nature neuroscience

Representation of aversive prediction errors in the human periaqueductal gray

Mathieu Roy[1,2], Daphna Shohamy[3], Nathaniel Daw[4], Marieke Jepma[1], G Elliott Wimmer[3,5] & Tor D Wager[1]

Pain is a primary driver of learning and motivated action. It is also a target of learning, as nociceptive brain responses are shaped by learning processes. We combined an instrumental pain avoidance task with an axiomatic approach to assessing fMRI signals related to prediction errors (PEs), which drive reinforcement-based learning. We found that pain PEs were encoded in the periaqueductal gray (PAG), a structure important for pain control and learning in animal models. Axiomatic tests combined with dynamic causal modeling suggested that ventromedial prefrontal cortex, supported by putamen, provides an expected value–related input to the PAG, which then conveys PE signals to prefrontal regions important for behavioral regulation, including orbitofrontal, anterior mid-cingulate and dorsomedial prefrontal cortices. Thus, pain-related learning involves distinct neural circuitry, with implications for behavior and pain dynamics.

Both appetitive and aversive primary reinforcers—pleasure and pain—fundamentally shape learning and decision-making. Neural processes that signal appetitive value, including responses in the mesolimbic dopamine system, drive reward-pursuit responses. Pain and other aversive processes drive avoidance and escape. In spite of its importance, however, pain avoidance is poorly understood, and the nature of the cerebral processes underlying pain's motivational functions is an important frontier[1,2].

Much progress in understanding motivational learning systems has come from the application of computational models of reinforcement learning to the analysis of animal brain circuitry[3] and human fMRI data[4]. Such models posit that learning occurs in proportion to the magnitude of the prediction error (PE)—the discrepancy between the predicted value and experienced reward or punishment—evaluated after each action[5]. Reinforcement learning models have been used to identify reward PE signals—which reflect 'better-than-expected' outcomes—in midbrain dopamine neurons[3], ventral striatum (VS) and medial orbitofrontal cortex (OFC). While fMRI activity in these and other areas correlates with parametric estimates of PEs, work examining such activity more carefully with respect to separate algebraic components of the PE[6–8]—or, in a related approach, testing activity against a set of axioms that together comprise the set of conditions that define a PE[9]—has so far validated only VS activity as satisfying all the criteria for appetitive PEs in humans.

Meanwhile, there has not yet been a similarly systematic decomposition of aversive PE-related activity. An emerging body of literature[2,10–13] has identified several candidate regions that may encode aversive PE signals (worse-than-expected outcomes) in humans, including the amygdala[12,14], VS[2,15,16] and lateral OFC[10,17]. However, it remains unclear whether this activity reflects PEs or, rather, related signals such as pain expectancies or aversive responses. In addition, recent animal studies have identified neurons in a different region, the midbrain periaqueductal gray (PAG), with several aversive PE–like properties[1], including elevated firing rates to unexpected versus expected punishment[1,18] and habituation as painful shocks become expected[1,18].

In this study, using a combination of computational modeling and axiomatic approaches with fMRI data, we sought to identify regions encoding aversive PE signals (worse-than-expected outcomes) and aversive value signals (pain expectancies). Participants ($N = 26$) performed a reinforcement-learning task during which they learned to avoid selecting the actions associated with a high probability of receiving pain. On each of 150 trials, participants chose between two options (**Fig. 1a**), each associated probabilistically with the delivery of painful heat (47.4 ± 1.71 °C). Probabilities for each option were governed by two independently varying random walks, so that participants learned to track the changing reinforcement values continuously throughout the task (**Fig. 1b**).

Our first objective was to identify brain regions that encode aversive PE signals and aversive value signals (pain expectancies), particularly in regions commonly thought to mediate PEs from human studies, including VS, and animal models (PAG). We reasoned that using an axiomatic testing approach could provide a stronger test for identifying aversive PEs and aversive value signals. We therefore considered whether (i) signals in PAG and VS correlate with PEs as predicted by a computational reinforcement learning model and (ii) they satisfy the three axiomatic properties that together define aversive PEs (ref. 9; see Results and Online Methods for a detailed description of the axioms).

Second, we sought to develop a brain-based model of how PE- and value-encoding regions interact during learning. The computational framework for reinforcement learning specifies dynamic interactions between brain regions encoding reinforcements, expected values and

[1]Department of Psychology and Neuroscience, University of Colorado, Boulder, Boulder, Colorado, USA. [2]PERFORM Centre, Concordia University, Montreal, Quebec, Canada. [3]Department of Psychology, Columbia University, New York, New York, USA. [4]Center for Neural Science, New York University, New York, New York, USA. [5]Department of Systems Neuroscience, University Medical Center Hamburg–Eppendorf, Hamburg, Germany. Correspondence should be addressed to T.D.W. (tor.wager@colorado.edu).

Received 2 July; accepted 3 September; published online 5 October 2014; doi:10.1038/nn.3832

PEs, but previous fMRI studies have not investigated the inter-region dynamics (that is, effective connectivity) implied by reinforcement learning models. In a strong form of the mapping between model and brain, PE-based value updating may be accomplished by direct connectivity between PE- and value-encoding regions. Here, we used dynamic causal modeling (DCM)[19] to test plausible models of effective connectivity among value-encoding and PE-encoding regions, with the goal of developing an empirically based model of how brain regions interact during pain-driven learning.

RESULTS

Behavioral results

As expected, participants switched options more frequently after receiving pain than no stimulus (40.5 ± 4.2% versus 6 ± 1.1% of trials, $P < 0.0001$). The effects of pain on switching also decayed exponentially with time, as evidenced by the results of logistic regressions assessing the effects of reinforcement history (pain delivered 1 to 6 trials back) on switches ($P < 0.001$ for one and two trials back; **Fig. 1c**).

We then used a standard temporal difference computational learning model to analyze subjects' choices as a function of pain. The model comprised a learning rate parameter (α), controlling the extent to which past feedback influences future predictions, and a softmax inverse temperature (β) parameter controlling the probability of selecting the most advantageous option. The analysis revealed learning rates ($\alpha = 0.63 \pm 0.26$), softmax inverse temperatures ($\beta = 4.74 \pm 2.74$) and model fits (negative log likelihood = 65 ± 21) comparable to those found in similar studies of reinforcement learning[20]. These results, along with the exponential form of the influences of previous pain (**Fig. 1c**), suggest that the temporal difference model captures pain avoidance learning in this task.

Aversive prediction error signals

Aversive PE signals should be phasically triggered at the moment when participants learn that punishment will be delivered and should correlate with computational model–derived PEs. Here we identified PE-correlated regions by regressing fMRI activity at outcome onset (see **Supplementary Fig. 1**) on model-derived PEs determined by fitting a temporal difference model to the individual's choice behavior (see Online Methods). Activity correlating with model-based aversive PEs (greater activity for worse-than-expected outcomes) was found in several areas (**Fig. 1d**). These included the left anterior insula, anterior and mid-cingulate cortices (ACC and MCC, respectively), the right pre- and post-central gyri, the right dorsolateral prefrontal cortex and a large cluster in the midbrain encompassing the periaqueductal gray (PAG). Negative correlations with PE (greater activity for better-than-expected outcomes) were found in the entorhinal and parahippocampal cortices, right inferior frontal gyrus, right temporal pole and right lateral thalamus (**Fig. 1d** and **Supplementary Table 1**).

In the reward domain, it has been shown that some signals that correlate with PE are better explained as relating to some other quantity, such as reward magnitude, that is intrinsically correlated with PE[6,8,9,14]. Here, activity that tracked aversive PEs was similar to activity related to pain onset versus no-stimulus onset (see **Supplementary Fig. 2** and **Supplementary Table 1**; note that both pain and no-stimulus trials were indicated by an identical change in the fixation cross to avoid temporal ambiguity). Thus, to ensure that the candidate PE-related fMRI signals truly integrate outcome and expectancy information into an aversive PE signal, we used an axiomatic approach[9] (see Online Methods), which specifies a set of three conditions that together define a PE. In the context of our task, these were as follows. Axiom 1: activity should be higher for received than avoided pain, unless pain is fully expected. Axiom 2: activity should decrease in proportion to expected pain (that is, expected aversive value), for both pain and no-stimulus trials. Axiom 3: activity on pain and no-stimulus trials should be equivalent if the outcome is completely predicted.

Here the first two conditions correspond to tests of effects of outcome and expectancy, respectively, the conjunction of which constitutes the algebraic definition of PEs ($r_t - V_t$). These tests are analogous to those in other recent work[6,8,14], while the third condition, less often explicitly examined, verifies that the magnitudes of these two separate and opposite effects are equivalent, so that fully predicted outcomes do not generate PE signals. Thus, the axioms as applied in our case reflect the requirements specified by the mathematical definition of an aversive PE.

Brain regions that satisfy all the axioms should show a distinctive profile of activity as a function of expectancy and pain delivery, whereas those that track only pain expectancy or delivery will show different patterns (**Fig. 2a**). Because axiom 3 depends on support for the null hypothesis, we conducted additional Bayesian analyses of the odds in favor of versus against the null hypothesis[21].

Region of interest (ROI) analyses revealed that PAG, but not VS, fulfilled all the axioms for aversive PE signals (**Fig. 2b**). The PAG

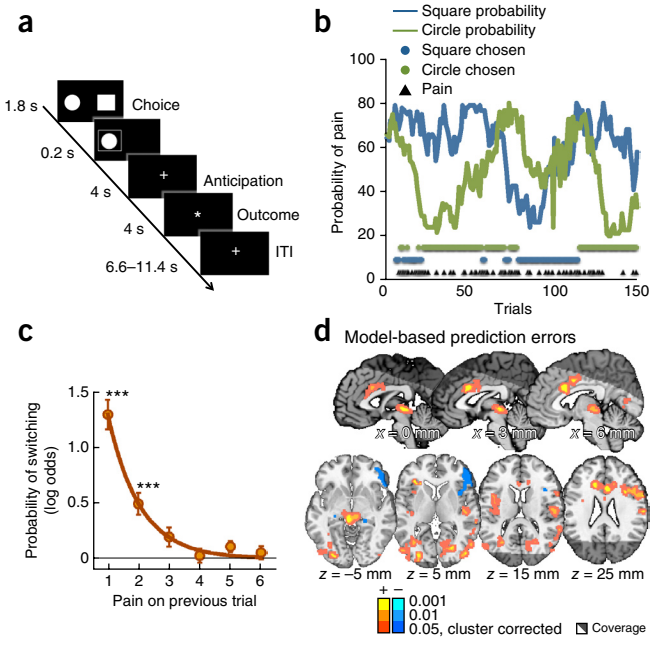


Figure 1 Pain avoidance learning task, behavioral and brain imaging results. (**a**) One experimental trial. Participants had 1.8 s to make their choice, after which their choice was displayed for 0.2 s. After an anticipation period of 4 s, participants either received a painful stimulus or nothing. The stimulation period was marked by a different fixation point. Trials were separated by a 6.6–11.4 s intertrial interval (ITI). (**b**) Data from one participant. The blue and green lines depict the probability of pain associated with each option over the 150 trials (one of four possible pairs of random walks). Blue and green dots represent the selected option and black triangles represent pain delivery. (**c**) Logistic regression model results (number of participants = 23). Probability of switching (mean log odds ± s.e.m.) as a function of pain one to six trials back decays exponentially and is significantly different from zero at one ($t(22) = 9.20$, $P < 0.001$) and two ($t(22) = 4.36$, $P < 0.001$) trials back. (**d**) Activity correlated with reinforcement learning model–based prediction errors at pain onsets (number of participants = 23), cluster-thresholded ($P < 0.05$, corrected for family-wise error rate (FWER), two-tailed) with cluster-defining thresholds of $P < 0.001$, $P < 0.01$ and $P < 0.05$.

