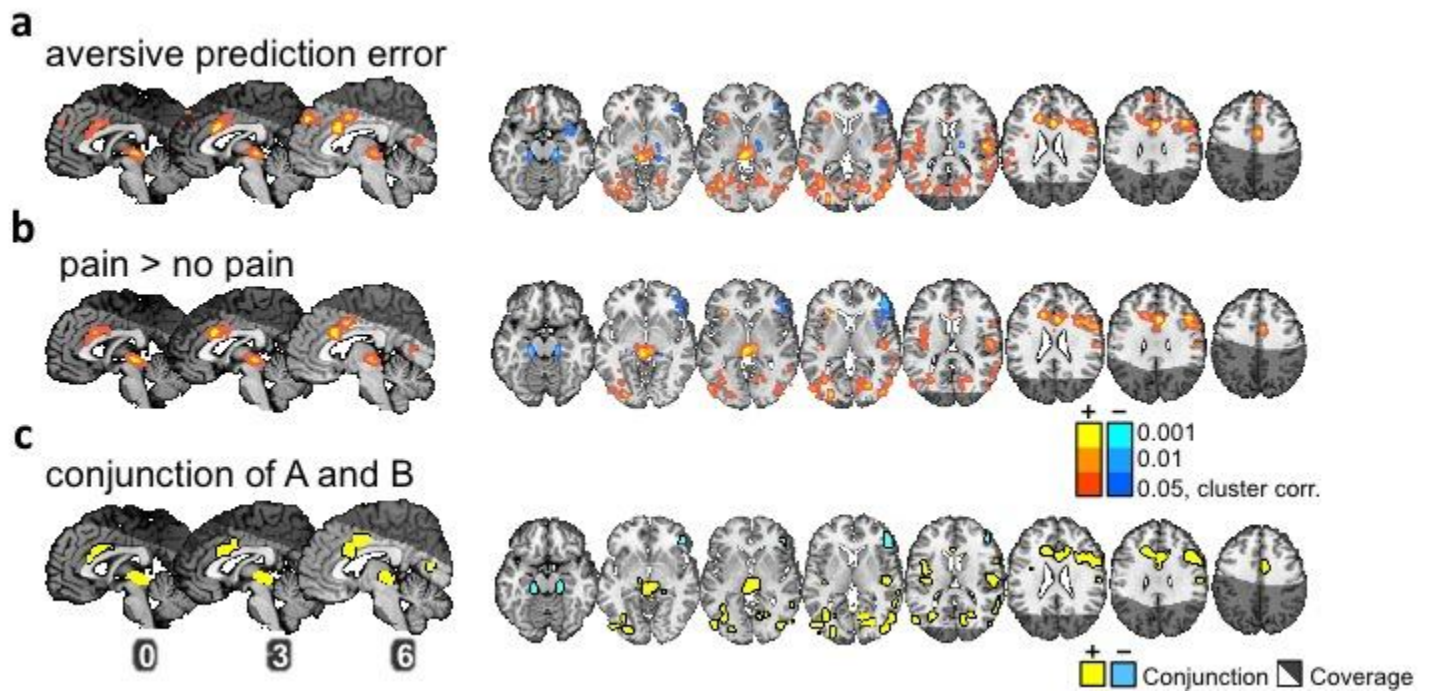


Supplementary Figure 1

Time course of subjective pain perception

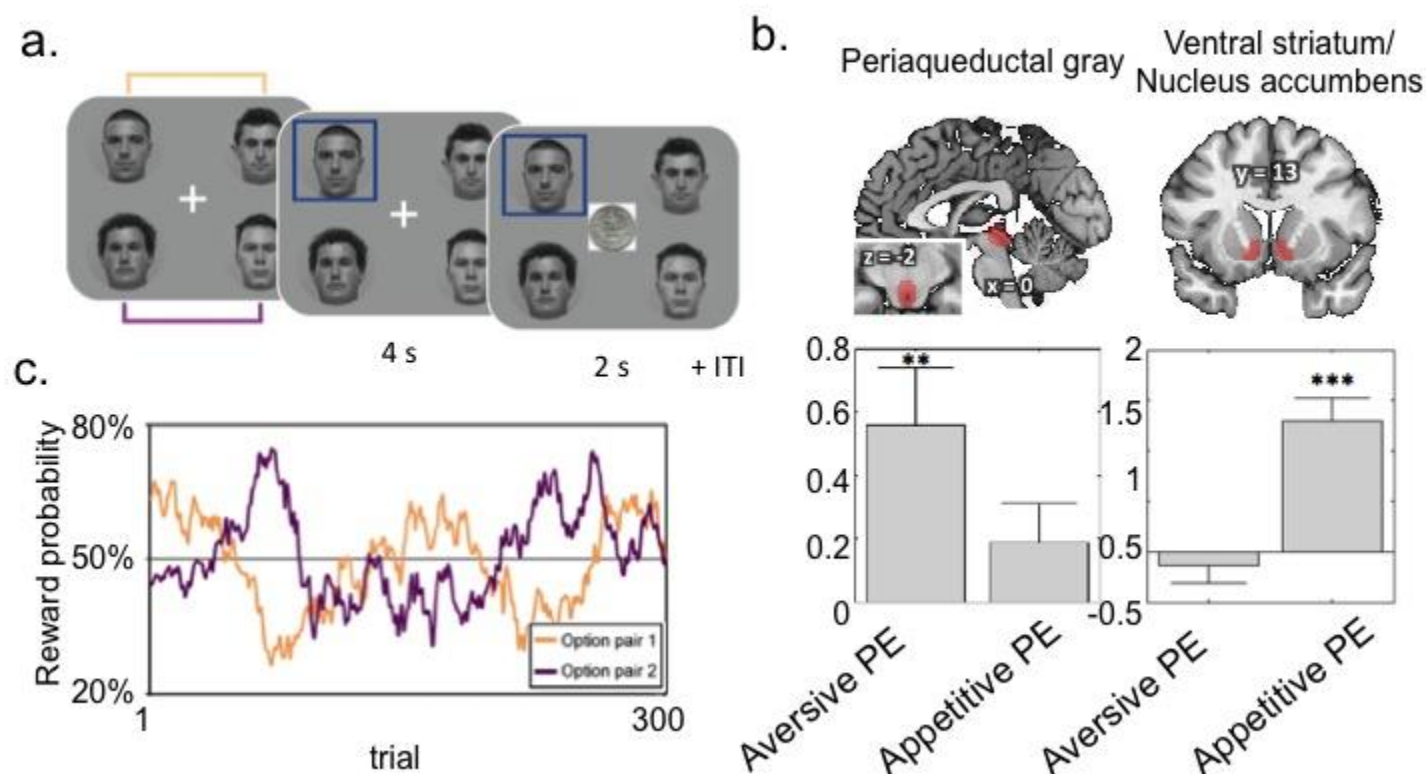
Mean online pain ratings obtained during the behavioral session superimposed on the temporal profile of the thermal stimulus (number of participants = 23; 4 seconds plateau, 2.5 seconds ramp-up/ramp-down). Ratings begin to rise in the first second of the stimulation (left of the red vertical line).



Supplementary Figure 2

Activity at outcome onset

