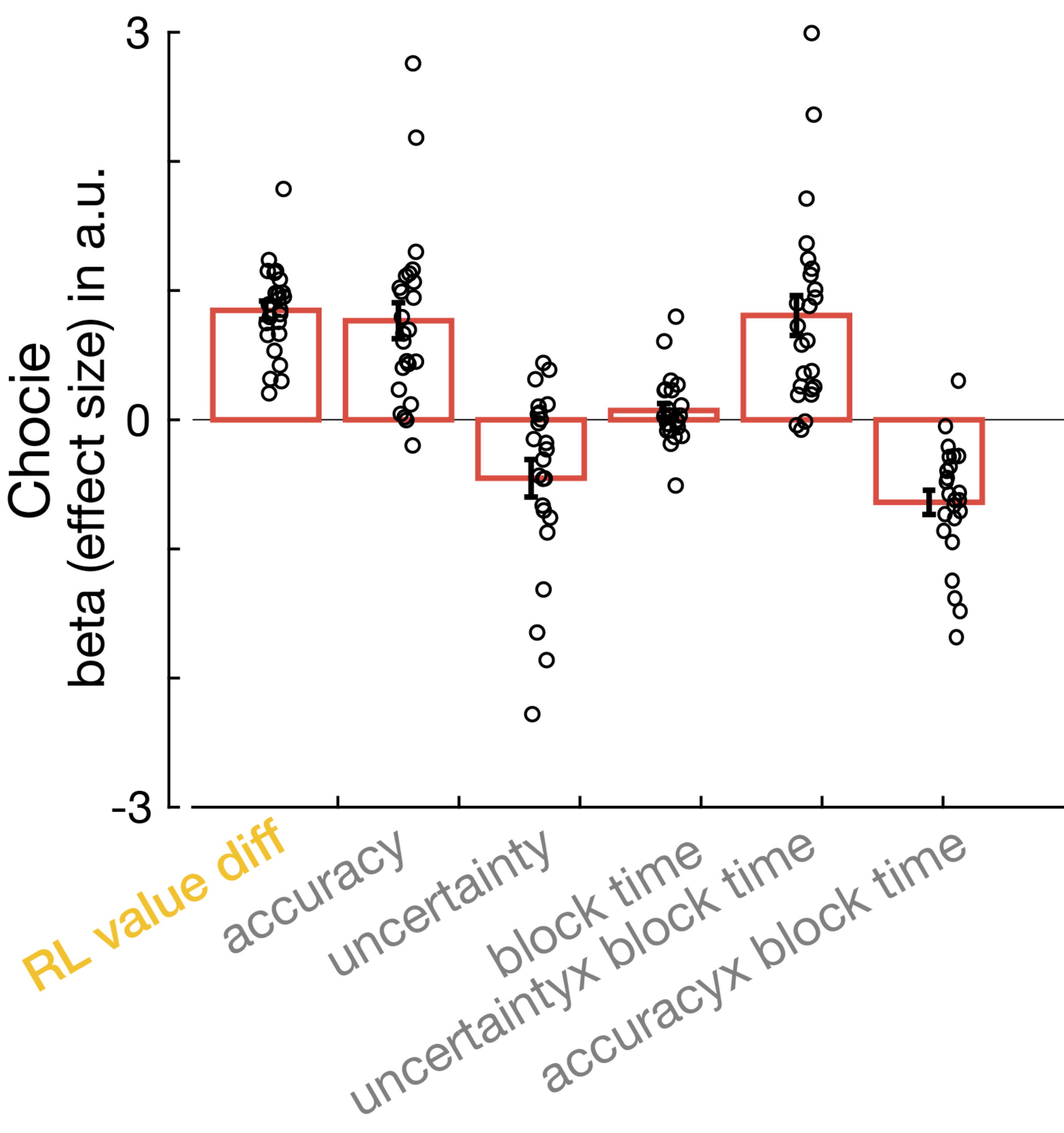


negative uncertainty
representation during
exploitation



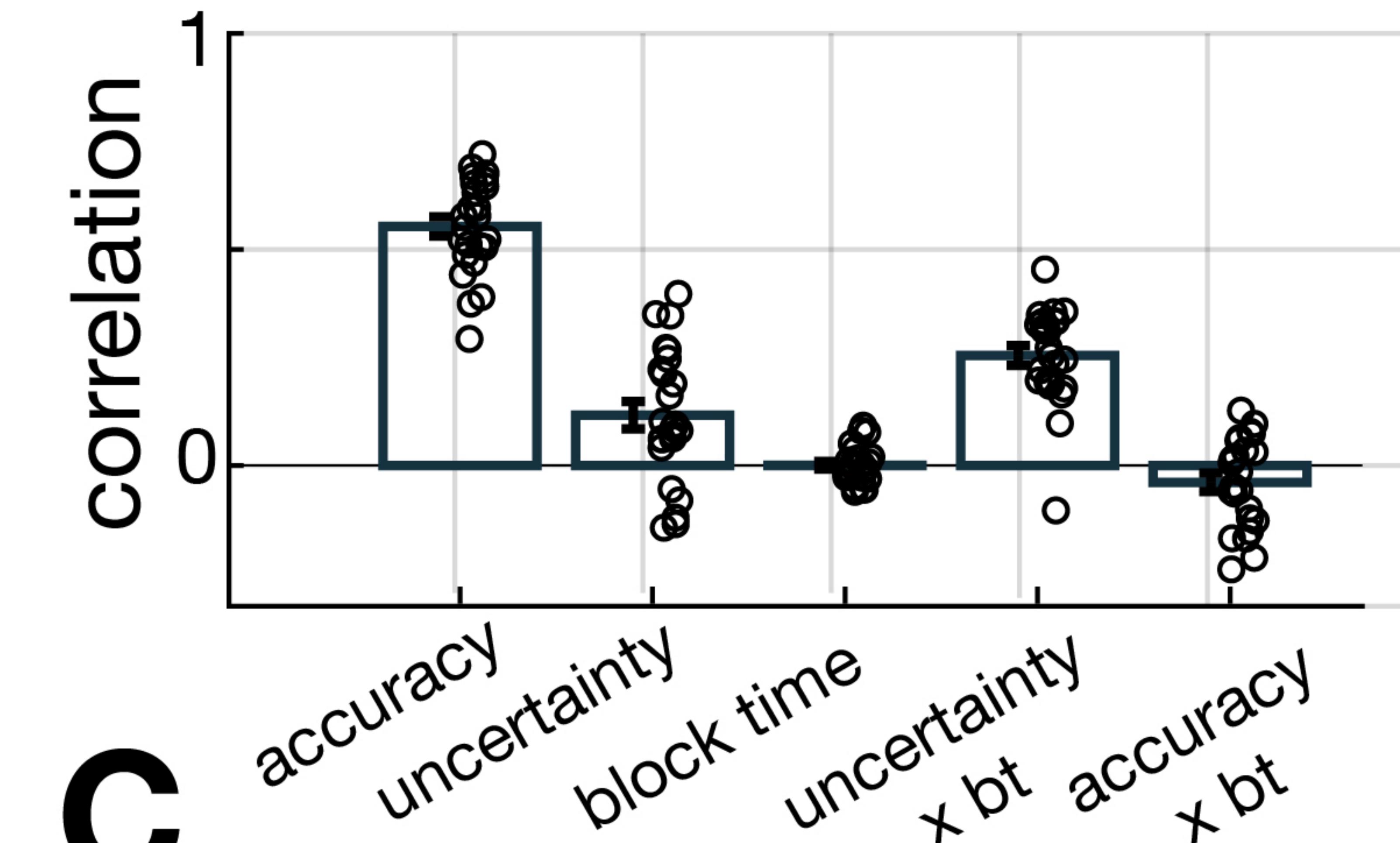
A