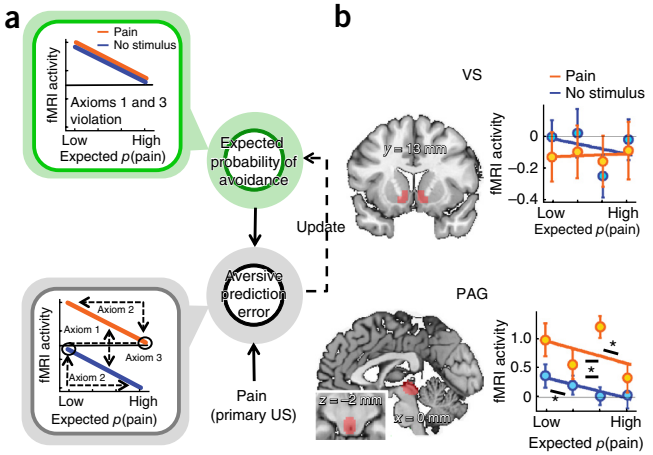


Figure 2 Pain avoidance learning model and axiomatic tests in VS and PAG ROIs. (**a**) Regions encoding the expected probability of avoidance (green) should display higher activity only for low expected probability of pain, regardless of outcome. Regions encoding aversive prediction errors (gray) should display higher activity for pain versus no-stimulus trials (axiom 1), higher activity for low expected probability of pain, regardless of outcome (axiom 2), and no difference in activity between highly predicted pain or no-stimulus outcomes (axiom 3). Aversive prediction errors result from the integration of pain-related information with prior expectations. This signal is then used to update future predictions. US, unconditioned stimulus. (**b**) Activity in a priori VS and PAG ROIs per quartile of expected probability of pain (number of participants = 23). Activity in the VS satisfied only axiom 3 (no difference between highly predicted pain or no-stimulus outcomes, Bayesian analyses, odds in favor of the null hypothesis = 7.07). By contrast, activity in the PAG satisfied all three axioms for aversive prediction errors (axiom 1 (pain > no stimulus): $t(22) = 3.67$, $P = 0.001$; axiom 2 (significance of slope for pain trials, sign permutation test): $t(22) = -2.05$, $P = 0.044$; axiom 2 (significance of slope for no-stimulus trials, sign permutation test): $t(22) = -1.98$, $P = 0.048$; axiom 3 (no difference between highly predicted pain or no-stimulus outcomes, Bayesian analyses, odds in favor of the null hypothesis = 5.48). Asterisk with horizontal bars indicates significant differences between pain and no-stimulus outcomes. Asterisks with diagonal bars indicate significant slopes ($*P < 0.05$). Error bars represent s.e.m.

responded more strongly to pain trials than no-stimulus trials, holding pain expectancy constant (axiom 1; $t(22) = 3.67$, $P < 0.05$). It showed reduced responses to outcomes with greater pain expectancy, for both pain and no-stimulus trials (axiom 2; pain trials: $t(22) = -2.05$, $P < 0.05$; no-stimulus trials: $t(22) = -1.98$, $P < 0.05$). And finally, it showed no difference between fully predicted pain and no-stimulus trials (axiom 3; $t(22) = 0.13$, $P = 0.90$, odds in favor of the null hypothesis = 5.48). The VS did not show any effect of pain expectancy (pain trials, $P = 0.87$; no-stimulus trials, $P = 0.40$), violating axiom 2, and did not respond to pain versus no-stimulus outcomes ($P = 0.62$), violating axiom 1.

To search for additional regions that might satisfy the axioms for aversive PEs, we conducted a whole-brain conjunction search for three relevant contrasts: (i) pain onset versus no-stimulus onset; (ii) expectancy effects—that is, parametric variation with the degree of model-based expectancy—on pain trials; and (iii) expectancy effects on no-stimulus trials (**Fig. 3**). Significant results in effect i satisfy axiom 1 and significant results for effects ii and iii satisfy axiom 2. A region of the PAG extending into the tectum (**Fig. 3** and **Supplementary Table 2**) was the only region to show significant results in all three tests ($P < 0.05$, cluster-extent corrected). To test axiom 3, we compared activity within that cluster for highly expected pain and no-stimulus outcomes. There was no activity difference between pain and no-stimulus trials when outcomes were highly predicted ($t(22)$: 0.56, $P > 0.4$, odds in favor of the null hypothesis = 5.13), thereby confirming axiom 3.

Studies 2 and 3: monetary rewards and varying pain levels

In this study, we chose not to include rewarding events, in part because many studies have demonstrated reward-related PEs linked to VS[9,22] and in part to avoid the complexity caused when participants directly compare rewarding and punishing events. However, to provide additional evidence on whether aversive and appetitive PEs are encoded in different brain circuits, we reanalyzed data from a published experiment[23] that used a similar experimental design with monetary rewards (**Supplementary Fig. 3**), focusing on the VS and PAG. As expected, in contradistinction to the main study results, appetitive PEs to monetary rewards were tracked by activity in the VS ($t(20) = 5.77$, $P < 0.001$), but not the PAG (PAG–appetitive: $t(20) = 1.54$, $P = 0.14$; see **Supplementary Fig. 3**). The signal-to-noise (SNR) ratios in the VS (171.13 ± 9.68) and PAG (163.20 ± 5.57) were not significantly different ($P > 0.23$). However, we note that

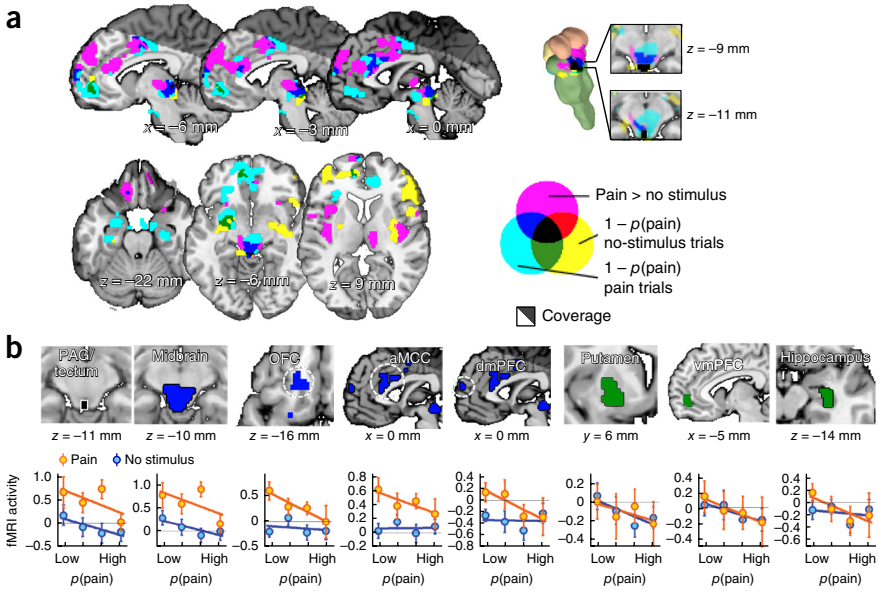


Figure 3 Results from the whole-brain conjunctive search (number of participants = 23). (**a**) Conjunction analysis of pain > no-stimulus effects (axiom 1) and expected avoidance probability (1 – pain probability) effects for both pain and no-stimulus outcomes (axiom 2). Clusters used for the conjunction analysis were cluster-thresholded ($P < 0.05$, FWER, one-tailed) with a cluster-defining threshold of $P < 0.05$. (**b**) Profiles of activity within ROIs defined from the conjunction analysis. For pain and no-stimulus outcomes, mean responses were extracted by quartiles of model-based probability. Error bars represent s.e.m.

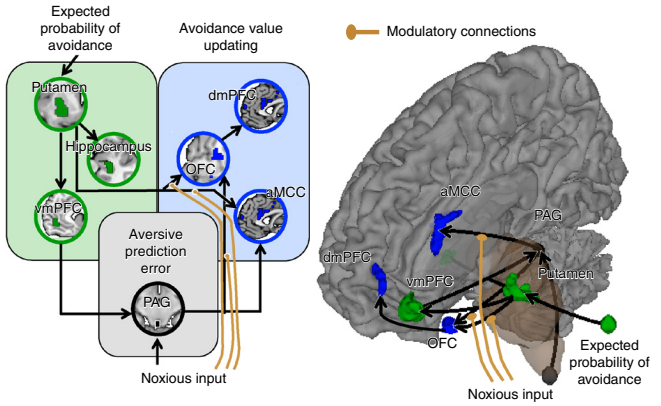


Figure 4 DCM of aversive prediction errors at outcome onset (number of participants = 23). This model was identified as the most likely of all the models tested (see **Supplementary Figs. 6–9**) by Bayesian model selection. The regions included in the model were identified by the previous conjunction analysis (**Fig. 3**) as reflecting 1 – expected probability of pain regardless of outcome (green), aversive prediction errors (black) or avoidance value updating (blue).

the dissociation between aversive and reward PEs depends on null findings in the PAG in study 2 and other reward learning studies. It is possible that high-resolution and brainstem-optimized imaging (for example, refs. 24,25) could yield additional reward-related signals that remain to be discovered. Moreover, differences in field strength (3 T versus 1.5 T) and other scanning parameters could also have impacted the ability to identify appetitive PE signals in the PAG in study 2.

Another important issue is whether putative PE-related signals are related to pain intensity or merely the presence versus absence of an aversive reinforcer. In study 3 (n = 50; **Supplementary Fig. 4**), we sought to replicate aversive PE–related findings in the PAG and test for activation related to noxious stimulus intensity. This study used three intensities of painful stimulation in the noxious range (46 °C, 47 °C and 48 °C) and two independent manipulations of expectations about pain intensity: (i) a classical conditioning procedure and (ii) unreinforced placebo instructions designed to induce expectations of relief (see **Supplementary Fig. 4**). Activity corresponding to the axiomatic requirements for prediction errors was analyzed in the time window during which the three stimulus intensities were subjectively differentiated (4–10 s after stimulus onset; see **Supplementary Fig. 4d**).

In accordance with axiom 1, PAG activity within that time window increased with temperature ($P < 0.001$). In accordance with axiom 2, activity was higher during pain for low versus high pain conditioned cues ($F(1,49) = 4.39$, $P < 0.05$) and for the placebo versus control condition ($F(1,49) = 16.03$, $P < 0.001$). In both of these conditions, the stimulus was higher than expected on the basis of cues and verbal instructions, respectively. Finally, it was not possible to definitively test axiom 3 in study 3 because cues never fully predicted outcomes. Overall, results from this supplementary experiment replicated and extended the findings of pain-related aversive prediction errors in the PAG, demonstrating sensitivity to the level of painful stimulus intensity and sensitivity to verbal instructions as well as predictive cues.

Expectancies and other learning-related variables

Regions that track expected avoidance value—a contributor to aversive PEs—should show effects of the expected probability of pain but no effects of pain itself (**Fig. 2a**). In terms of neural effects, this translates into greater activity with low pain expectancy for both pain and no-stimulus trials, but no difference in activity between pain and no-stimulus trials. We identified clusters in the left putamen, ventromedial prefrontal cortex (vmPFC) and right hippocampus in which activity fit this profile, consistent with expectancy effects. These regions displayed increased activity when pain was expected to be avoided (low pain expectancy) but did not respond differentially to pain versus no-stimulus outcomes (**Fig. 3a**). *Post hoc* analyses confirmed that no areas showed significant pain versus no-stimulus effects (left putamen: $P = 0.43$, Bayes factor in favor of the null hypothesis = 7.13; vmPFC: $P = 0.93$, Bayes factor in favor of the null hypothesis = 10.03; right hippocampus: $P = 0.81$, Bayes factor in favor of the null hypothesis = 9.44).

The conjunction analyses we conducted can identify regions that do not conform precisely to all elements of the reinforcement learning model but may nonetheless be important for guiding behavior and learning. Several regions identified in the conjunction analysis correlated with PEs only on pain trials (**Fig. 3a**), including the left OFC, anterior MCC (aMCC), dorsomedial prefrontal cortex (dmPFC) and a larger dorsal midbrain cluster comprising the PAG, tectum, nucleus cuneiformis (NCF), dorsal raphe nucleus (DRN) and red nucleus. Though there are several potential interpretations, we suggest that these areas reflect updating of the value of switching away from the punished option on the next trial, a decision participants only have to make after pain delivery.