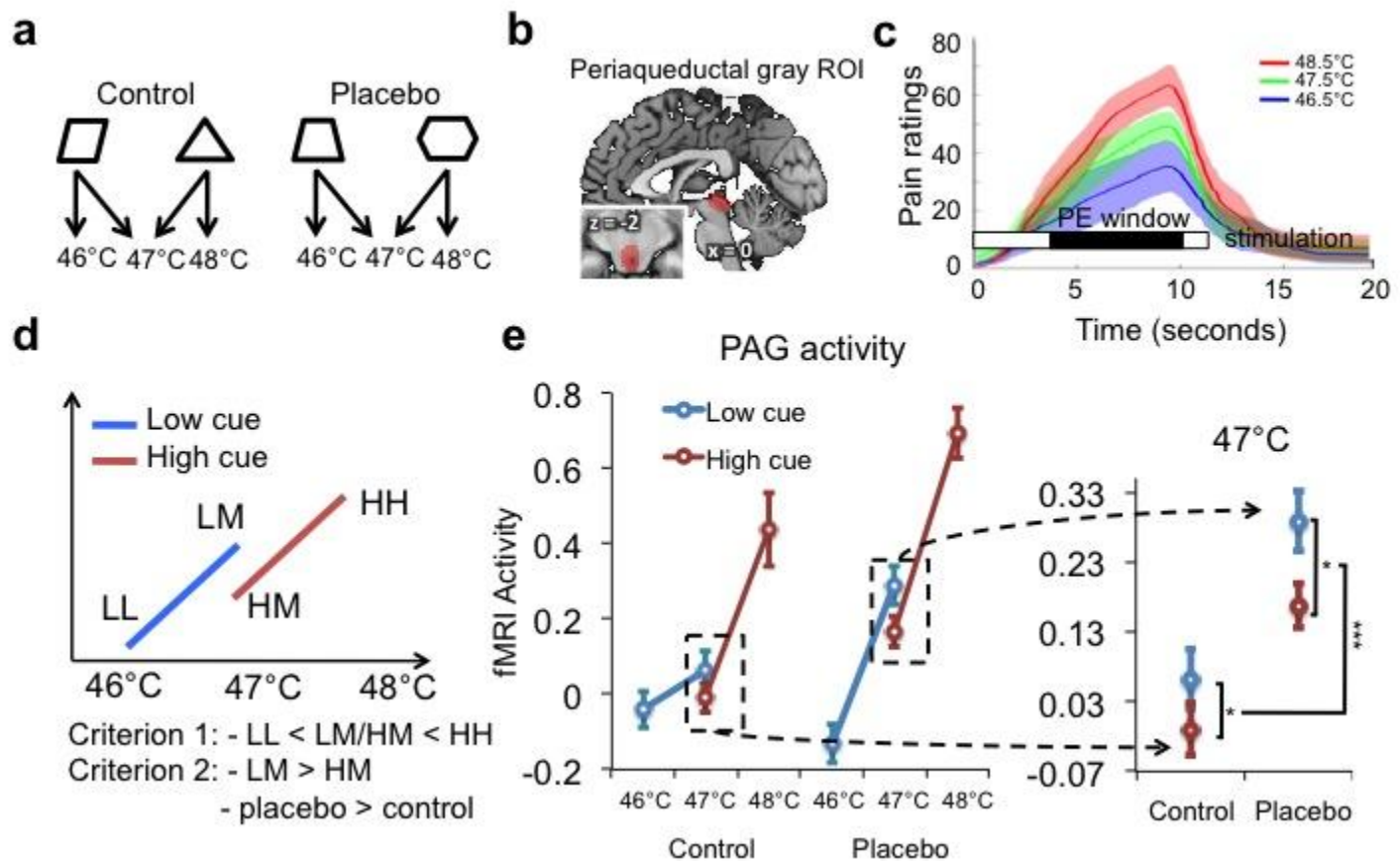
Activity at outcome onset (1st second of outcome; number of participants = 23) related to (A) model-based aversive prediction error (outcome worse than expected), (B) pain > no stimulus. Displayed activations are cluster-thresholded ($p < 0.05$, FWE, two-tailed) with cluster-defining thresholds of $p < 0.001$, $p < 0.01$ and $p < 0.05$. (C) Conjunction of model-based prediction error and pain effects. Conjunctions of positive/negative effects are in yellow/blue.