Replication of Choice GLM when controlling for RL value difference



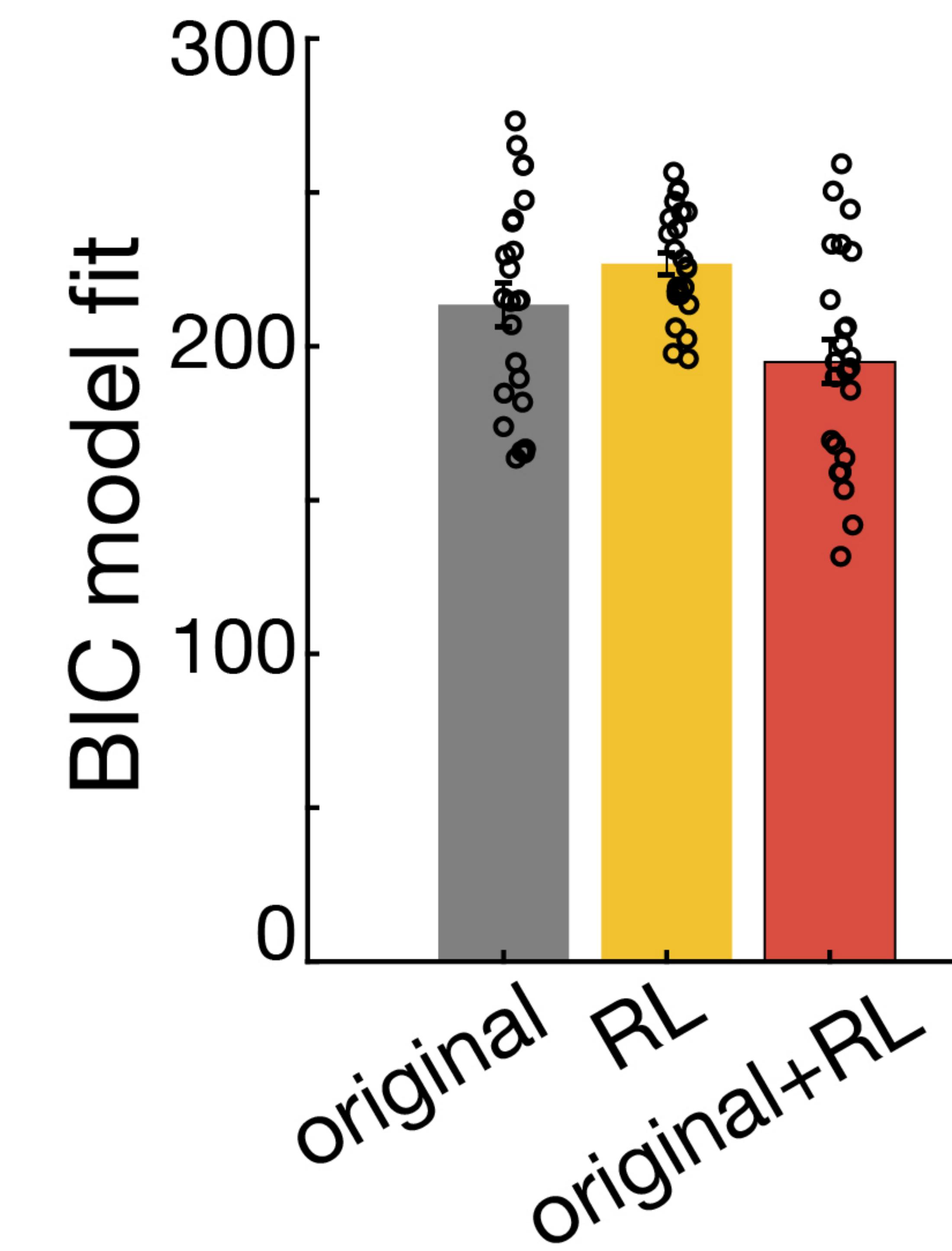
B

Correlation: RL value difference and ...