Finally, follow-up ROI analyses of response patterns within these regions revealed that the midbrain showed a significant correlation with expected value on no-stimulus trials ($t(22) = -1.82$, $P = 0.05$, one-tailed) that did not meet the whole-brain threshold (**Fig. 3b**). Thus, findings in this larger midbrain cluster are consistent with aversive PE signals, though the dorsal PAG region was the only portion to survive whole-brain correction in all three contrasts. The other three regions showed little evidence for expectancy effects on no-stimulus trials (Bayes factors in favor of the null hypothesis: aMCC, 2.50; OFC, 4.07; dmPFC, 2.26) and thus are more likely to reflect avoidance value updating or other motivational processes. Finally, we note that although the current results relate signal in the PAG as a whole to aversive PEs, it is possible that high-resolution and brainstem-optimized imaging could reveal a finer-grained distribution of PAG subregions with functionally distinct response profiles[24,26], including portions that respond only to expectancies. More broadly, our results do not imply that aversive PEs are the only signal represented in the PAG.

The previous analyses examined fMRI activity at pain onset, when PEs are generated. Brain regions that encode expected value should also be active earlier, when decisions are made and the expected value is computed. To identify such regions, we examined activity that parametrically tracked the expected probability of avoidance at the time of decision (**Supplementary Fig. 5** and **Supplementary Table 3**). Positive effects (that is, greater activity with high avoidance value or low pain expectancy) were observed in the ventromedial prefrontal cortex (vmPFC), and in particular in the medial OFC and perigenual ACC. Conversely, negative activations were observed in the aMCC, lateral frontal pole, parietal operculum, cerebellum and visual cortex.

Network dynamics underlying aversive PE signals

To develop a brain-based model of the learning process, we used DCM[19] to explore how the seven regions identified in the previous analysis (**Fig. 3b**; note that the larger midbrain cluster was not included in the DCM analyses) interact during learning. On the basis of the principles governing reinforcement learning models (**Fig. 2a**), regions that encode aversive PEs (PAG) should receive converging input from those that encode expectancies (vmPFC, putamen, hippocampus) and primary reinforcement (nociceptive) signals. Afferent nociceptive signals in PE-encoding regions should be cancelled out by expectancy-related information when those signals are fully

predicted[1,18]. Regions important for action value and decision-making (aMCC, OFC, dmPFC) may receive converging PE and primary reinforcement signals.

As is increasingly common with DCMs, we tested a family of similar models to identify the most likely configuration of connections based on the data. This is conceptually analogous to optimizing parameter values (for example, in linear regression), except that we search over models, identifying the most likely pattern of connections given the data using Bayesian model selection[27]. We defined a model limited to brain correlates of reinforcement learning model–based effects (PAG, vmPFC, putamen, hippocampus) and then extended the model to include other regions that may encode avoidance value and related properties, testing 72 plausible models in total (see Online Methods and **Supplementary Figs. 6–9**). Hence, the final model was jointly constrained by a priori theoretical constraints and the evidence in the data.

In the final, most likely model (**Fig. 4**), vmPFC projects most directly to PAG among value-encoding regions, and avoidance value is most closely related to the putamen, which transmits value information to vmPFC. Then PE and expected value signals from the midbrain and putamen are transmitted to OFC and aMCC, with effects on dmPFC mediated by OFC. Noxious input had direct effects on PAG, in keeping with the known anatomy of the ascending spino-mesencephalic nociceptive pathway[28] and modulatory effects on ascending putamen-to-OFC, putamen-to-aMCC and midbrain-to-aMCC connections. These modulatory effects are plausible given that the spino-thalamic tract and other pathways provide separate channels of ascending nociceptive input that are distinct from spino-mesencephalic inputs to the PAG.

DISCUSSION

Toward a neural systems model for aversive PEs

Pain has obvious motivational functions, both shaping and being shaped by learning, but we still know very little about the basic neural processes underlying its influence on human behavior. Using a combination of reinforcement learning computational models, axiomatic tests and DCM models, we identified a candidate system that allows humans to avoid actions associated with pain. In this model, a system of interconnected forebrain regions including the putamen, hippocampus and vmPFC encodes expected value signals. Expectations are then compared with primary nociceptive inputs in the PAG to generate aversive PE signals. These signals shape expectations maintained in medial prefrontal-temporal-striatal systems and are also conveyed to forebrain structures involved in behavioral decisions and choice (aMCC, dmPFC and OFC).

Our data and connectivity models identify the PAG as a primary site for aversive PEs, in contrast to previous neuroimaging findings and theories that a single system drives appetitive and aversive PEs[29]. The centrality of the PAG for aversive PEs is consistent with both anatomical and neurophysiological evidence in animals. The PAG receives monosynaptic inputs from both nociceptive spinal projection neurons[28] and top-down projections from the vmPFC[30], positioning it as a potential comparator of bottom-up aversive sensations with expectations. It also sends monosynaptic, reciprocal projections to the vmPFC—which is essential for value updating in the reinforcement learning framework and likely for behavioral choice as well[31]—and to aMCC[32], OFC and other areas involved in determining action value and coordinating defensive behavior[18,33]. The aMCC in particular is critical for pain avoidance[18] and is heavily connected to motor and premotor centers.

The central role of the PAG in aversive PEs is also consistent with several prominent animal models of aversive learning[34]. These models suggest that the PAG is critical for integrating expectations with ascending nociceptive information. Though animal studies have not formally tested the axioms that satisfy aversive PEs directly, as we have here, several functional properties of the PAG in animals are consistent with PE signaling[1], such as higher firing rates to unexpected versus expected punishment[1,18]. These expectancy effects seem to be mediated by inhibition of ascending nociceptive inputs through the release of endogenous opioids in the PAG[1], which blocks nociceptive responses when pain is expected. This converging animal evidence suggests that opioidergic modulation of the PAG may be a critical element of aversive PEs.

Overlap in systems for reward and aversive PEs

The nature and degree of overlap between appetitive and aversive learning is intensely debated. On the one hand, some proponents of a unitary system for reward and aversion have stressed the close coexistence of neuronal populations signaling appetitive and aversive value in structures including the striatum and ventral tegmental area[35]. Other arguments in favor of a unitary system come from neuroimaging studies showing that a common set of regions activates (vmPFC, striatum) or deactivates (ACC, insula, dorsolateral prefrontal cortex) parametrically with increasing outcome value across both aversive (monetary losses) and appetitive (monetary gains) domains[36,37]. However, these findings may be caused by framing of the outcomes as relative gains or losses compared to the alternative and hence not truly reflect categorical similarities between appetitive and aversive learning systems. That is, losing endowed money may not engage learning systems adapted for primary punishments like pain, and both the nature of the reinforcer (primary versus secondary) and the specific type of reinforcement (thermal pain versus loss) may be important.

On the other hand, imaging studies using primary aversive reinforcers such as pain have converged on a set of candidate regions that are potentially specific to aversive learning, including the brainstem, amygdala, OFC, insula and ACC[2,10,12,14,38], but the results across studies have also been mixed. Part of this variability could be due to heterogeneity in the response properties of different neuronal subpopulations[39], but also to the fact that latent variables derived from reinforcement learning models, such as PEs, are by definition correlated with related signals, such as expected values or outcome information. As a result, PE signals within a given voxel can be highly correlated with expected values or outcome information. In the case of our study, regions tracking parametric estimates of aversive PEs strongly overlapped with regions signaling pain onsets (**Supplementary Fig. 2**), making them indistinguishable without more fine-grained tests.

Here the use of axiomatic tests allowed us to dissociate regions tracking aversive PEs from similar, intrinsically correlated signals such as expectancy and nociception or pain. Only the PAG showed consistent evidence for aversive PEs in all axiomatic tests. By contrast, activity in a VS ROI previously shown to fulfill all axiomatic requirements for an appetitive PE signal[9] showed no evidence for aversive PE signals, although it encoded appetitive PEs to monetary rewards in a separate experiment using a similar design. Conversely, activity in the PAG did not correlate with appetitive PEs to monetary rewards, suggesting that there is at least partial segregation between aversive and appetitive systems at the level of PEs. Indeed, the functional neuroanatomy of the PAG strongly indicates it is highly specialized in the treatment of intrinsically aversive stimuli[26,40]. Among other primary aversive reinforcers, PAG is activated by painful events[41], aversive images[24,42] and social threats[43]. By contrast, a recent meta-analysis of over 200 neuroimaging studies found no reliable reward-related signal in the PAG[44].

By contrast, vmPFC activity seems to reflect expected positive value in a variety of contexts and paradigms[45], including reward

learning, economic choice[46], pain avoidance[11] and extinction/extinction recall[47]. Our results are consistent with the vmPFC as the most direct representation of value as related to choice and learning (**Fig. 3** and **Supplementary Fig. 5**) and most closely connected to the PAG in the retained DCM model. This is consistent with recent findings that the vmPFC may act as a hub or convergence point for different types of expected value signals, such as experience-based expected value signals computed in the putamen and model-based value signals computed in the caudate[48]. Hence, our results seem to indicate both a convergence between appetitive and aversive systems in value representations in the vmPFC and a segregation between the two systems when these expected values signals are integrated with ascending aversive unconditioned stimulus inputs in the PAG to generate PE signals.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We would like to thank D. Abraham and A. Pingree for help with data collection and A. Krishnan, L. Schmidt and L. Atlas for help with data analyses. This work was supported by a grant from the US National Institutes of Health to T.D.W. (R01DA035484 and R01MH076136) and by Canadian Institute of Health Research and Fonds de Recherche en Santé du Québec fellowships to M.R. N.D. was supported by a Scholar Award from the James S. McDonnell Foundation.