Supplementary Figure 3

Study 2: comparison with reward prediction errors

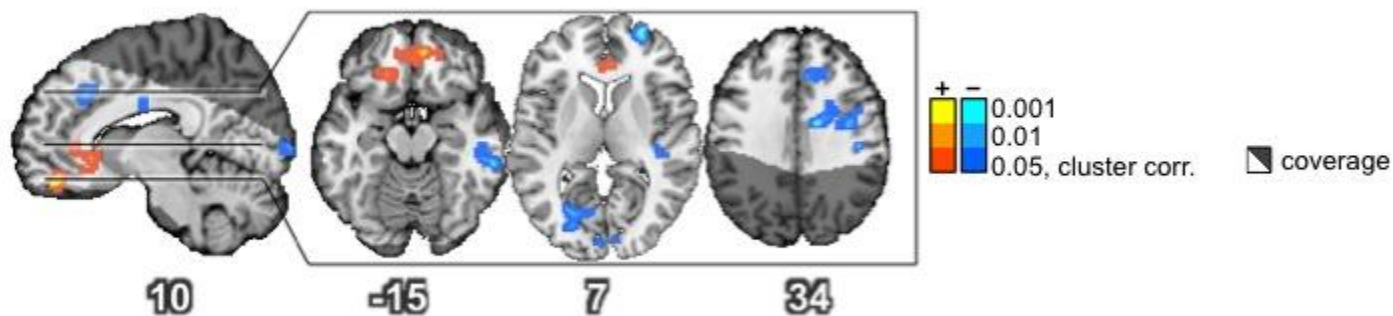
A) On each trial, participants ($n = 24$) chose one of four face options. After a delay, the outcome (\$0.25 or \$0.00) was revealed. Faces are paired together such that the probability of receiving a reward on a given trial is the same for both faces of the pair. In colored brackets, one example of option pairing is indicated (reproduced from Wimmer et al., 2012). B) Comparison of pain aversive prediction errors (PE) and monetary appetitive PE in periaqueductal gray (PAG) and nucleus accumbens (Nacc) regions of interest (see figure 2 in main article; PAG-aversive: $t(22) = 3.07$, $p = 0.006$; PAG-appetitive: $t(20) = 1.54$, $p = 0.14$; NAcc-aversive = $t(22) = -0.80$, $p = 0.44$; NAcc-appetitive = $t(20) = 5.77$, $p < 0.001$). C) Drifting reward probability distribution defining the reward equivalence for one example pairing (reproduced from Wimmer et al., 2012). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. Error bars represent standard errors of the mean.