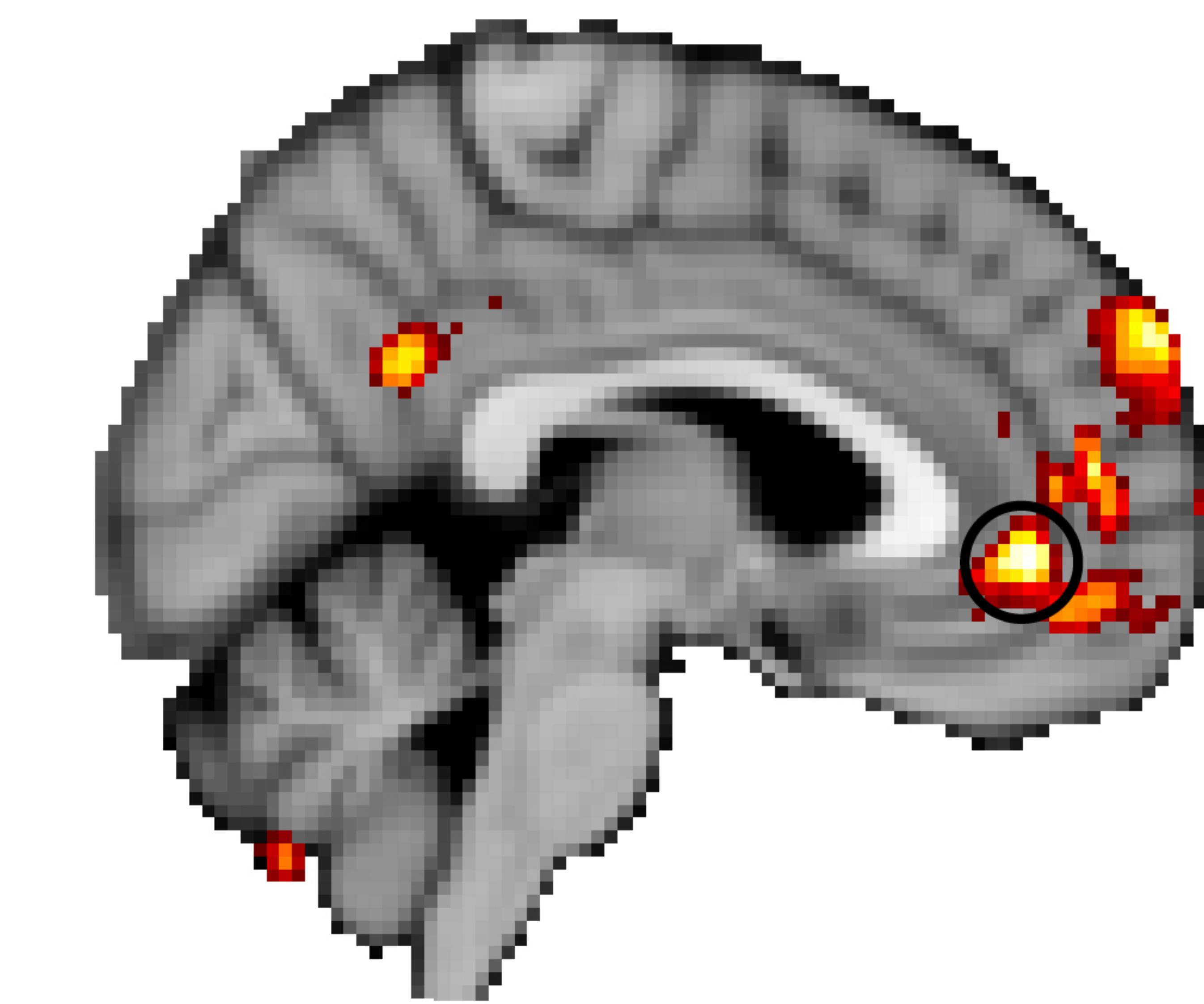
C

Model comparison

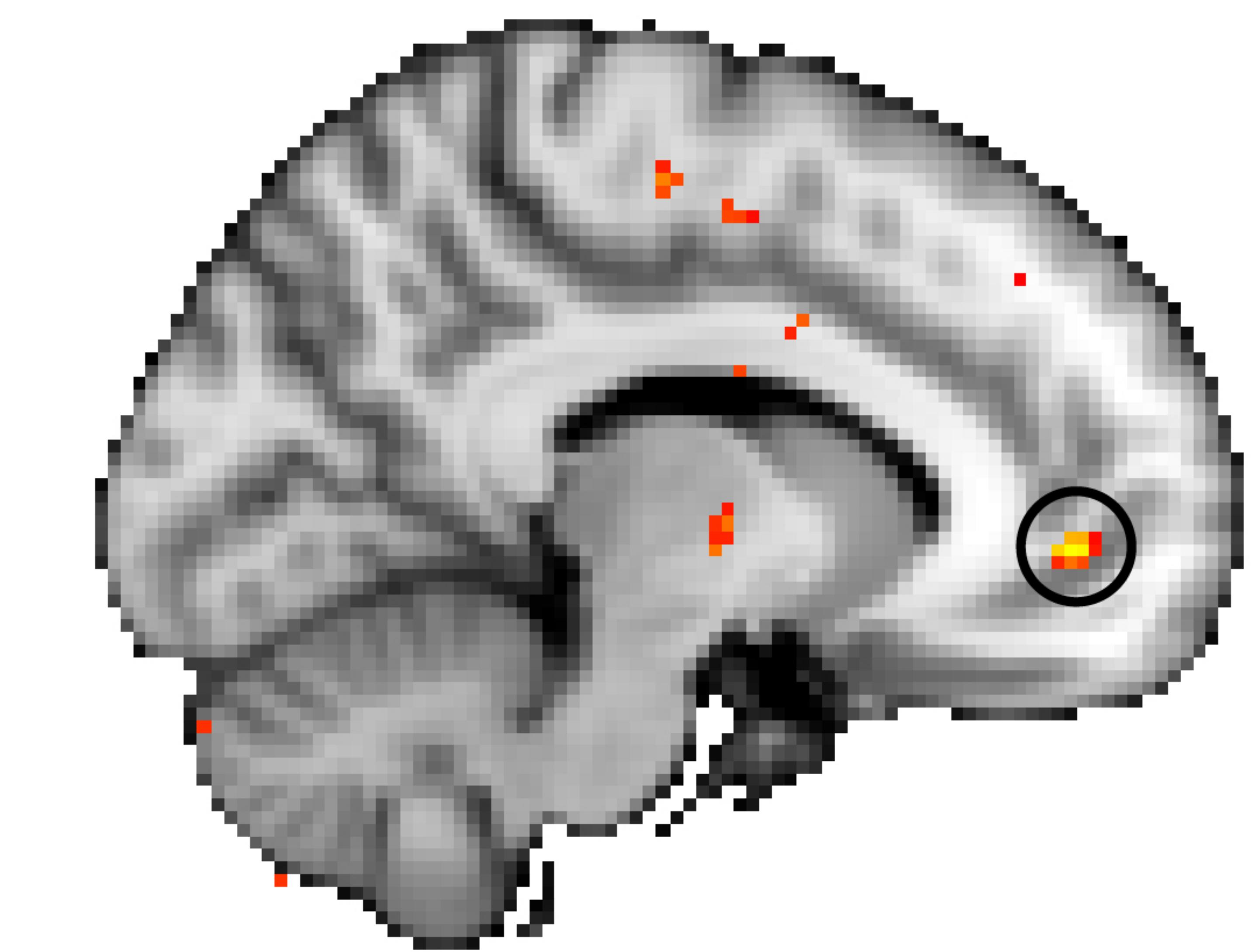


D

Replication of neural results