AUTHOR CONTRIBUTIONS

M.R., N.D., D.S. and T.D.W. designed the study. M.R. performed the study. M.R., N.D. and T.D.W. analyzed the data, and M.R., N.D., D.S. and T.D.W. wrote the paper. G.E.W., N.D. and D.S. designed study 2, G.E.W. performed study 2, and G.E.W., N.D., D.S. and M.R. analyzed the data. M.J. and T.D.W. designed study 3, M.J. performed the study, and M.J., T.D.W. and M.R. analyzed the data.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. McNally, G.P., Johansen, J.P. & Blair, H.T. Placing prediction into the fear circuit. *Trends Neurosci.* **34**, 283–292 (2011).
2. Seymour, B. *et al.* Temporal difference models describe higher-order learning in humans. *Nature* **429**, 664–667 (2004).
3. Hollerman, J.R. & Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* **1**, 304–309 (1998).
4. O'Doherty, J.P., Hampton, A. & Kim, H. Model-based fMRI and its application to reward learning and decision making. *Ann. NY Acad. Sci.* **1104**, 35–53 (2007).
5. Daw, N.D. in *Decision Making, Affect and Learning.* (eds. Delgado, M.R., Phelps, E.A. & Robbins, T.W.) 3–38 (Oxford Univ. Press, 2011).
6. Behrens, T.E.J., Hunt, L.T., Woolrich, M.W. & Rushworth, M.F.S. Associative learning of social value. *Nature* **456**, 245–249 (2008).
7. Li, J. & Daw, N.D. Signals in human striatum are appropriate for policy update rather than value prediction. *J. Neurosci.* **31**, 5504–5511 (2011).
8. Niv, Y., Edlund, J.A., Dayan, P. & O'Doherty, J.P. Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *J. Neurosci.* **32**, 551–562 (2012).
9. Rutledge, R.B., Dean, M., Caplin, A. & Glimcher, P.W. Testing the reward prediction error hypothesis with an axiomatic model. *J. Neurosci.* **30**, 13525–13536 (2010).
10. Seymour, B. *et al.* Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nat. Neurosci.* **8**, 1234–1240 (2005).
11. Seymour, B., Daw, N.D., Roiser, J.P., Dayan, P. & Dolan, R. Serotonin selectively modulates reward value in human decision-making. *J. Neurosci.* **32**, 5833–5842 (2012).
12. Yacubian, J. *et al.* Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *J. Neurosci.* **26**, 9530–9537 (2006).
13. Ploghaus, A. *et al.* Learning about pain: the neural substrate of the prediction error for aversive events. *Proc. Natl. Acad. Sci. USA* **97**, 9281–9286 (2000).
14. Li, J., Schiller, D., Schoenbaum, G., Phelps, E.A. & Daw, N.D. Differential roles of human striatum and amygdala in associative learning. *Nat. Neurosci.* **14**, 1250–1252 (2011).
15. Schiller, D., Levy, I., Niv, Y., LeDoux, J.E. & Phelps, E.A. From fear to safety and back: reversal of fear in the human brain. *J. Neurosci.* **28**, 11517–11525 (2008).
16. Delgado, M.R., Li, J., Schiller, D. & Phelps, E.A. The role of the striatum in aversive learning and aversive prediction errors. *Phil. Trans. R. Soc. Lond. B* **363**, 3787–3800 (2008).
17. Hindi Attar, C., Finckh, B. & Büchel, C. The influence of serotonin on fear learning. *PLoS ONE* **7**, e42397 (2012).
18. Johansen, J.P., Tarpley, J.W., LeDoux, J.E. & Blair, H.T. Neural substrates for expectation-modulated fear learning in the amygdala and periaqueductal gray. *Nat. Neurosci.* **13**, 979–986 (2010).
19. Stephan, K.E. *et al.* Dynamic causal models of neural system dynamics: current state and future extensions. *J. Biosci.* **32**, 129–144 (2007).
20. Schönberg, T., Daw, N.D., Joel, D. & O'Doherty, J.P. Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J. Neurosci.* **27**, 12860–12867 (2007).
21. Gallistel, C.R. The importance of proving the null. *Psychol. Rev.* **116**, 439–453 (2009).
22. Garrison, J., Erdeniz, B. & Done, J. Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neurosci. Biobehav. Rev.* **37**, 1297–1310 (2013).
23. Wimmer, G.E., Daw, N.D. & Shohamy, D. Generalization of value in reinforcement learning by humans. *Eur. J. Neurosci.* **35**, 1092–1104 (2012).
24. Satpute, A.B. *et al.* Identification of discrete functional subregions of the human periaqueductal gray. *Proc. Natl. Acad. Sci. USA* **110**, 17101–17106 (2013).
25. Beissner, F. & Baudrexel, S. Investigating the human brainstem with structural and functional MRI. *Front. Hum. Neurosci.* **8**, 116 (2014).
26. Keay, K.A. & Bandler, R. Parallel circuits mediating distinct emotional coping reactions to different types of stress. *Neurosci. Biobehav. Rev.* **25**, 669–678 (2001).
27. Schmidt, L., Lebreton, M., Cléry-Melin, M.-L., Daunizeau, J. & Pessiglione, M. Neural mechanisms underlying motivation of mental versus physical effort. *PLoS Biol.* **10**, e1001266 (2012).
28. Millan, M.J. The induction of pain: an integrative review. *Prog. Neurobiol.* **57**, 1–164 (1999).
29. Brooks, A.M. & Berns, G.S. Aversive stimuli and loss in the mesocorticolimbic dopamine system. *Trends Cogn. Sci.* **17**, 281–286 (2013).
30. Price, J.L. Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. *Ann. NY Acad. Sci.* **1121**, 54–71 (2007).
31. Rangel, A. & Hare, T. Neural computations associated with goal-directed choice. *Curr. Opin. Neurobiol.* **20**, 262–270 (2010).
32. Herrero, M.T., Insausti, R. & Gonzalo, L.M. Cortically projecting cells in the periaqueductal gray matter of the rat. A retrograde fluorescent tracer study. *Brain Res.* **543**, 201–212 (1991).
33. Shackman, A.J. *et al.* The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* **12**, 154–167 (2011).
34. Krasne, F.B., Fanselow, M.S. & Zelikowsky, M. Design of a neurally plausible model of fear learning. *Front. Behav. Neurosci.* **5**, 41 (2011).
35. Reynolds, S.M. & Berridge, K.C. Emotional environments retune the valence of appetitive versus fearful functions in nucleus accumbens. *Nat. Neurosci.* **11**, 423–425 (2008).
36. Tom, S.M., Fox, C.R., Trepel, C. & Poldrack, R.A. The neural basis of loss aversion in decision-making under risk. *Science* **315**, 515–518 (2007).
37. Kim, H., Shimojo, S. & O'Doherty, J.P. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol.* **4**, e233 (2006).
38. Boll, S., Gamer, M., Gluth, S., Finsterbusch, J. & Büchel, C. Separate amygdala subregions signal surprise and predictiveness during associative fear learning in humans. *Eur. J. Neurosci.* **37**, 758–767 (2013).
39. Paton, J.J., Belova, M.A., Morrison, S.E. & Salzman, C.D. The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* **439**, 865–870 (2006).
40. Linnman, C., Moulton, E.A., Barmettler, G., Becerra, L. & Borsook, D. Neuroimaging of the periaqueductal gray: state of the field. *Neuroimage* **60**, 505–522 (2012).
41. Buhle, J.T. *et al.* Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging studies. *Cereb. Cortex* doi:10.1093/cercor/bht154 (2013).
42. Buhle, J.T. *et al.* Common representation of pain and negative emotion in the midbrain periaqueductal gray. *Soc. Cogn. Affect. Neurosci.* **8**, 609–616 (2013).
43. Wager, T.D. *et al.* Brain mediators of cardiovascular responses to social threat, part II: Prefrontal-subcortical pathways and relationship with anxiety. *Neuroimage* **47**, 836–851 (2009).
44. Bartra, O., McGuire, J.T. & Kable, J.W. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* **76**, 412–427 (2013).
45. Roy, M., Shohamy, D. & Wager, T.D. Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends Cogn. Sci.* **16**, 147–156 (2012).
46. Chib, V.S., Rangel, A., Shimojo, S. & O'Doherty, J.P. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J. Neurosci.* **29**, 12315–12320 (2009).
47. Milad, M.R. *et al.* Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biol. Psychiatry* **62**, 446–454 (2007).
48. Wunderlich, K., Dayan, P. & Dolan, R.J. Mapping value based planning and extensively trained choice in the human brain. *Nat. Neurosci.* **15**, 786–791 (2012).

ONLINE METHODS

Participants. Twenty-six healthy, right-handed participants completed the study (mean age = 26.7 ± 7.6 years, 14 females). The sample consisted of 52% Caucasian, 20% Asian, 16% Hispanic and 12% African-American participants. All participants provided informed consent. The study was approved by the Columbia University Institutional Review Board. Preliminary eligibility was assessed with a general health questionnaire, a pain safety screening form and an fMRI safety screening form. Participants reported no history of psychiatric, neurological or pain disorders. Three participants were excluded from the analysis because of poor performance on the task (see section on reinforcement model-based analysis below).

Thermal stimulation. Thermal stimulation was delivered to the volar surface of the left (nondominant) inner forearm using a 16 × 16 mm Peltier thermode (Medoc). To minimize the effects of peripheral sensitization/habituation, the thermode was moved to a new skin spot after each run. Each stimulus lasted 9 s with 2.5-s ramp-up and ramp-down periods and 4 s at target temperature. Temperatures were individually calibrated to be at a level 7 on a continuous scale ranging from 0 to 8 (0, no sensation; 1, nonpainful warmth; 2, low pain; 5, moderate pain; 8, maximum tolerable pain) during a practice session performed on a separate day before the imaging session. On the basis of this procedure, a single temperature level was selected within each participant's tolerance limit. The average temperature of the stimuli was 47.4 ± 1.71 °C.

Experimental task. The pain avoidance instrumental learning task comprised 150 trials (divided in 6 runs of 25 trials), during which subjects had to select the option with the lowest probability of being followed by a painful thermal stimulation. The probabilities associated with each option were independent of one another and varied from trial to trial according to pairs of random walks. Four pairs of random walks were selected on the basis of the criterion that they must cross (reverse) at least one time (see **Supplementary Fig. 10**); each participant was randomly administered one of the four pairs.

Each trial (see **Fig. 1a**) started with the presentation of the two options (circle or square, randomly displayed to the left or right) for 1,800 ms, during which participants had to enter their decision by pressing the left or right button of the response unit. If the participant did not have time to make a choice (<1% of trials), the computer randomly selected a response for them. After a feedback period of 200 ms and an anticipation period of 4,000 ms, the fixation point changed from an asterisk (*) to a cross (+) that stayed on the screen for 9,000 ms to mark the period during which participants could receive a painful thermal stimulation. After that stimulation period, the fixation point changed back to an asterisk for a jittered inter-trial interval of 6,600, 7,800, 9,000, 10,200 or 11,400 ms. On a day before the imaging session, participants performed a practice session with a different pair of random walks and options (diamond and triangle) from the ones they received during the imaging session. During that practice session, they were carefully instructed about all aspects of the experiment, except the actual probabilities of pain that they had to infer. Participants provided on-line continuous ratings of pain (0, no sensation; 1, nonpainful warmth; 2, low pain; 5, moderate pain; 8, maximum tolerable pain) for the practice, but not imaging, session.

fMRI data acquisition and preprocessing. *Data acquisition.* Whole-brain fMRI data were acquired on a 1.5-T GE Sigma TwinSpeed Excite HD scanner (GE Medical Systems) at the Functional MRI Research Center at Columbia University. Functional images were acquired with a T2*-weighted, two-dimensional gradient echo spiral in/out pulse sequence[49] (repetition time (TR) = 3,000 ms; echo time = 30 ms; flip angle = 84°; field of view = 224 mm; 64 × 64 matrix, 3.5 × 3.5 × 2.2 mm voxels, 64 slices). To maximize signal in the vmPFC, slices were tilted by 30° from AC–PC axis, resulting in a loss of coverage in dorsoposterior parietal areas, including S1 in the arm area. We were therefore unable to assess the contribution of S1 in pain avoidance learning. Each run lasted 10 min 20 s (206 TRs). Stimulus presentation and data acquisition were controlled using E-Prime software (Psychology Software Tools). Responses were made with the right hand via an MRI-compatible response unit (Resonance Technologies). Visual stimuli were presented through goggles positioned on the scanner head coil (Avotech).

Preprocessing. Before preprocessing, global outlier time points (that is, 'spikes' in BOLD signal) were identified by computing both the mean and the s.d. (across voxels) of values for each image for all slices. Mahalanobis distances for the matrix of slice-wise mean and s.d. values (concatenated) × functional volumes (time) were computed, and any values with a significant χ^2 value (corrected for multiple comparisons based on the more stringent of either false-discovery-rate or Bonferroni method) were considered outliers. Less than 1% of images were outliers. The output of this procedure was later used as a covariate of noninterest in the first-level models.

Functional images were slice-acquisition-timing and motion corrected using SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK). Structural T1-weighted images were registered to the first functional image for each subject using an iterative procedure of automated registration using mutual information co-registration in SPM8 and manual adjustment of the automated algorithm's starting point until the automated procedure provided satisfactory alignment. Structural images were normalized to MNI space using SPM8, interpolated to 2 × 2 × 2 mm voxels, and smoothed using a 6-mm full-width at half maximum Gaussian kernel.

Reinforcement model-based analysis. Participants' decisions were modeled as a function of previous choices and rewards using a temporal difference algorithm. Specifically, the predicted value for options "square" and "circle" (V_{square} or V_{circle}) were updated in the direction of the obtained reward using a delta rule with learning rate α whenever that option was chosen [$V_{\text{chosen option}(t+1)} = V_{\text{chosen option}(t)} + \alpha \times (r_t - V_{\text{chosen option}(t)})$], where r_t is the reward (pain = −1; no stimulus = 0) obtained at trial t. The probability of choosing option i over j at trial t was determined by a softmax distribution, where the inverse temperature parameter β controls the strength of the competition between the two options: $p(\text{choice}_t = \text{"square"} \mid V_{\text{square}(t)}, V_{\text{circle}(t)}) = \exp(\beta V_{\text{square}(t)})/[(\exp(\beta V_{\text{square}(t)}) + (\exp(\beta V_{\text{circle}(t)})]$. Model fits were estimated by negative log likelihoods (smaller values indicate better fit).

The temporal difference model could not be fitted or gave aberrant α or β values ($\alpha = 1$ or 0; $\beta = 0$) for three subjects. This was caused by complete reliance on a win-stay, lose-shift strategy (one subject), frequent switches in choice following absence of pain, which was caused by use of an irrelevant strategy (one subject) or numerous missing responses (20% missing; one subject). These three subjects were excluded from further analyses because their choices revealed that they were not behaving in accordance with the experience-based, incremental type of learning under study[5]. The average α and β values for the remaining participants were then used to estimate their trial-by-trial expected values ($V_{\text{square}(t)}$ and $V_{\text{circle}(t)}$) and prediction errors ($r_t - V_{\text{chosen option}(t)}$). Note that the temporal difference model does not make any assumption about participants' conscious expectations. Rather, expected values estimates reflect latent variables that are necessary for learning to avoid pain, but the conscious or unconscious nature of this learning process remains unspecified.