Supplementary Figure 4

Study 2: comparison of different pain levels

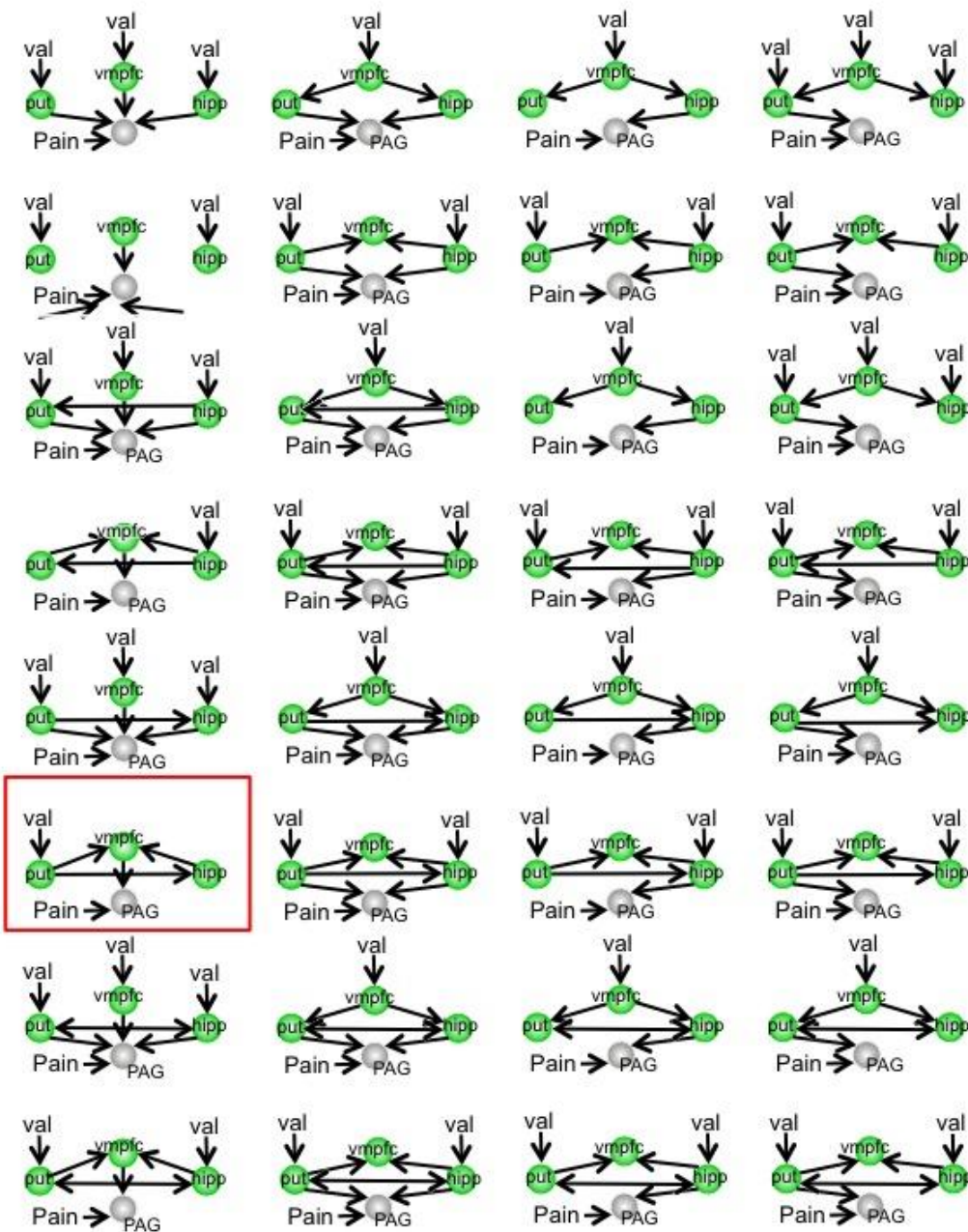
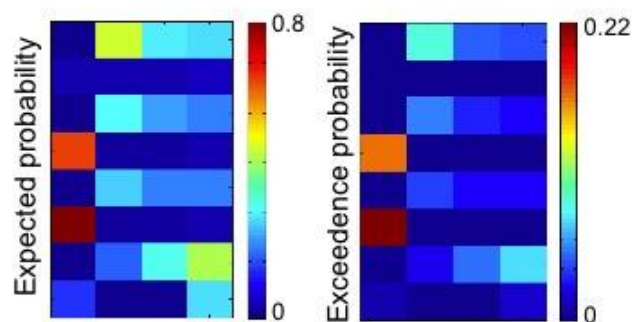
A) Experimental conditions. In both the control and placebo runs, participants ($n = 50$) are presented two predictive cues. The low cue is followed 50% of the time by low pain (46°C) and 50% of the time by medium pain (47°C). The high cue is followed 50% of the time by high pain (48°C) and 50% of the time by medium pain (47°C). In placebo runs the thermode is installed on a skin spot pre-treated with a cream participants are told has analgesic properties. B) Periaqueductal gray (PAG) region of interest (ROI). C) Continuous pain ratings from 30 independent subjects for 11-s thermal stimulations at 46.5°C, 47.5°C and 48.5°C. The window of analysis for aversive prediction error signals was set between 4 and 10 seconds, i.e. between the time the temperatures can be differentiated and the peak of pain. D) Axiomatic predictions for aversive prediction error. Axiom #1 stipulates that aversive prediction error signals should increase with temperature intensity, regardless of expectations. Axiom #2 stipulates that lower pain expectations should be associated with higher prediction errors, regardless of temperature. Therefore, axiom #2 can only be tested on the medium temperature. Moreover, if the same prediction error signals are also influenced by instruction-based expectations, we should observe higher activity for the placebo vs. control condition. E) Activity in the PAG during the PE window. The left panel shows a clear effect of temperature (low < medium, medium < high, all p 's < 0.001). The right panel shows effects of cues and condition for stimulations at 47°C, which are in conformity with axioms #2 and 3. Activity in the PAG ROI for medium pain stimulations is higher for low vs. high pain cues ($F(1,49) = 4.39$, $p < 0.05$) and for the placebo analgesia condition vs. control ($F(1,49) = 16.03$, $p < 0.001$). Error bars represent standard errors of the mean.



