Different brain systems support the aversive and appetitive sides of human pain-avoidance learning

Marieke Jepma^{1,2,3*}, Mathieu Roy^{4,5}, Kiran Ramlakhan^{2,6}, Monique van Velzen⁷, Albert Dahan⁷

¹ Department of Psychology, University of Amsterdam, the Netherlands

² Department of Psychology, Leiden University, the Netherlands

³ Leiden Institute for Brain and Cognition, the Netherlands

⁴ Department of Psychology, McGill University, Montreal, QC, Canada

⁵ Alan Edwards Centre for Research on Pain, McGill University, Montreal, QC, Canada

⁶ Pharos - Dutch Centre of Expertise on Health Disparities, Utrecht, the Netherlands

⁷ Department of Anesthesiology, Leiden University Medical Center, Leiden, the Netherlands

* corresponding author

Abstract

Both unexpected pain and unexpected pain absence can drive avoidance learning, but whether they do so via shared or separate neural and neurochemical systems is largely unknown. To address this issue, we combined an instrumental pain-avoidance learning task with computational modeling, functional magnetic resonance imaging (fMRI) and pharmacological manipulations of the dopaminergic (100 mg levodopa) and opioidergic (50 mg naltrexone) systems (N=83). Computational modeling provided evidence that untreated participants learned more from received than avoided pain. Our dopamine and opioid manipulations negated this learning asymmetry by selectively increasing learning rates for avoided pain. Furthermore, our fMRI analyses revealed that pain prediction errors were encoded in subcortical and limbic brain regions, whereas no-pain prediction errors were encoded in frontal and parietal cortical regions. However, we found no effects of our pharmacological manipulations on the neural encoding of prediction errors. Together, our results suggest that human pain-avoidance learning is supported by separate threat- and safety-learning systems, and that dopamine and endogenous opioids specifically regulate learning from successfully avoided pain.

Introduction

Learning to avoid actions that cause damage to our body is critical for health and survival. The experience of pain is an important teaching signal in this learning process, such as when a child learns to avoid touching a hot stove, or when a patient who underwent knee surgery learns to avoid bending his or her knee. However, the *absence* of otherwise expected pain can be an equally important teaching signal. When, for example, some weeks after surgery a patient realizes that bending his or her knee is *not* painful anymore, this suggests that particular movements are safe again and no longer needs to be avoided. Adaptive behavior in situations associated with pain thus requires an optimal balance between threat and safety learning when confronted with, respectively, the unexpected presence and absence of pain.

Previous studies have made considerable progress in our understanding of the neural basis of passive cue-pain-association learning (Ploghaus et al. 2000, Seymour et al. 2004, Seymour et al. 2005) and—more recently—active pain-avoidance and -relief learning (Roy et al. 2014, Eldar et al. 2016, Zhang et al. 2018) in humans. A key aspect of these studies was the application of reinforcement-learning models to the analysis of neuroimaging data. According to reinforcement-learning theory, learning is driven by *prediction errors*, which signal the difference between the actual and expected outcomes of an action (Sutton and Barto 1998). Thus, actions that result in the unexpected presence versus absence of pain yield oppositely signed prediction errors which, respectively, increase and decrease the aversive value associated with that action. However, whether these opponent teaching signals drive learning via one underlying brain system, or via separate ones, is still largely unknown.

One possibility is that prediction errors elicited by the unexpected presence and absence of pain are encoded as opposite activity patterns in the same brain regions (i.e., one learning system). If this is the case, we may also expect that—at the neurochemical level—learning from these two outcomes is supported by the same neuromodulator(s). Furthermore, these two outcomes may then be equally effective in driving learning, such that they are associated with the same learning rate. Most previous studies, including our own, have assumed that this is the case. For example, in a previous functional magnetic resonance imaging (fMRI) study, we identified brain activity encoding *general* aversive prediction errors, signaling the degree to which *both* received- and avoided-pain outcomes are relatively worse (or less good) than expected (Roy et al. 2014). Another possibility, however, is that learning from received and successfully avoided pain are subserved by two separate brain

systems. In this case, learning from these two outcomes may also be supported by different neuromodulatory systems, and associated with different learning rates. Although the notion of two systems is broadly consistent with theoretical accounts of avoidance learning, such as two-factor theory (Mowrer 1951), there is not much empirical evidence for this idea, especially not in humans.