Logistic regression model. Participants' choices were also analyzed with a logistic regression model predicting the chances of switching choices as a function of pain delivered over the six previous trials.

fMRI data analyses. *Model-based PE analysis.* Statistical analyses were conducted using the general linear model framework implemented in SPM8. In a first model, boxcar regressors, convolved with the canonical hemodynamic response function, modeled periods of decision (onset of decision period to response; mean reaction time = 732 ± 251 ms), anticipation (4 s), outcome onset (1 s) and outcome period (8 s). The decision to use the first 1 s of the stimulation as representing the onset of the stimulation was based on continuous pain ratings obtained in the first, pre-scan session suggesting that this is the moment when subjects begin to feel the thermal stimulation (see **Supplementary Fig. 1**). Outcome (pain = 1, no stimulus = −1) and aversive PE estimates were added as parametric modulators on all regressors (SPM orthogonalization option turned off). The inter-trial interval was used as an implicit baseline. The six runs were concatenated for each subject. A high-pass filter of 180 s was used. Other regressors of non-interest (nuisance variables) included (i) dummy regressors coding for each run (intercept for each run); (ii) linear drift across time within each run; (iii) the six estimated head movement parameters (x, y, z, roll, pitch and yaw), their mean-zeroed squares, their derivatives and squared derivative for each run (total 24 columns); and (iv) indicator vectors for outlier time points identified on the basis of their multivariate distance from the other images in the sample (see above).

Results were cluster-corrected ($P < 0.05$, FWER, two-tailed) with cluster-defining thresholds of $P < 0.001$, $P < 0.01$ and $P < 0.05$ using AFNI's alphasim.

ROI axiomatic response profile analysis. In a second set of analyses aiming to characterize the profiles of activation across outcomes and expected probability of pain, trials within each type of outcome were binned into quartiles of expected probability of pain, resulting in 8 types of outcomes: 2 outcomes (pain or no stimulus) × 4 quartiles (from least to highest expected probability of pain). Mean activity was then extracted for each of the 8 regressors within either a priori PAG and VS ROIs, or ROIs defined by the conjunction analysis (see below). The PAG a priori ROI was constructed by aligning three overlapping 6-mm spheres along the central aqueduct (in mm [0 −24 −4; 0 −26 −6; 0 −29 −8]) and closely matched the findings of a recent meta-analysis on pain processing in the PAG[40]. The VS ROI was based on the Rutledge *et al.* (2010)[9] nucleus accumbens ROI and comprised three 5-mm spheres for each hemisphere (in mm [8 13 −3; 12 13 −8; 9 13 −7; −8 13 −3; −12 13 −8; −9 13 −7]).

To test whether or not activity profiles in the PAG and VS ROIs integrate outcome information with prior expectations into an aversive PE signal, we used an axiomatic approach initially developed to test necessary and sufficient activity for a broad class of PE models. Because our objective was to identify regions that encode aversive PEs, we specialized these axioms for the case of an aversive prediction error and a learned, continuously graded punishment expectancy by making particular assumptions about the sign and monotonicity of effects. This approach reduces the quite general axioms to more familiar algebraic tests, notably separate tests for magnitude and expectation effects, which correspond to the two algebraic components of PEs (PE = magnitude − expectation; see also refs. 6,8 for a similar approach). Moreover, in addition to the reward and expectancy components tested in those studies, we also test a third axiom, which specifies that expectation and magnitude effects are properly registered to one another, resulting in identical response amplitudes when outcomes are fully predicted. Finally, one difference between our axiomatic approach and the one previously used by Rutledge *et al.*[9] is that we define reward expectancy as estimated from the fit of learning model to choice behavior. This is because, in our task, participants' expectations were not explicitly instructed but were instead derived from their reinforcement history and therefore had to be computationally estimated before being used to test the axioms for PEs.

The three axiomatic tests used to identify aversive PE signals are described below. They are derived from the more general axioms of Caplin and Dean[50] by introducing specific, plausible assumptions about the aversive case—namely, that higher pain is always more aversive than lower pain (monotonicity in pain intensity) and that high versus low expectancies are similarly monotonic. With these assumptions, axiom 1 (consistent prize ordering: the outcome effect) stipulates that activity for pain outcomes should be higher than for no-stimulus outcomes. This axiom was tested by a simple *t*-test of the difference between averaged values for the four pain and four no-pain quartiles. Axiom 2 (consistent lottery ordering: the expectancy effect) stipulates that activity should decrease with increasing expected probability of pain. This axiom was tested by separately testing the slopes of regressions lines passing through the four quartiles for pain and no-pain trials, using a nonparametric multilevel sign permutation test (1,000 bootstrap samples). Finally, axiom 3 (no surprise equivalence: that the expectancy and outcome effects have the correct relationship to one another) stipulates that completely predicted outcomes should generate equivalent responses. This axiom was tested by a simple *t*-test comparing activity for the highest quartile of expected pain for pain trials and lowest quartile of expected pain for no-pain trials. We note that pain adaptation processes such as sensitization and habituation can cause pain itself to behave like PEs in some respects; for example, both pain and aversive PEs may decrease across trials and vary inversely with the intensity of prior pain[51], causing a stronger partial overlap between aversive PE signals and pain itself. However, this effect explains only some of the effects tested in axiom 2 (those on pain trials). It does not account for the effects tested under axiom 1 or axiom 3, or effects on no-pain trials tested under axiom 2. In addition, it does not account for experimental effects such as effects of placebo instructions, tested in study 3. Thus, the axiomatic tests provide a strong test of aversive PE-related signal properties.

Conjunction analysis. To specifically test for the expected probability of pain within pain and no-stimulus trials (axiom 2), we modeled separately pain and no-stimulus trials and included expected probability of pain as parametric modulators for pain and no-stimulus trials. We then looked at the conjunction between the three relevant contrast maps (pain > no stimulus, expected probability of pain within pain trials, expected probability of pain within no-stimulus trials), which were cluster-corrected ($P < 0.05$, FWER, one-tailed) with a cluster-defining threshold of $P < 0.05$ using AFNI's alphasim.

Dynamic causal models. To explore how the seven different regions identified in the previous analysis (aversive PE: PAG; pain-specific PE: aMCC, OFC, dmPFC; expected value: vmPFC, putamen, hippocampus) interacted to generate aversive PE signals, we compared several probable dynamic causal models (DCMs) with a Bayesian model selection procedure[27]. On the basis of the principles governing reinforcement learning models (**Fig. 2a**), regions that encode aversive PEs (PAG) should receive converging input from those that encode expectancies (vmPFC, putamen, hippocampus) and primary reinforcement (nociceptive) signals. Regions important for action value and decision-making (aMCC, OFC, dmPFC) may receive converging PE and primary reinforcement signals.

Because of the large number of possible models, we began by defining a model limited to brain correlates of reinforcement learning model–based effects (PAG, vmPFC, putamen, hippocampus). We constrained this model by making two assumptions: (i) primary nociceptive afferents directly project to PAG[28], and (ii) expected avoidance value is conveyed to the PAG through one or more of the three expected value structures (green in **Fig. 3**). We used Bayesian model selection to evaluate 32 plausible models, which varied systematically in their projections to the midbrain and connections among expected value-related regions (see **Supplementary Fig. 6**), and tested the most likely model against 7 other close variants[27] (see **Supplementary Fig. 7**). Overall, the most likely configuration given our data is shown in **Figure 4** (black and green portions only). In this model, vmPFC projects most directly to PAG, and avoidance value is most closely related to the putamen, which transmits value information to vmPFC. Though this procedure cannot definitively isolate causal relationships among regions, this model provides a plausible working model considering the direct anatomical projections from vmPFC to PAG[30].

We then extended the model to include other regions that may encode avoidance value and related properties. On the basis of existent animal models of fear conditioning[1,18], we posited that these regions receive the PE signals generated in PAG. Moreover, on the basis of known anatomical projections of the PAG, we constrained the space of possible models by assuming that PE signals could be directly conveyed to the OFC and aMCC[32], and indirectly to the dmPFC through either aMCC or OFC. However, as the sources of expected value signals to these regions are less informed by the existent literature, we allowed these regions to be functionally connected to any of the three regions encoding expected value signals (that is, vmPFC, putamen and hippocampus). Within these constraints, we evaluated 27 models that systematically varied connections among avoidance updating–related regions (blue) and relationships with expected value–related regions (green; see **Supplementary Fig. 8**) and 16 additional models closely related to the best-fitting model and including modulatory nociceptive inputs (see **Supplementary Fig. 9**). The best model overall (**Fig. 4**) included (i) direct connections from both putamen and midbrain to OFC and aMCC, with dmPFC effects mediated by OFC, and (ii) modulatory effects of noxious input to putamen → OFC, putamen → aMCC, and midbrain → aMCC connections.

Study 2: comparison with reward prediction errors. *Participants.* Twenty-one participants (mean age, 19.3 years; range, 18–28; ten female) took part in the study. Informed consent was obtained in a manner approved by the New York University Committee on Activities Involving Human Subjects.

Monetary reward task. In the experimental task (**Supplementary Fig. 3a**,**b**; see also ref. 23), on each of 300 trials, participants chose one of four presented face stimuli and then received monetary feedback. Participants then received binary reward feedback, a $0.25 'win' outcome represented by an image of a quarter-dollar and a $0.00 'miss' outcome represented by a phase-scrambled image of a quarter-dollar. Participants were instructed that each face option was associated with a different probability of reward, that these probabilities could change slowly, and that their goal was to attempt to find the most rewarding option at a given time to earn the most money. Across the 300 trials in the experiment, the reward probabilities diffused gradually according to Gaussian random walks, so as to encourage continual learning. Unbeknownst to the participants, the faces were grouped into equivalent pairs.

doi:10.1038/nn.3832

Imaging procedure. Whole-brain imaging was conducted on a 3.0-T Siemens Allegra head-only MRI system at NYU's Center for Brain Imaging, using a Nova Medical NM-011 head coil. Functional images were collected using a gradient echo T2*-weighted echoplanar (EPI) sequence with BOLD contrast (TR = 2,000 ms, TE = 15 ms, flip angle = 82, 3 × 3 × 3 mm voxel size; 33 contiguous oblique-axial slices), tilted on a per-participant basis approximately 23 degree off of the AC–PC axis to optimize sensitivity to signal in the orbitofrontal cortex and the medial temporal lobe. The task was scanned in four blocks each of 310 volumes (10 min 20 s).

Behavioral analysis. Participants' choices were analyzed with a similar temporal difference model to the one used for the analysis of pain-related aversive PEs, with the exception that an additional parameter accounted for the generalization of learned values from one face of a pair to the other.

Imaging analyses and results. Preprocessing and data analysis was performed using Statistical Parametric Mapping software (SPM5; Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). After realignment and normalization, images were resampled to 2-mm cubic voxels, smoothed with an 8-mm FWHM Gaussian kernel, and filtered with a 128-s high-pass filter. To identify the structures encoding appetitive prediction errors, activity at outcome delivery was correlated with the trial-by-trial reward PE estimates derived from the computational temporal difference model. Finally, we extracted the mean activity related to these appetitive PEs in the periaqueductal gray (PAG) and ventral striatum (VS) regions of interest (ROIs) used in previous analyses (see **Fig. 2**) and compared it to pain-related aversive PE signals (see **Supplementary Fig. 3**) extracted from the same ROIs.

Study 3: comparison of different pain levels. *Participants.* Fifty healthy participants completed the study (mean age = 25.1, range = 18–52 years; 27 females). All participants gave informed consent and the experiment was approved by the institutional review board of the University of Colorado Boulder.

Thermal stimulation. Thermal stimulation was delivered to the volar surface of the left inner forearm using a 16 × 16 mm Peltier thermode (Medoc). Each stimulus lasted 11 s with 1.75-s ramp-up and ramp-down periods and 7.5 s at target temperature. Stimulation temperatures were 46, 47 and 48 °C, and in between stimuli the thermode maintained a baseline temperature of 32 °C.