Supplementary Figure 5

Activity during decision-making

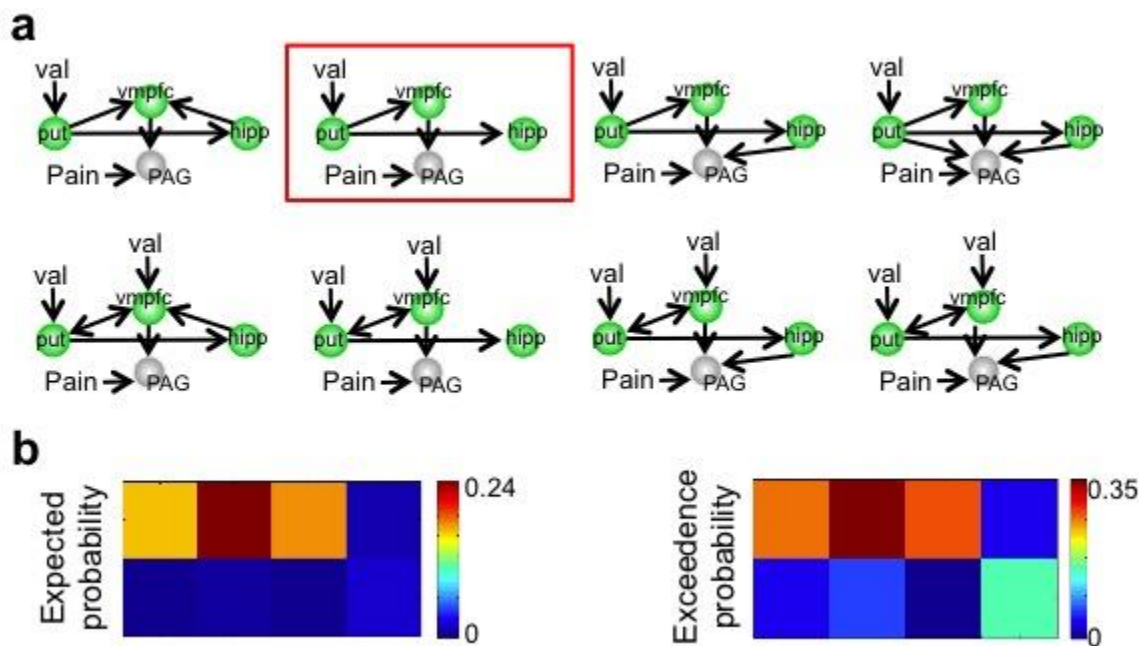
Activity during decision-making (number of participants = 23) correlating positively (red) or negatively (blue) with the expected value of the chosen option (warm/cold colors indicate low/high subjective (model-based) probability of pain. Displayed activations are cluster-thresholded ($p < 0.05$, FWE, two-tailed) with cluster-defining thresholds of $p < 0.001$, $p < 0.01$ and $p < 0.05$.

a**b**

Supplementary Figure 6

DCM optimizing the connectivity of the aversive prediction error structure: step 1a

(A) In all models, the driving inputs are the pain > no stimulus and expected value parametric modulators on outcome onsets. The models tested systematically varied the structure(s) receiving the expected value driving inputs and conveying this information to the midbrain. The model with the highest exceedance probability is highlighted in red. (B) Expected (expected posterior probability) and exceedance (probability compared with other tested models) probabilities associated with each model. Val = expected value, str = striatum, hipp = hippocampus, mb = midbrain. number of participants = 23.