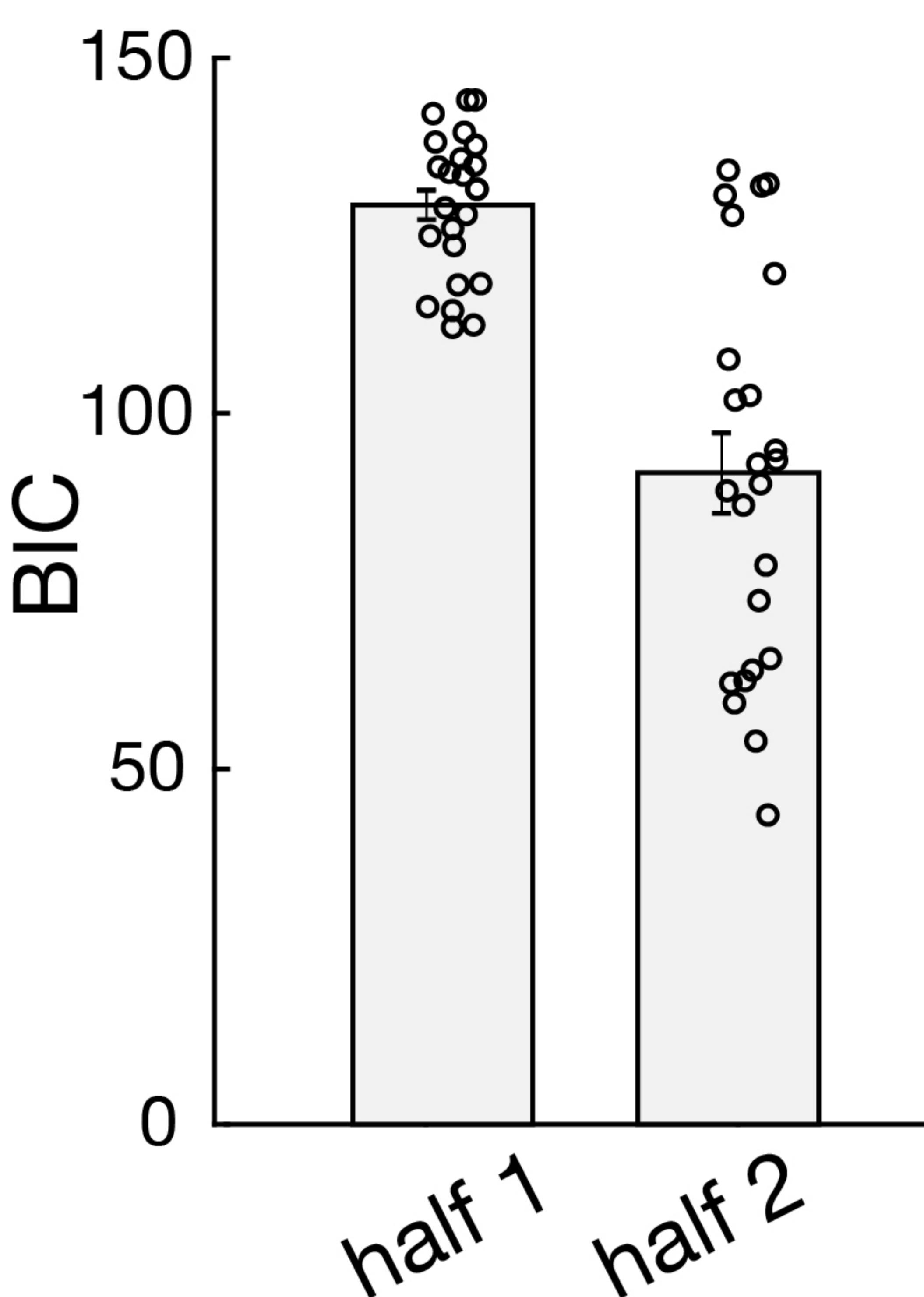


domain general prediction difference
* cluster corrected



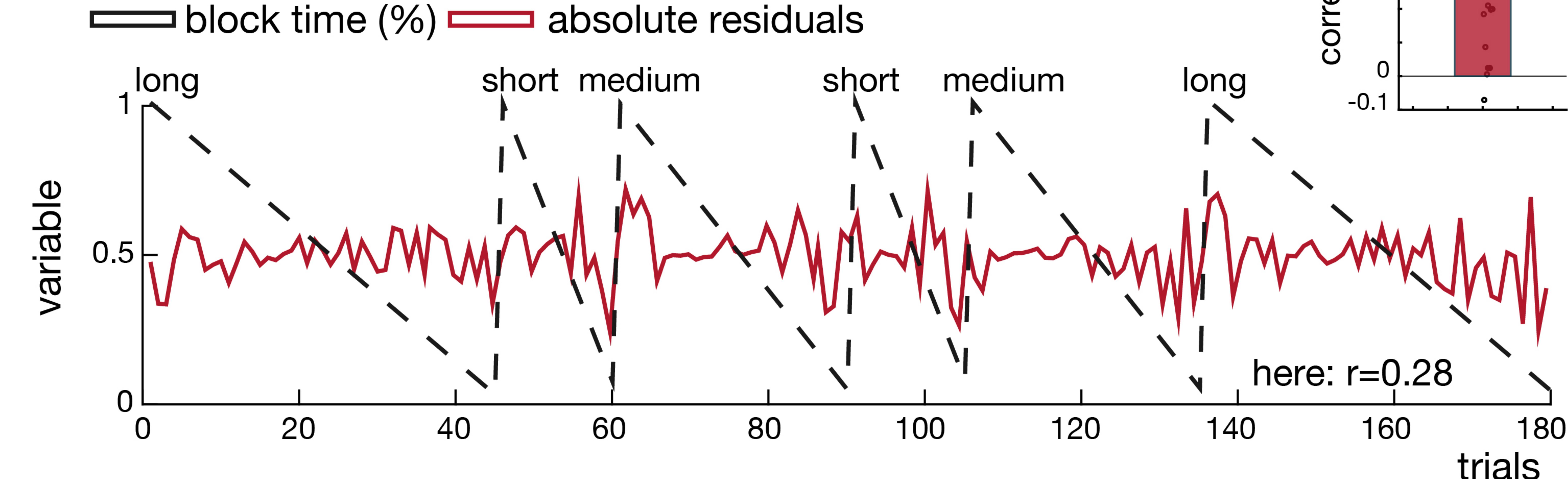