Experimental task. This pain-learning task consisted of 6 runs of 8 trials each and alternated between placebo and control runs (in counterbalanced order). In the placebo runs, the thermode was placed on a skin site that had been pretreated with a placebo analgesic cream. In the control runs, the thermode was placed on a site that had not been pretreated. During both the placebo and control runs, participants were presented with two visual cues (geometric shapes). One cue was always followed by a 46 °C (low pain) or a 47 °C (medium pain) thermal stimulus and the other cue by a 47 °C (medium pain) or a 48 °C (high pain) thermal stimulus, in 50% of the trials each (see **Supplementary Fig. 4a**). Participants were not informed about these contingencies.

Each trial started with the presentation of the two cues randomly displayed at the left and right side of the screen for 4 s, during which participants selected the cue that they thought was predictive of the least pain, by means of a left or right button press. One to 3 s later the computer selected a cue, alternating between the high and the low cue. The computer's selection was shown for 3 s and was immediately followed by a thermal stimulation. Note that the stimulation temperature was contingent on the computer's—not the participant's—cue selection. Nine to 13 s after the thermal stimulation, a pain-rating scale was presented for 6 s, and participants rated their experienced pain using a trackball. The rating period was followed by a 9–13 s inter-trial interval. During the stimulation, post-stimulation and inter-trial intervals, a fixation cross was presented at the center of the screen.

Imaging procedure. Whole-brain fMRI data were acquired on a Siemens 3-T Trio scanner at the Center for Innovation and Creativity (CINC) in Boulder. Functional images were acquired with an echo-planar imaging sequence (TR = 1,300 ms, TE = 25 ms, field of view = 220 mm, 3.4 × 3.4 × 3.0 mm voxels, 26 slices). Each run lasted 394 s (303 TRs).

Imaging analyses and results. The preprocessing procedure was identical to the one used in the main experiment (see above). Boxcar regressors, convolved with the canonical hemodynamic response function, were constructed to model (i) the periods in which visual stimuli other than the fixation cross were presented (that is, the cues and the rating scale), (ii) participants' cue-selection times, and (iii) participants' pain-rating times. Because the onset of the stimulation is uninformative of the pain level participants received, we used continuous pain ratings for three levels of thermal stimulations of identical durations (11 s; 46.5 °C, 47.5 °C, 48.5 °C) to identify the time at which the different temperatures could be clearly distinguished. We identified the period between 4 and 10 s as the one conveying information about the pain level received, and we therefore modeled thermal stimuli as three successive time-windows: (iv) onset (0–4 s), (v) middle (4–10 s) and (vi) offset (10–11 s) (see **Supplementary Fig. 4d**).

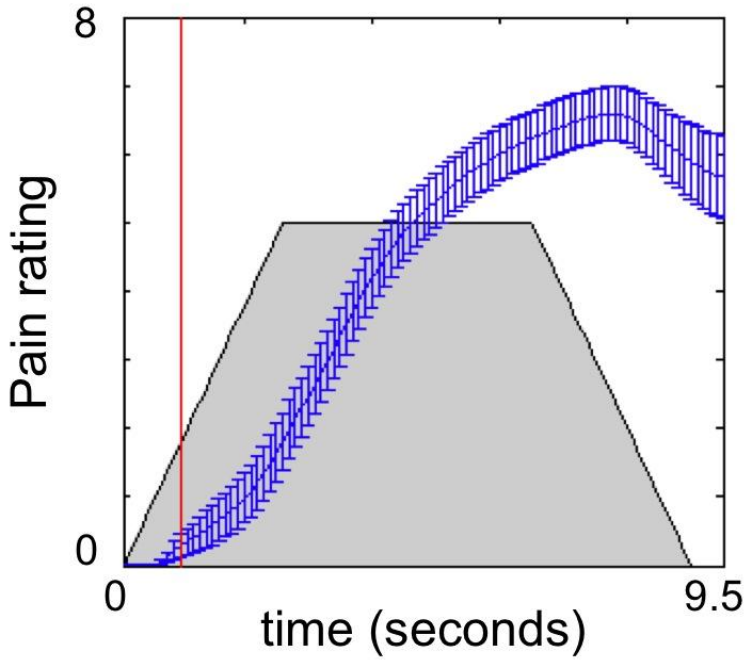
We then extracted mean activity in the PAG ROI (see **Fig. 2** and **Supplementary Fig. 4c**) during this middle, pain-informative window for the three different levels of temperature, the two different levels of predictive cues and the placebo versus control condition. To test whether the PAG also encoded aversive PEs in an intensity-dependent manner, we adapted study 1 axioms by making the additional assumption that more intense noxious stimulus intensities should be more aversive (that is, monotonicity of aversiveness with stimulus intensity).

To test axiom 2, we looked more specifically at activity in response to the medium temperature, which could be preceded by either low or high predictive cues. Axiom 2 requires that aversive PEs should be higher when less pain is predicted. In the current experiment, this should translate into higher activity for low versus high cues. Moreover, if PEs are also sensitive to explicit predictions about pain, PEs should be higher during the placebo versus control condition (see **Supplementary Fig. 4b**).

Finally, axiom 3 stipulates that there should be no difference in signal strength between fully expected outcomes of different intensities. Unfortunately, this axiom cannot be fully tested here because the outcome is never fully predicted by the cue (50%–50%), and the nonlinear relationship between temperature and pain makes it difficult to precisely estimate expectations. Minimally, there should be a partial overlap between low and high cue lines allowing certain temperature levels to be associated with equivalent prediction error signals, which again entails that responses to the medium temperature should be higher for low versus high cues.

A **Supplementary Methods Checklist** is available.

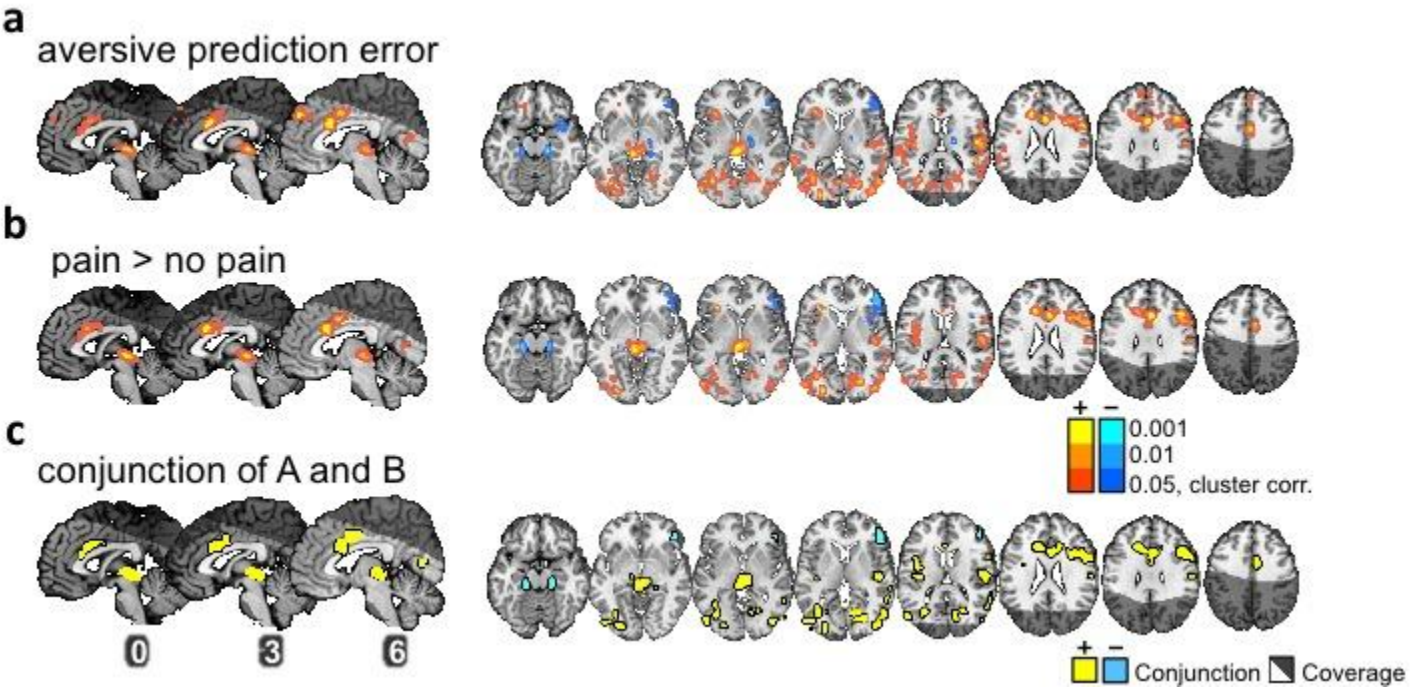
49. Glover, G.H. & Law, C.S. Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Mag. Reson. Med.* **46**, 515–522 (2001).
50. Caplin, A. & Dean, M. Axiomatic methods, dopamine and reward prediction error. *Curr. Opin. Neurobiol.* **18**, 197–202 (2008).
51. Jepma, M., Jones, M. & Wager, T.D. The dynamics of pain: evidence for simultaneous site-specific habituation and site-nonspecific sensitization in thermal pain. *J. Pain* **15**, 734–746 (2014).



Supplementary Figure 1

Time course of subjective pain perception

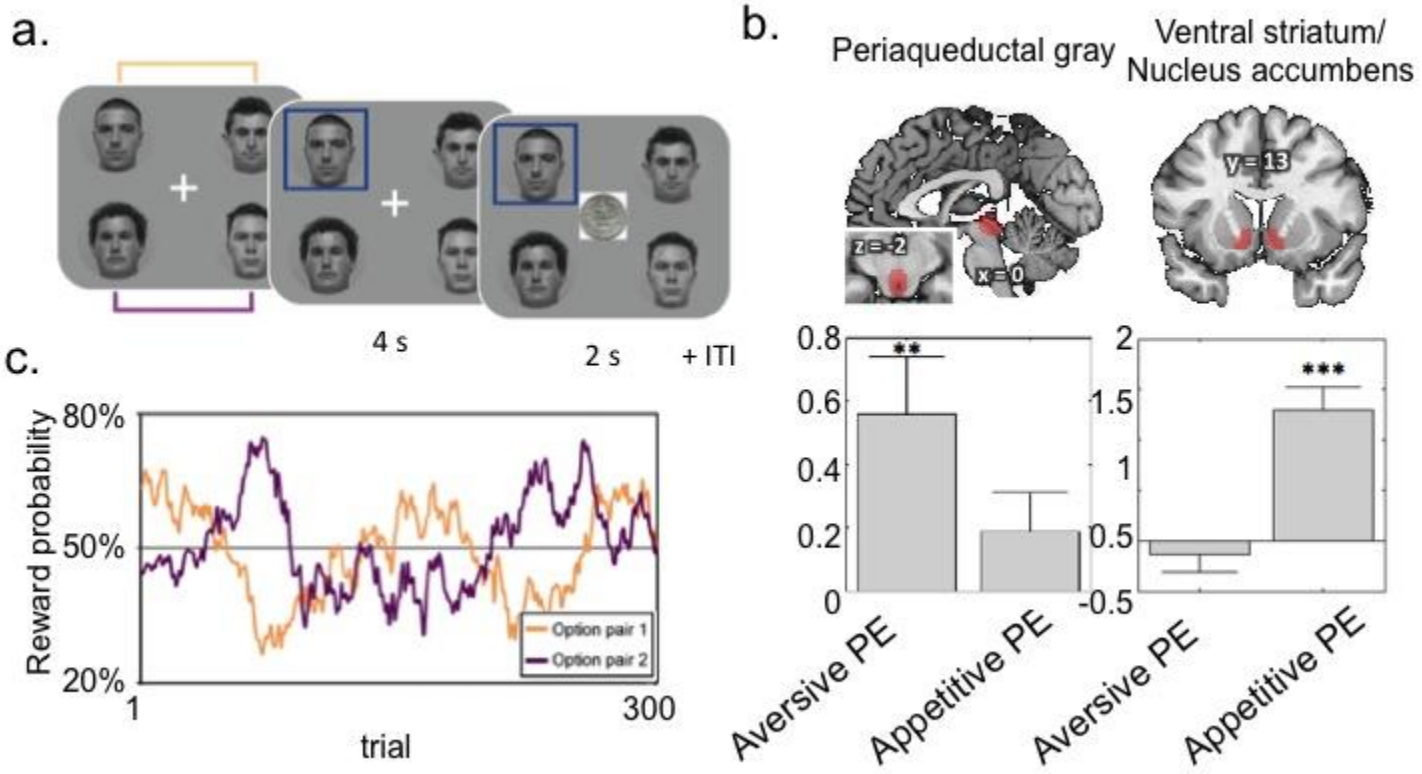
Mean online pain ratings obtained during the behavioral session superimposed on the temporal profile of the thermal stimulus (number or participants = 23; 4 seconds plateau, 2.5 seconds ramp-up/ramp-down). Ratings begin to rise in the first second of the stimulation (left of the red vertical line).