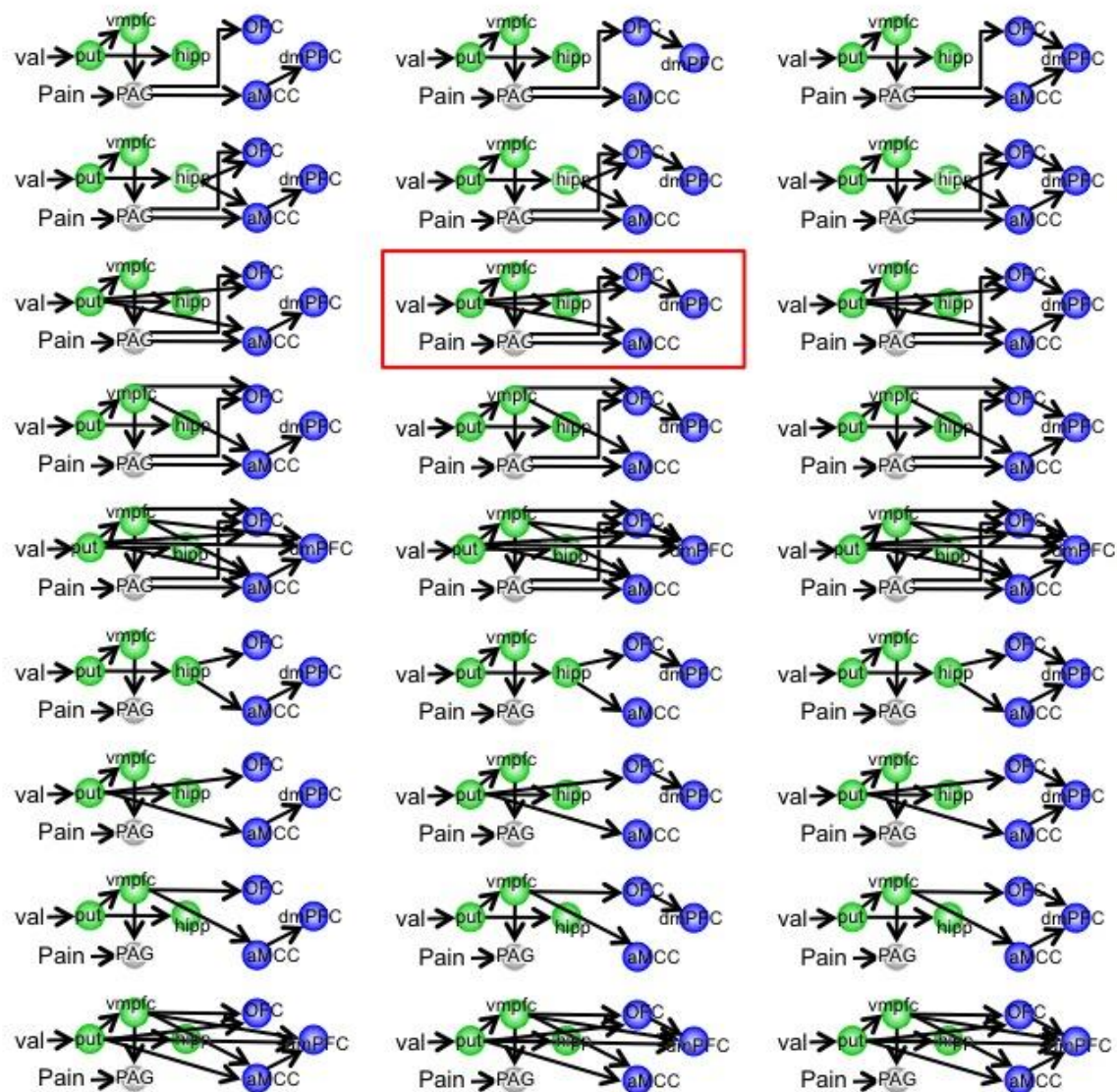


Supplementary Figure 7

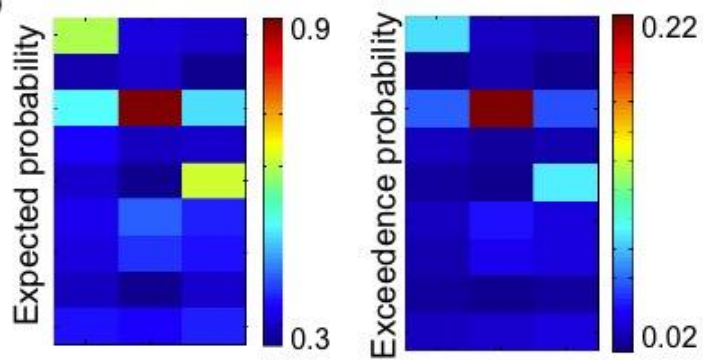
DCM optimizing the connectivity of the aversive prediction error structure: step 1b