RL value difference
* sub-threshold

A Modelfit



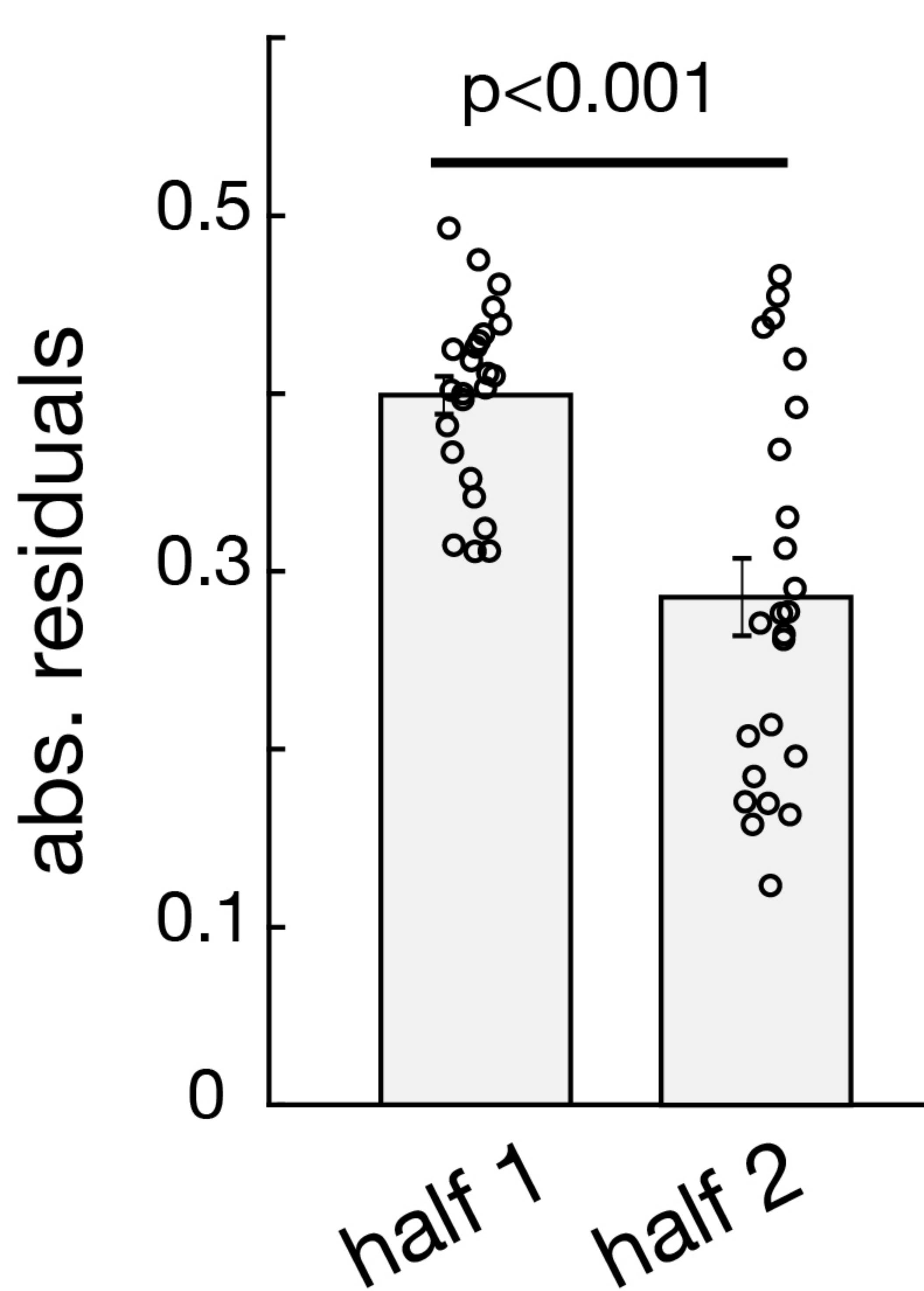
B Correlation between abs. residuals and block time