Supplementary Figure 2

Activity at outcome onset

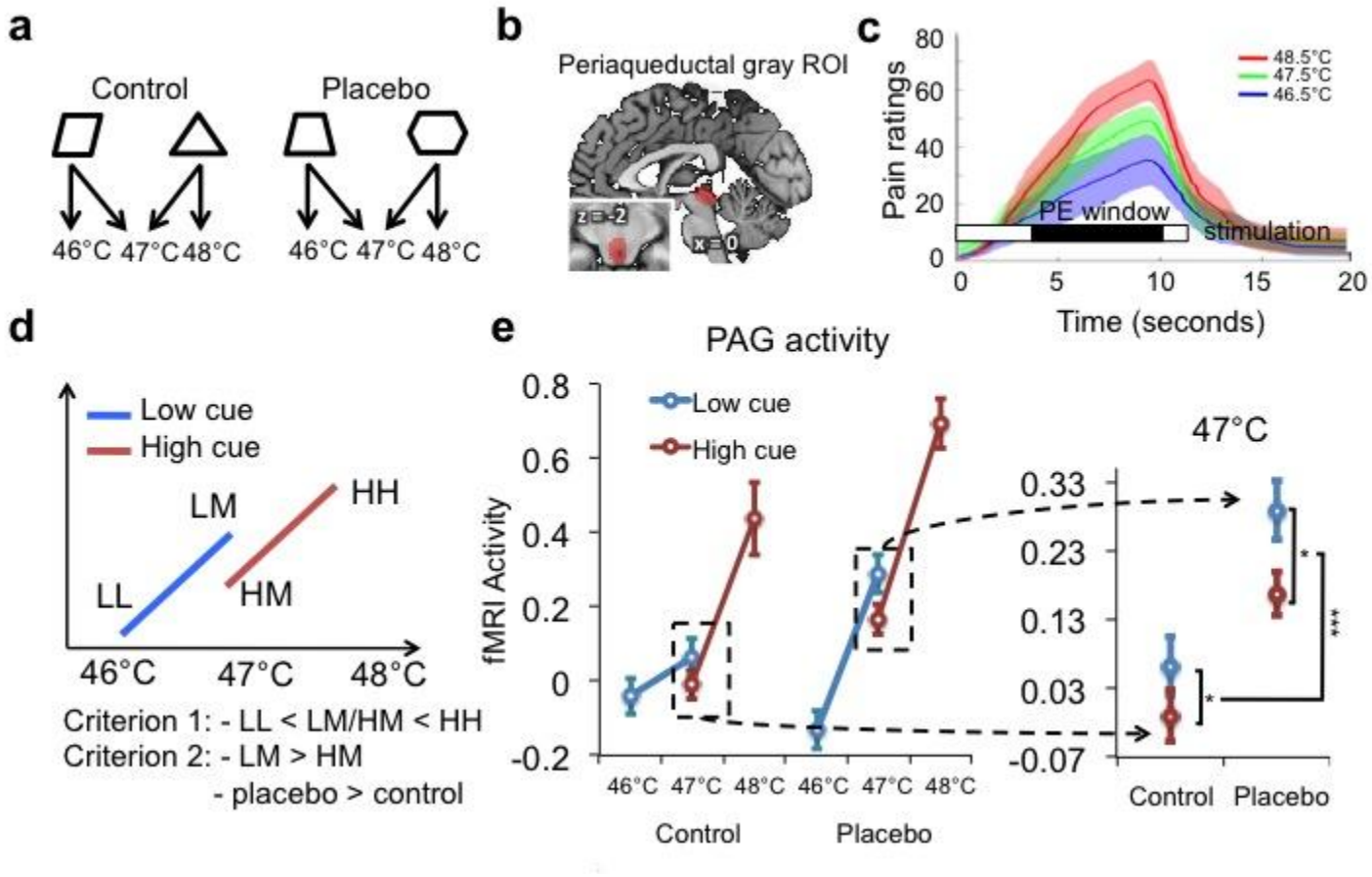
Activity at outcome onset (1st second of outcome; number or participants = 23) related to (A) model-based aversive prediction error (outcome worse than expected), (B) pain > no stimulus. Displayed activations are cluster-thresholded ($p<0.05$, FWE, two-tailed) with cluster-defining thresholds of $p<0.001$,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

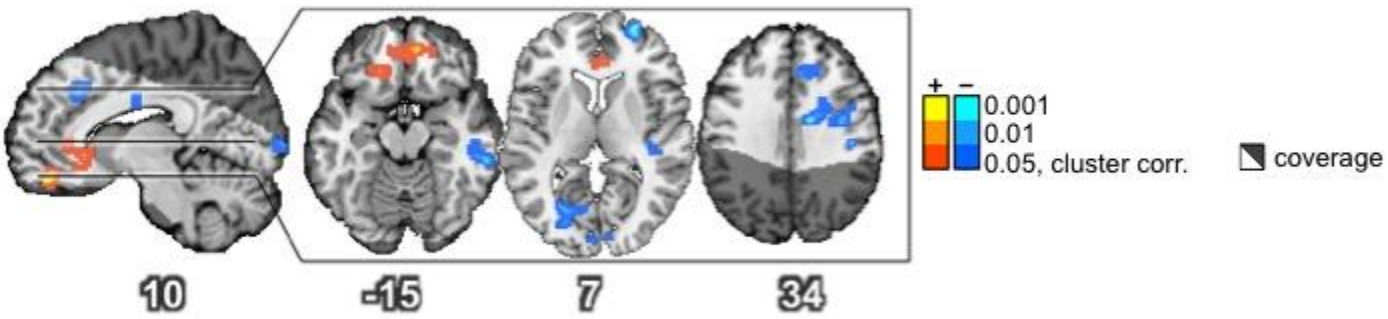
A) On each trial, participants (n = 24) chose one of four face options. After a delay, the outcome ($0.25 or $0.00) was revealed. Faces are paired together such that the probability of receiving a reward on a given trial is the same for both faces of the pair. In colored brackets, one example of option 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

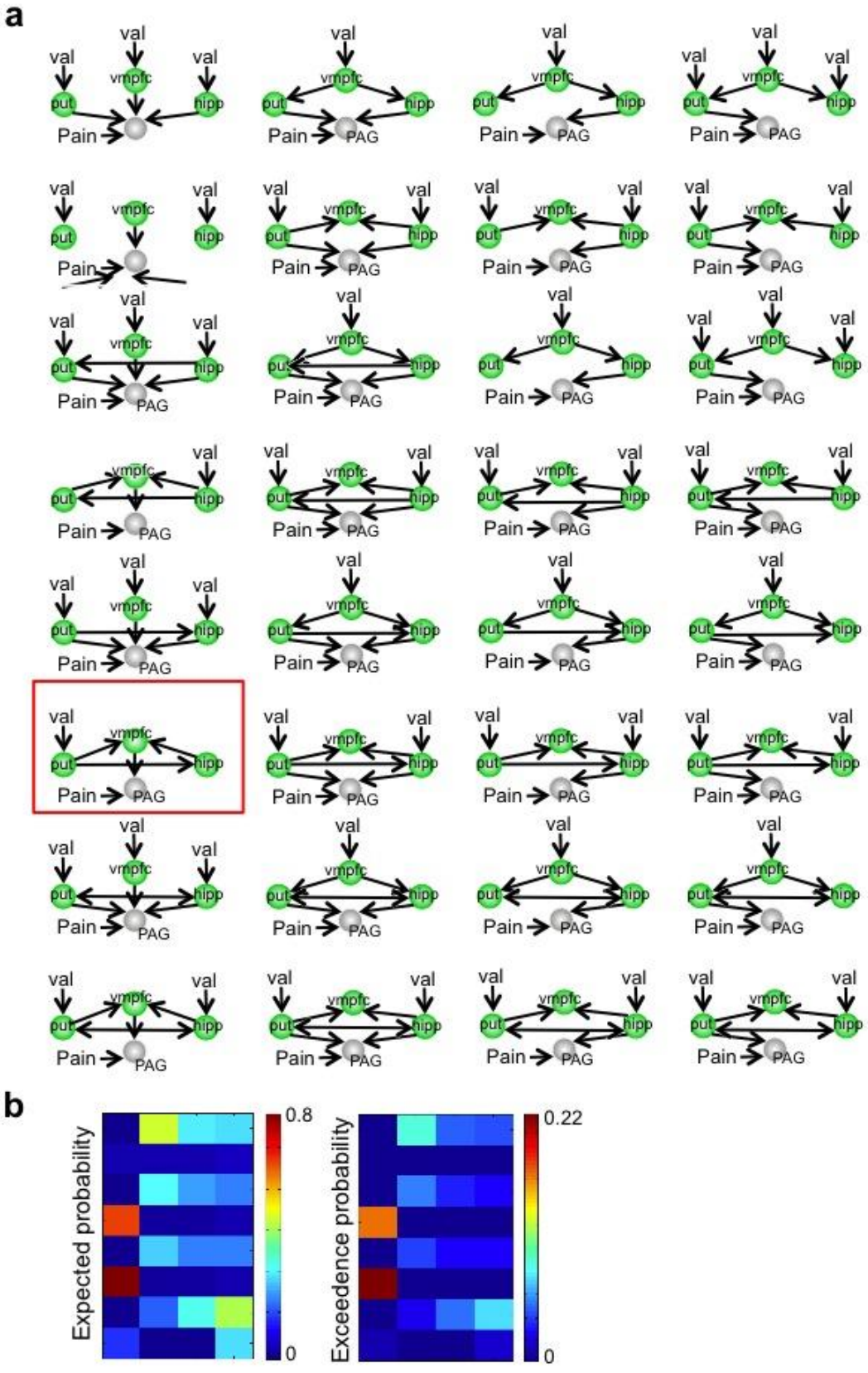
A) Experimental conditions. In both the control and placebo runs, participants (n = 50) are presented two predictive cues. The low cue is followed 50% of the time by low pain (46°C) and 50% of the time by medium pain (47°C). The high cue is followed 50% of the time by high pain (48°C) and 50% of the 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

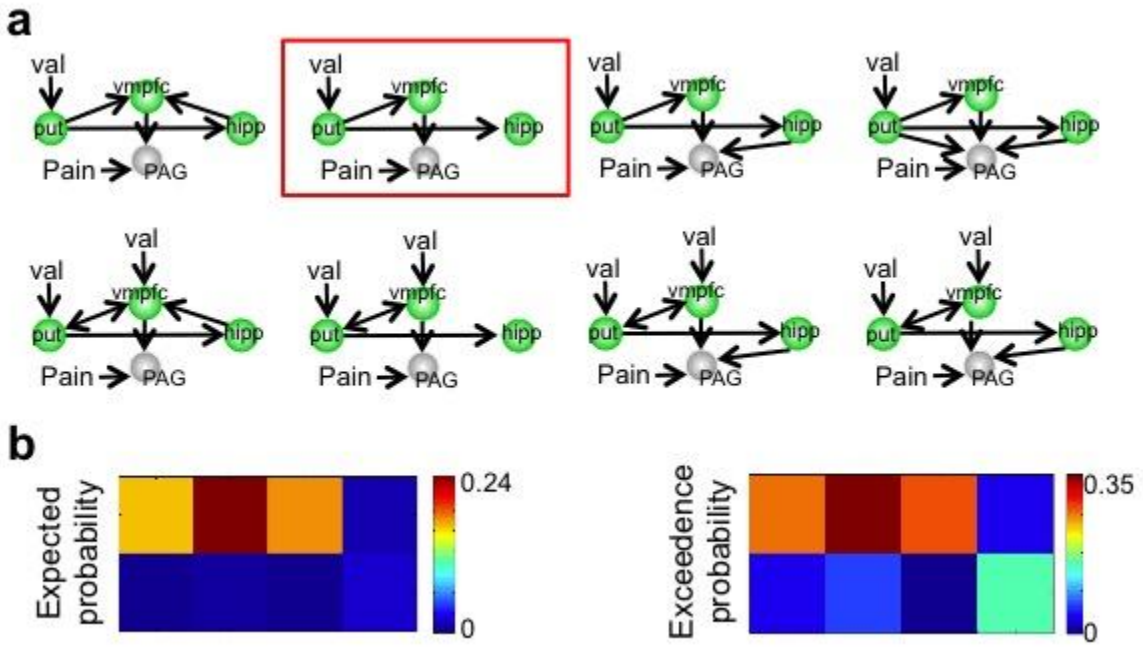
Activity during decision-making (number or participants = 23) correlating positively (red) or negatively (blue) with the expected value of the chosen option (warm/cold colors indicate low/high subjective (model-based) probability of pain. Displayed activations are cluster-thresholded ($p<0.05$, FWE,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


a
val
vmpfc
put
hipp
Pain
PAG
b
Expected probability
0.8
0
Exceedence probability
0.22
0