(A) In all models, the driving inputs are the pain > no stimulus and expected value parametric modulators on outcome onsets. The model selected in the previous step is the first one (top left). From left to right, the tested models varied hippocampus targets (vmPFC, PAG, or nothing), or added a link between the striatum and midbrain. Models in the second row additionally include expected value as a driving input to the vmPFC, and a link from the vmPFC to the striatum. (B) Expected (expected posterior probability) and exceedance (probability compared with other tested models) probabilities associated with each model. Val = expected value, put = putamen, hipp = hippocampus, PAG = periaqueductal gray. number of participants = 23.

a



b

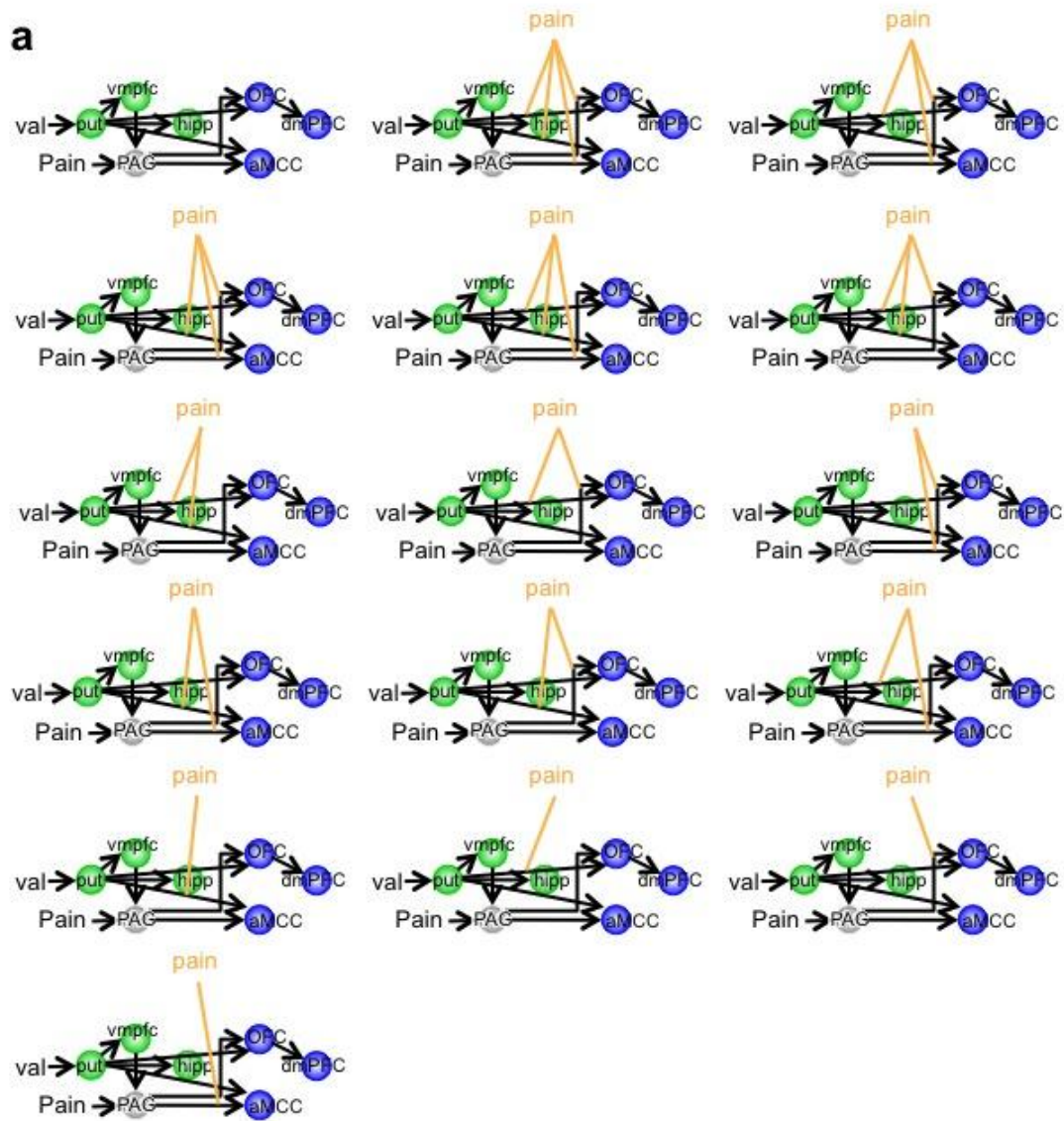


Supplementary Figure 8

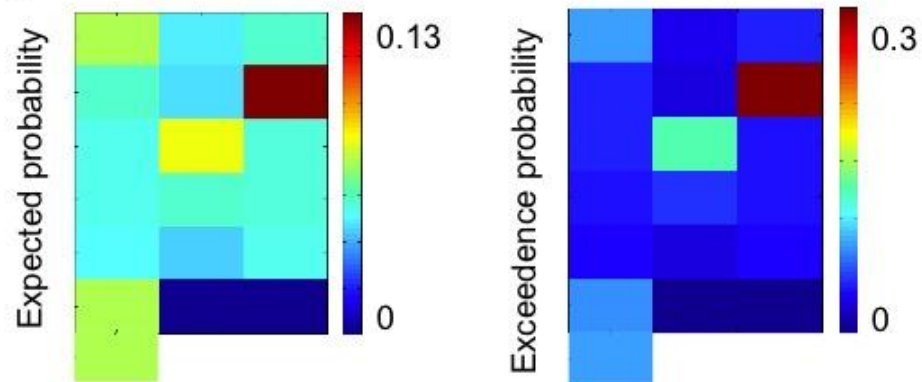
DCM optimizing the connectivity of the aversive prediction error structure: step 2a.

(A) In all models, the structures generating PE signals (put, vmPFC, hipp, PAG) are arranged according to the best model selected from the previous model selection steps. The links between these structures and the pain-specific PE structures are systematically varied. The model with the highest exceedance probability is highlighted in red. (B) Expected (expected posterior