i) example participant

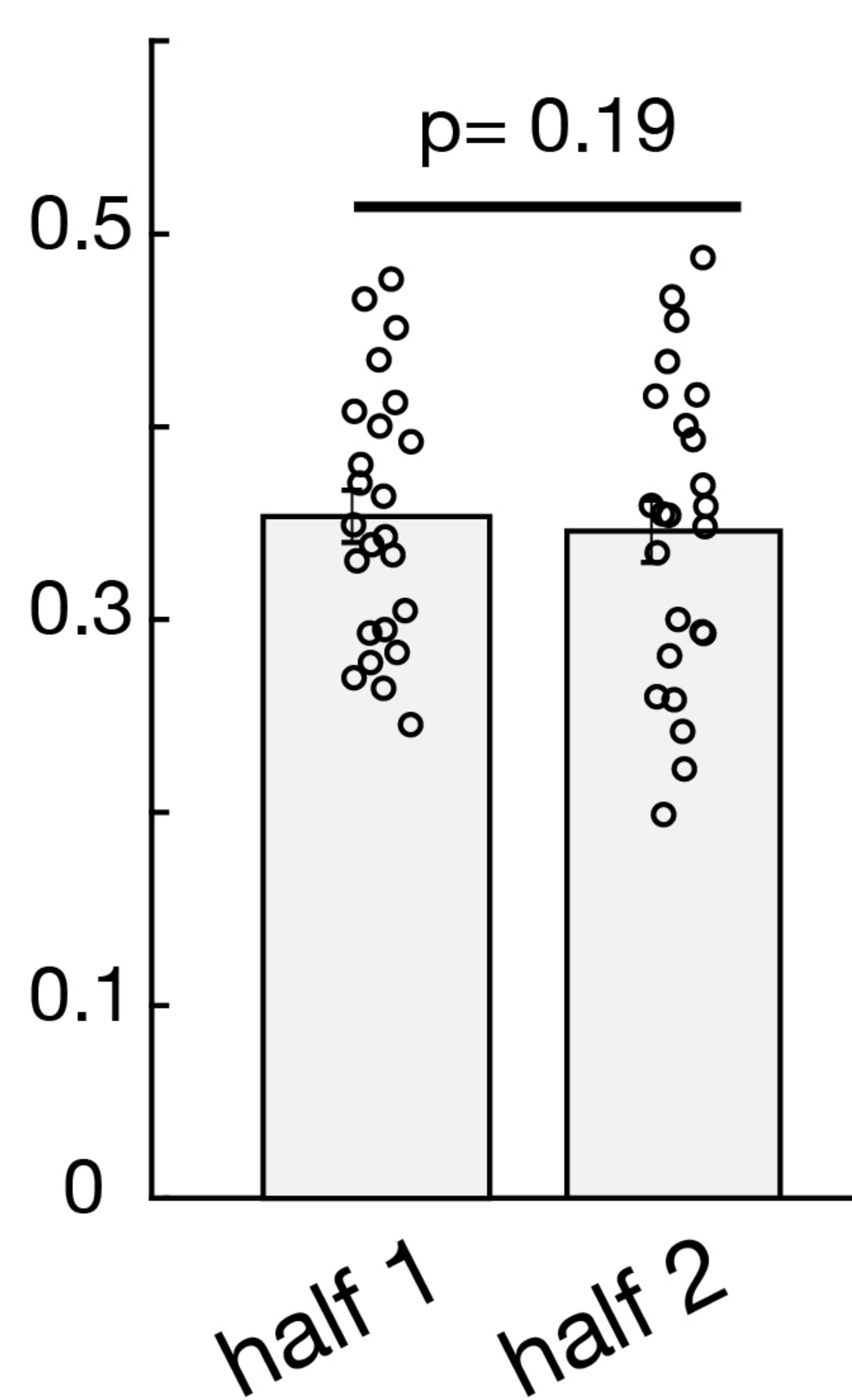


C Absolute residuals per block halves

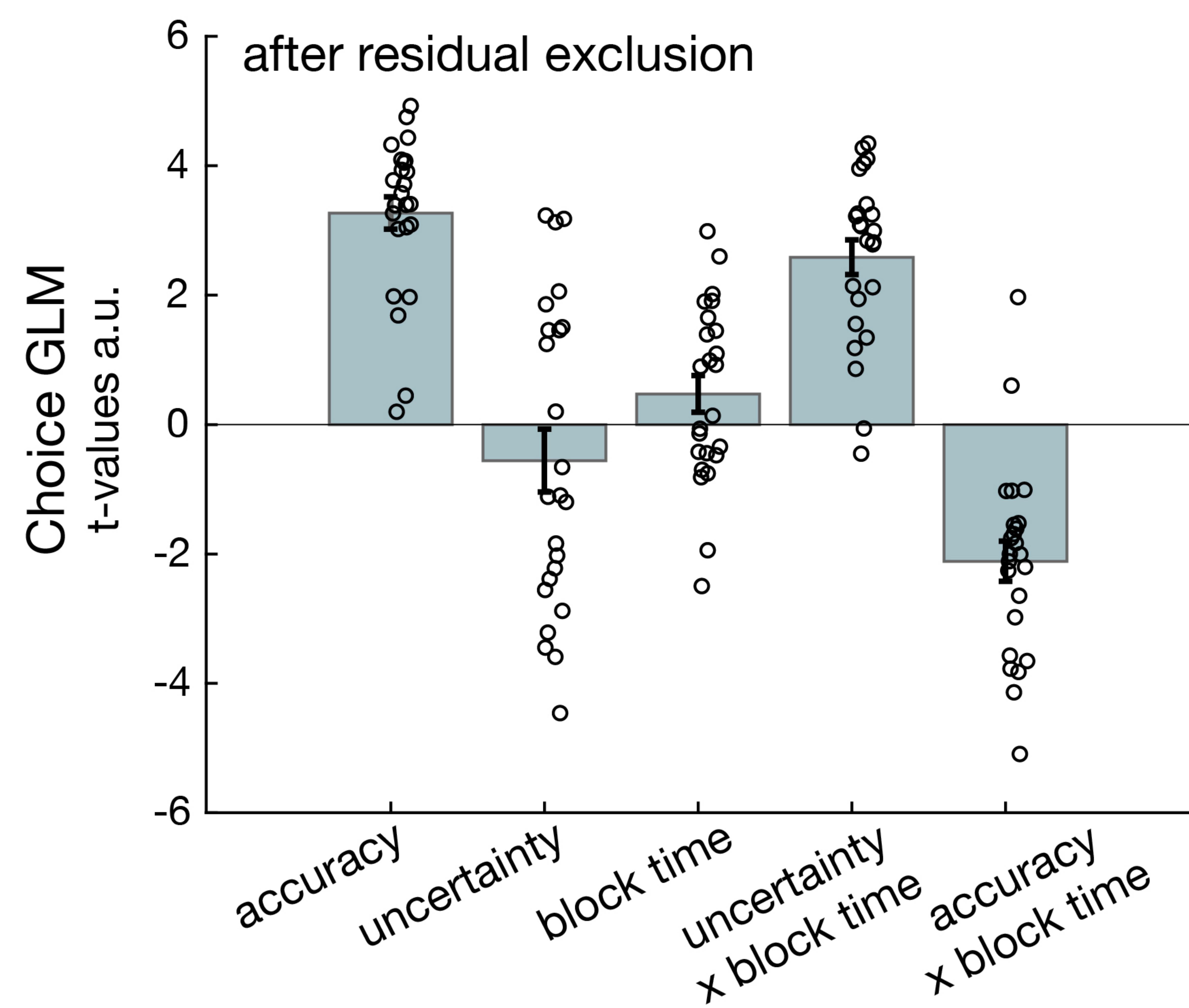
i) before residual exclusion



ii) after residual exclusion



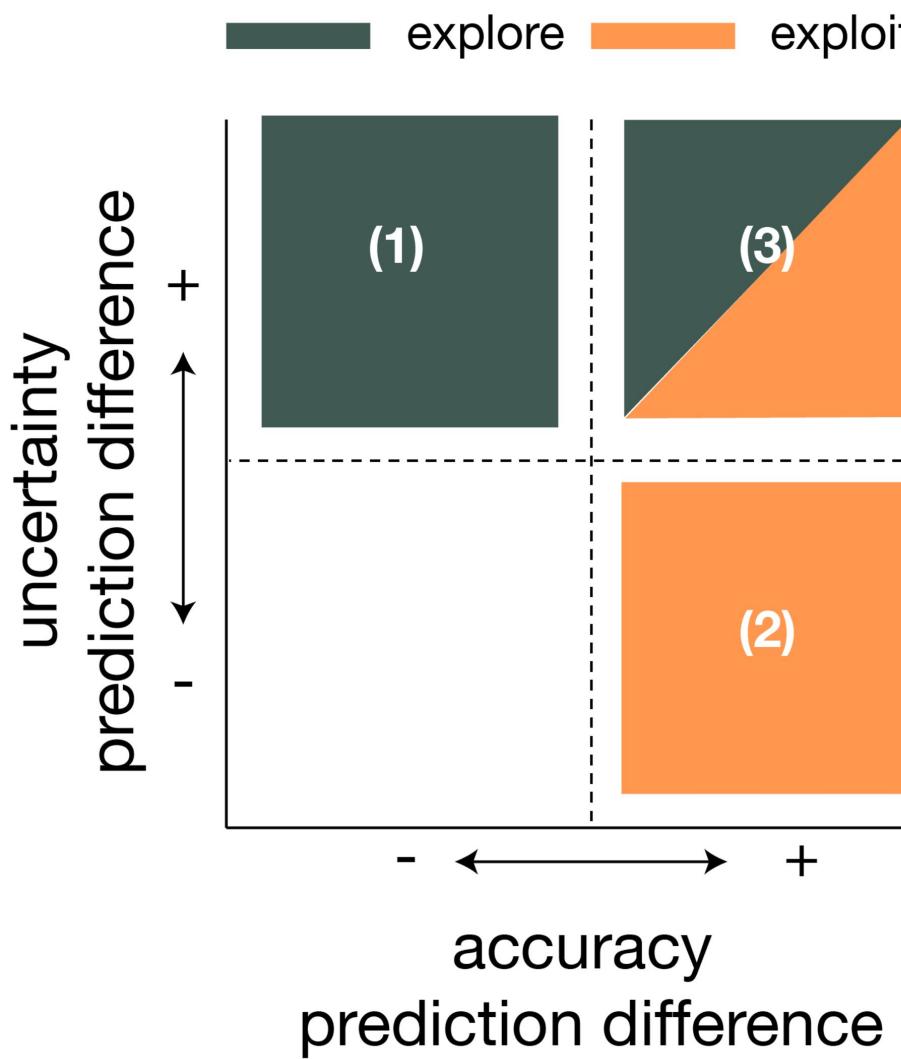