Supplementary Figure 6

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

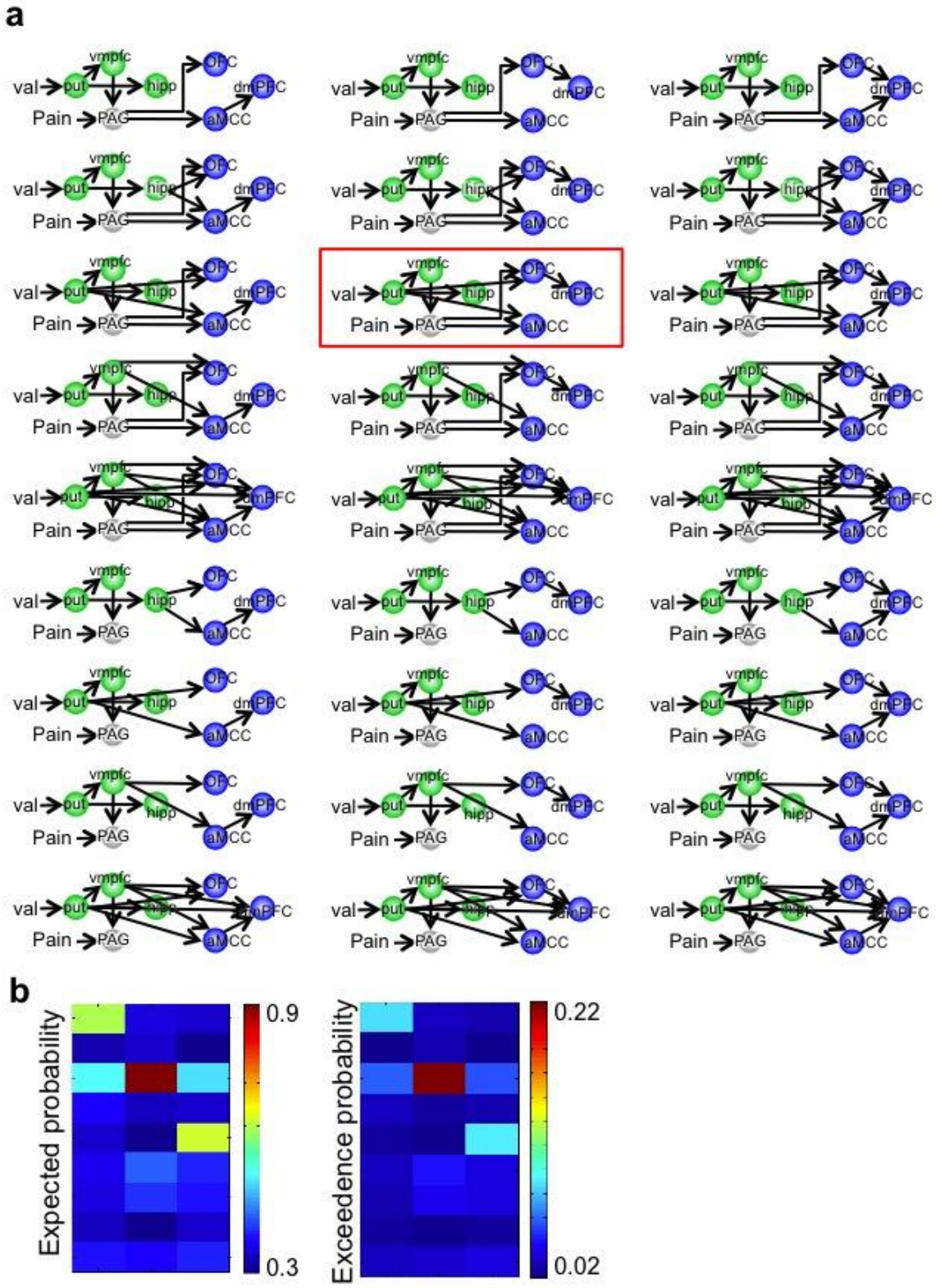
(A) In all models, the driving inputs are the pain > no stimulus and expected value parametric modulators on outcome onsets. The models tested systematically varied the structure(s) receiving the expected value driving inputs and conveying this information to the midbrain. The model with the highest 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 or 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 or participants = 23.

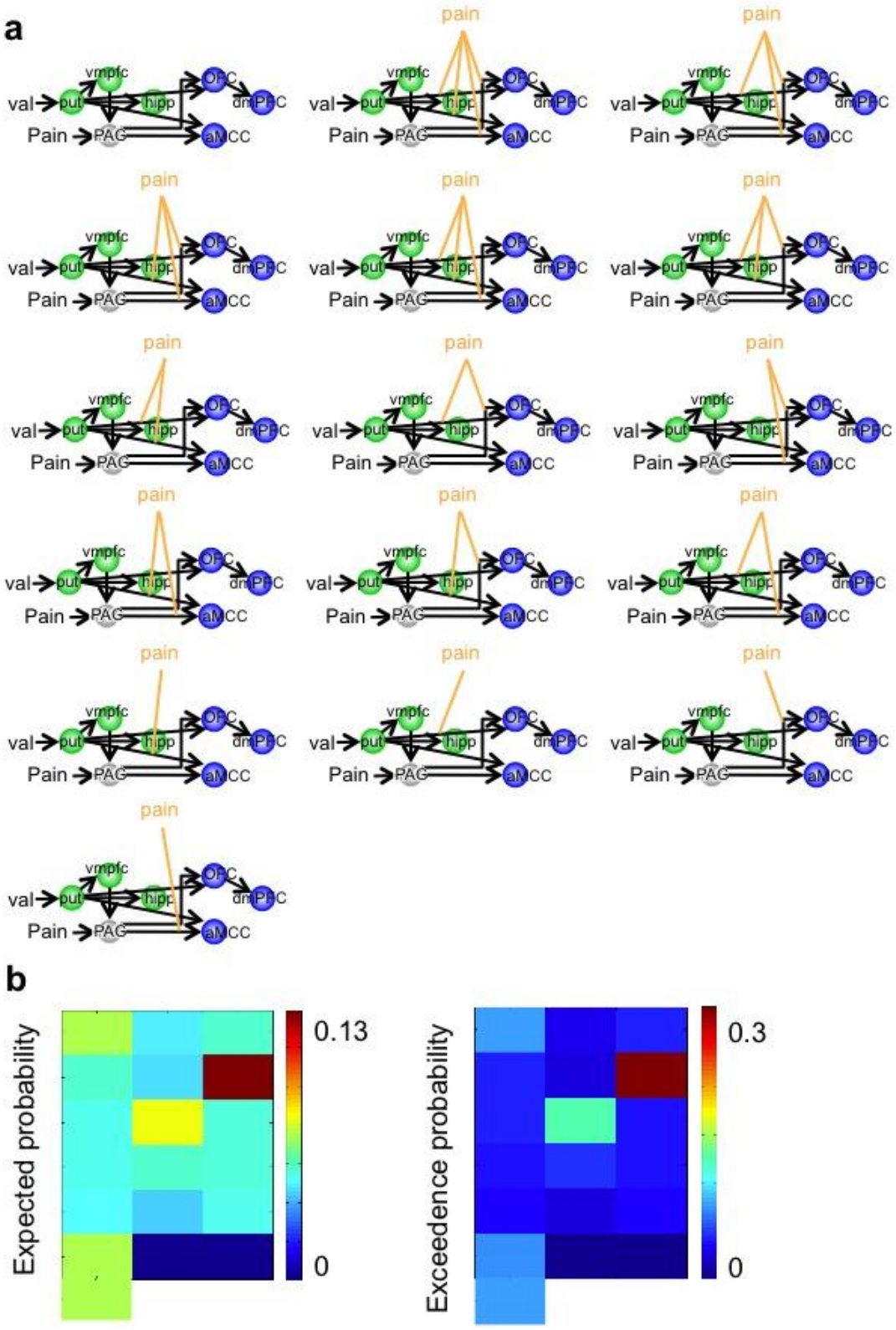


Nature Neuroscience: doi:10.1038/nn.3832

Supplementary Figure 8

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

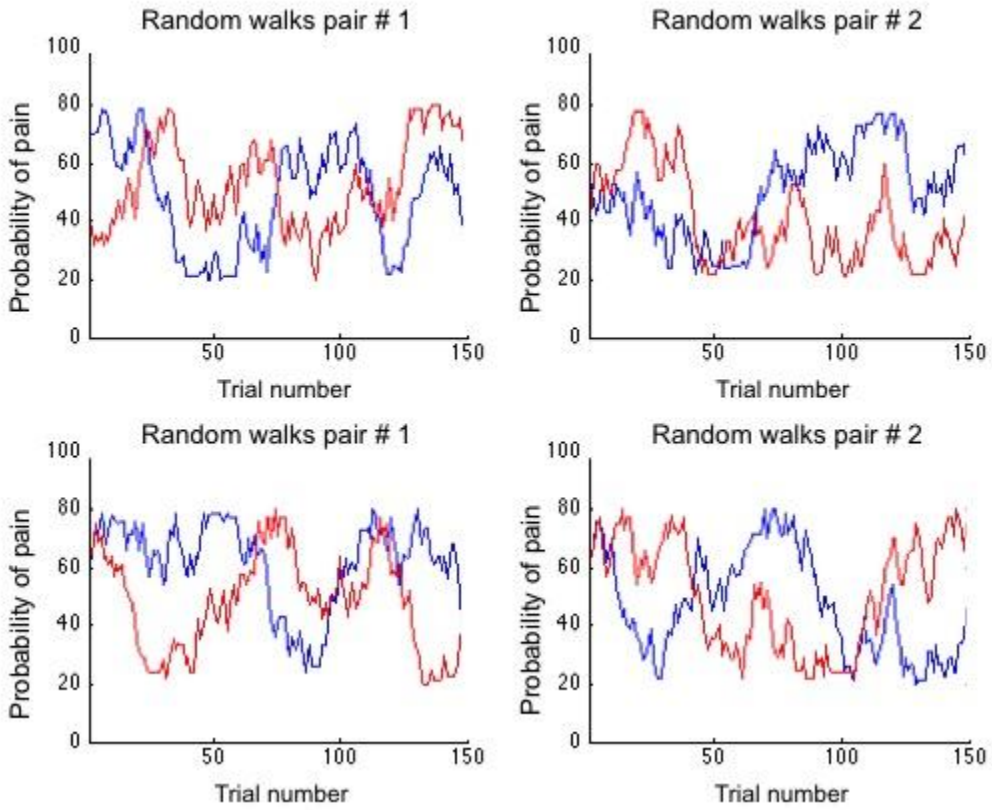
(A) In all models, the structures generating PE signals (put, vmPFC, hipp, PAG) are arranged according to the best model selected from the previous model selection steps. The links between these structures and the pain-specific PE structures are systematically varied. The model with the highest exceedance 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 or participants = 23.


a
pain
vmpfc
val
put
hipp
OFC
dmPFC
Pain
PAG
aMCC
b
Expected probability
0.13
0
Exceedence probability
0.3
0

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.