probability) and exceedance (probability compared with other tested models) probabilities associated with each model. Val = expected value, put = putamen, hipp = hippocampus, PAG = periaqueductal gray, OFC = orbitofrontal cortex, aMCC = anterior cingulate cortex, dmPFC = dorsomedial prefrontal cortex. number of participants = 23.

a



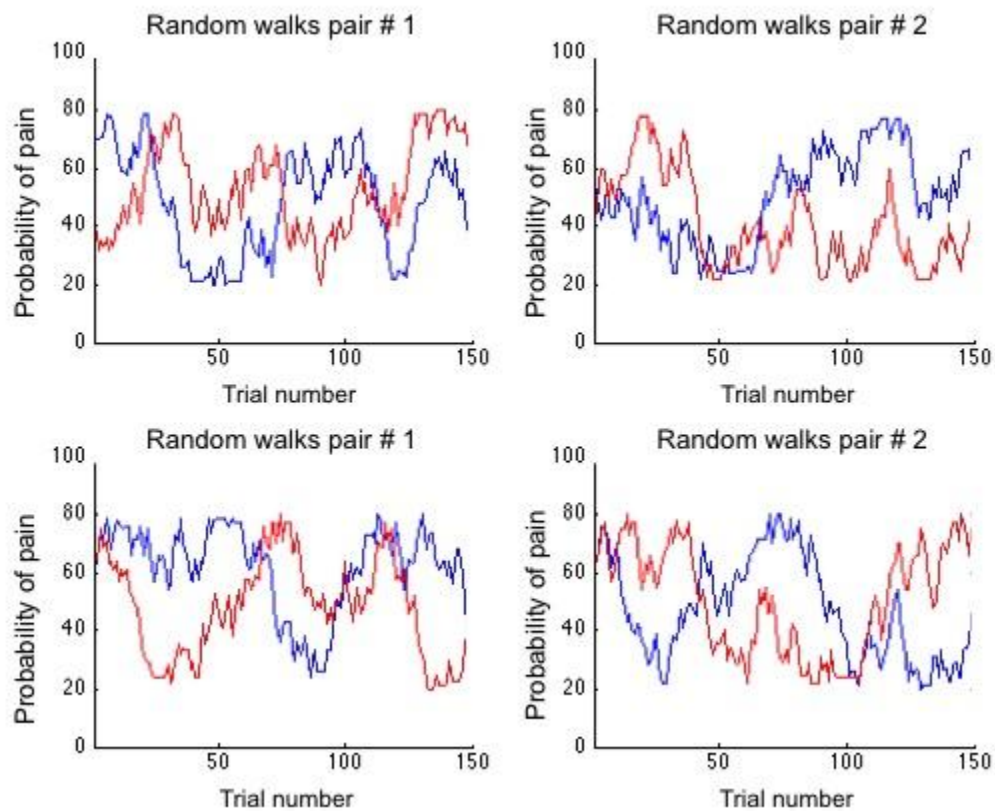
b



Supplementary Figure 9

DCM optimizing the connectivity of the aversive prediction error structure: step 2b.

(A) Modulatory influences were systematically added to the connections from the striatum or midbrain to the OFC or aMCC. (B) Expected (expected posterior probability) and exceedance (probability compared with other tested models) probabilities associated with each model. Val = expected value, str = striatum, hipp = hippocampus, mb = midbrain, OFC = orbitofrontal cortex, aMCC = anterior cingulate cortex, dmPFC = dorsomedial prefrontal cortex.



Supplementary Figure 10

Probabilities associated with each option

The four sets of random walks used in the current study. Probabilities associated with each option (blue and red lines) varied independently and slowly from trial-to-trial according to random walks. Probabilities were bounded 20% and 80%, and had to cross at least once over the course of the experiment.