D GLM on new subset of trials



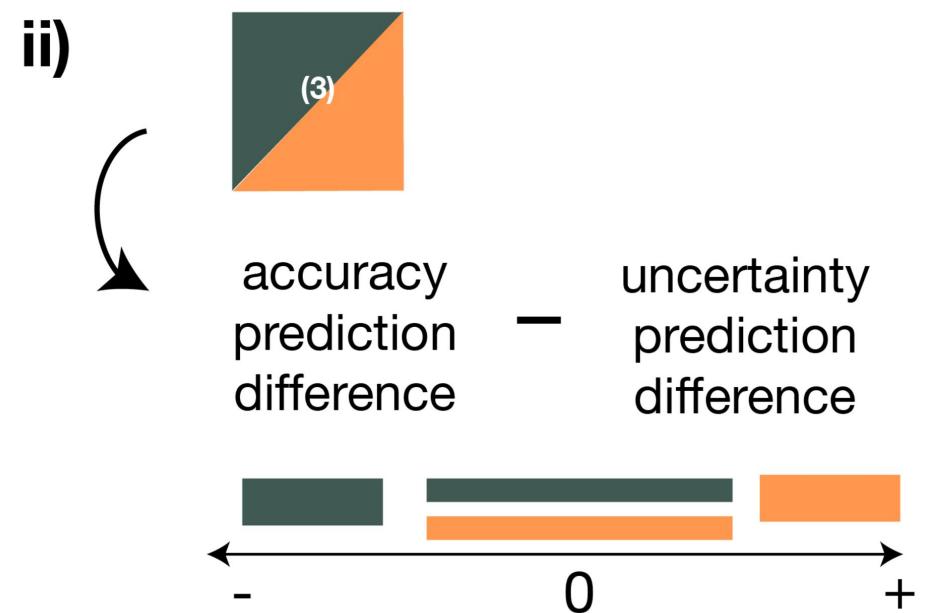
A

Trial separation

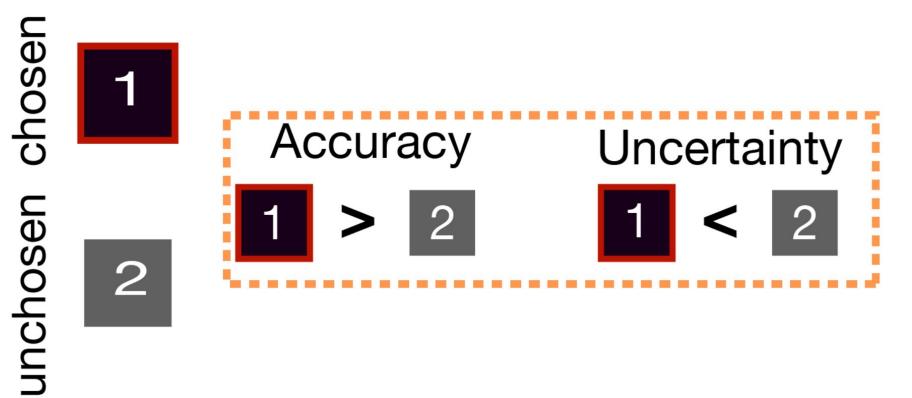
i) Explore vs exploit trials



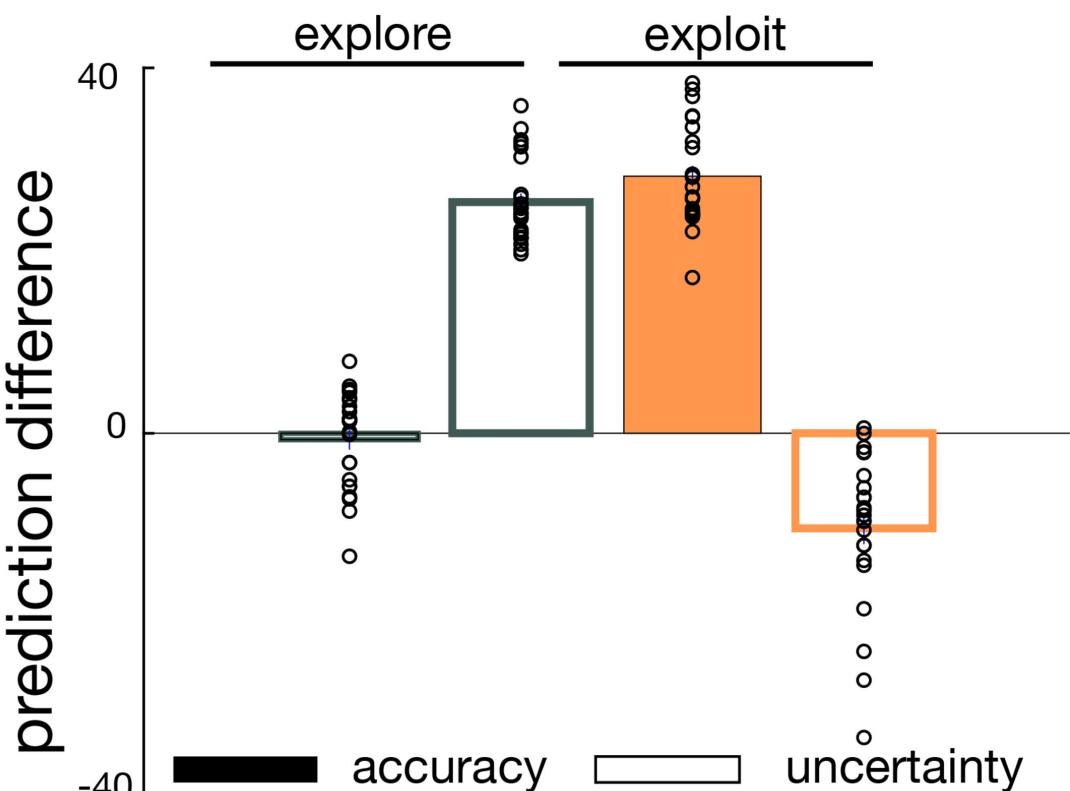
ii)



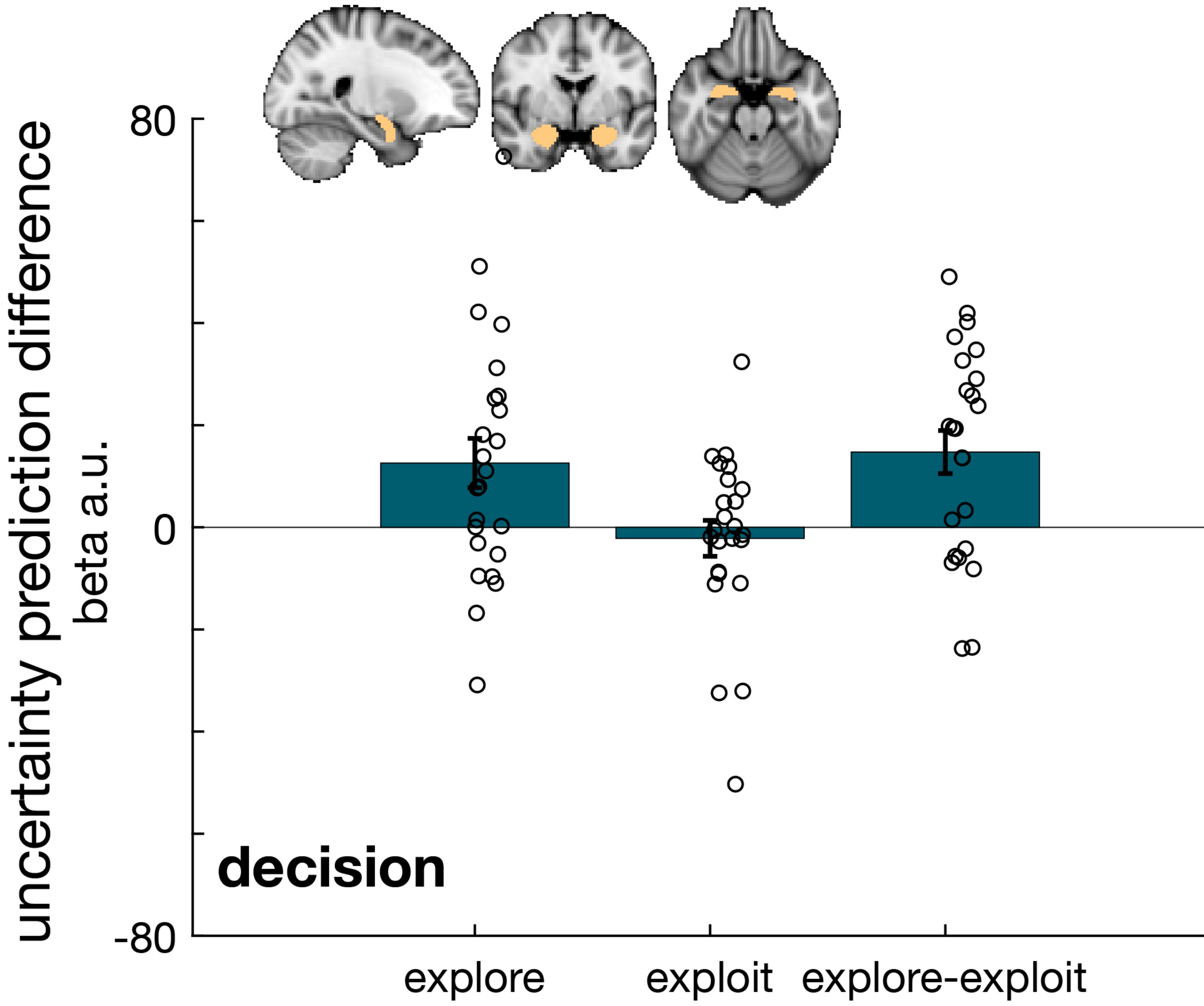
iii) Choice example

**B**

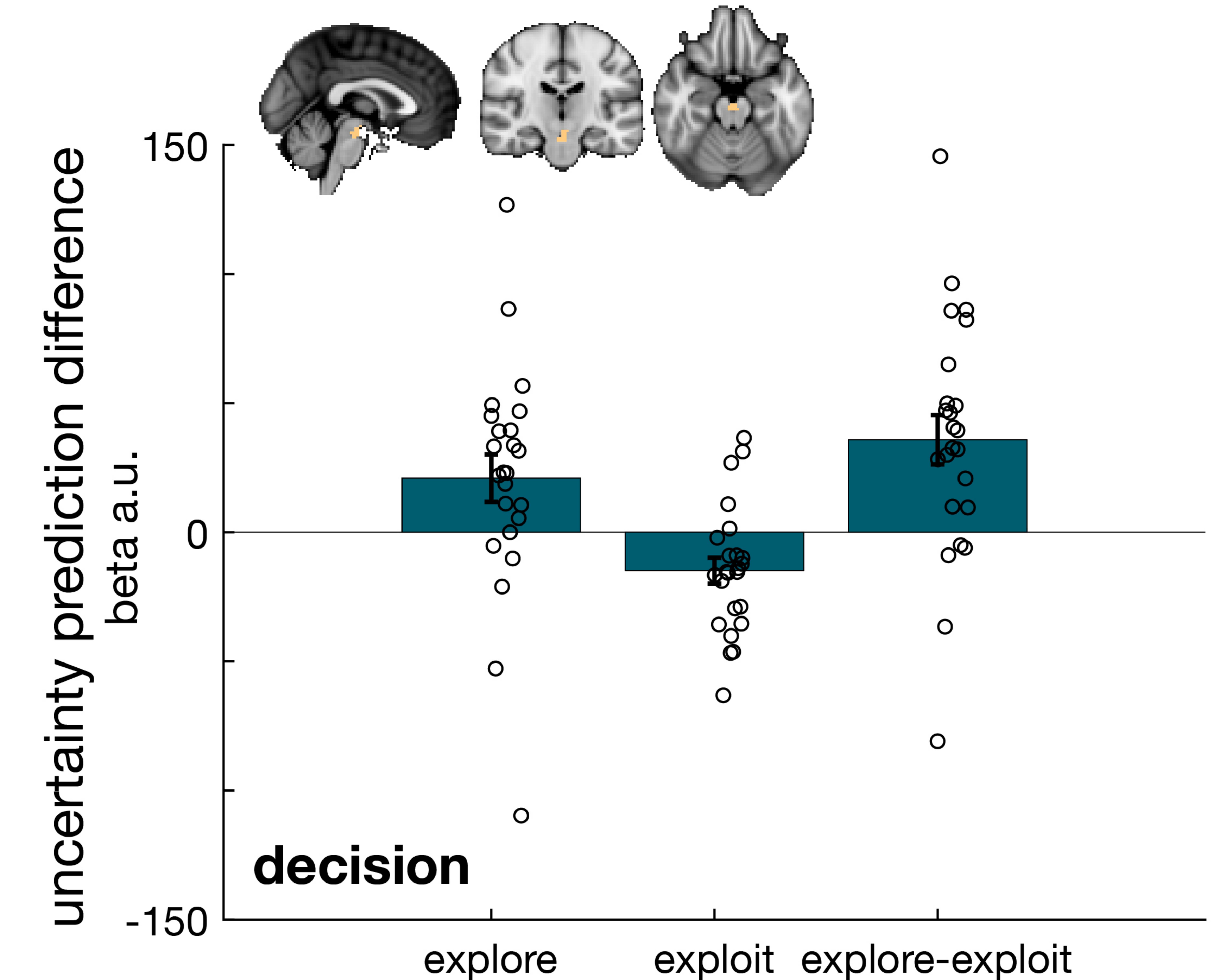
Manipulation check



A Amygdala



C Ventral tegmental area



B Ventral striatum