Regarding the role of neuromodulators, there is a wealth of evidence that reward prediction errors, which are thought to drive reward-pursuit behaviors, are signaled by the phasic activity of midbrain dopamine neurons (Schultz et al. 1997) that project to the ventral striatum (O'Doherty et al. 2003, Rutledge et al. 2010). Whether the dopamine system also has a role in aversive learning is controversial, but several hypotheses have been proposed, mostly based on animal studies (Palminteri and Pessiglione 2017). According to one prominent account, the same dopaminergic prediction-error response that reinforces actions associated with reward also reinforces actions associated with the omission of aversive outcomes (Mowrer 1956, Dinsmoor 2001, Moutoussis et al. 2008). This account thus predicts a role for dopamine in learning from successfully avoided pain. However, animal studies have provided mixed evidence for this idea (Josselyn et al. 2005, Oleson et al. 2012, Wietzikoski et al. 2012, Dombrowski et al. 2013, Fernando et al. 2013, Salinas-Hernandez et al. 2018, Wenzel et al. 2018, Stelly et al. 2019) and evidence in humans is scarce (Raczka et al. 2011).

The endogenous opioid system is another interesting candidate neuromodulatory system for pain-avoidance learning. Pavlovian fear conditioning studies in animals have provided evidence for a causal role of opioidergic activity—specifically via μ -opioid receptors in the periaqueductal gray (PAG)—in both fear conditioning (McNally and Cole 2006, Cole and McNally 2007, Cole and McNally 2009, McNally et al. 2011) and fear-extinction learning (McNally and Westbrook 2003, McNally et al. 2004, McNally et al. 2005, Kim and Richardson 2009, Parsons et al. 2010). These findings suggest that the endogenous μ -opioid system may mediate learning from received and/or avoided pain, although this remains to be studied in instrumental learning tasks and in humans.

In the present study, we aimed to dissociate learning from received and avoided pain in terms of behavior (learning rates), neural encoding of prediction errors, and the roles of dopamine and endogenous opioids. To this end, we combined an instrumental pain-avoidance learning task with computational modeling, fMRI, and pharmacological manipulations of the dopamine and opioid systems, in a randomized, double-blind, between-subject design. Specifically, participants completed the learning task under one of three drug conditions: 100

mg levodopa (a dopamine precursor), 50 mg naltrexone (an opioid antagonist, with highest affinity for the μ -opioid receptor), or placebo. PET studies in humans suggest that levodopa increases phasic dopamine bursts (Floel et al. 2008), but not tonic dopamine activity (Black et al. 2015). Thus, if dopaminergic prediction-error responses support learning from successfully avoided pain, we expect levodopa to enhance learning rates and neural prediction error signaling when pain is avoided. Naltrexone blocks the majority of μ -opioid receptors in the brain (Lee et al. 1988, Preston and Bigelow 1993, Schuh et al. 1999, Weerts et al. 2013). Thus, if μ -opioid receptor activity supports learning from received and/or avoided pain outcomes, we expect naltrexone to reduce learning rates and neural prediction error signaling for the corresponding outcome(s).

Results

Eighty-three healthy human participants completed a pain-avoidance learning task during fMRI, under one of three treatment conditions (levodopa, naltrexone, or placebo). On each of 144 trials of the pain-avoidance learning task, participants chose between two options (Figure 1A), each probabilistically associated with the delivery of painful heat (49 or 50°C, 1.9 s duration) to their left lower leg. Pain probabilities for each option were governed by two independently varying random walks (Figure 1B). Thus, participants had to track the changing reinforcement values, and kept experiencing prediction errors throughout the task.

Figure 1. Pain-avoidance learning task. **A.** Outline of one trial. **B.** Example of pain probabilities and choice data for one participant. The green and blue lines show the trial-specific probabilities of receiving a painful heat stimulus when choosing each option. The green and blue circles below the graph indicate the participant's choices, and the red stars indicate trials on which pain was delivered. **C.** Probability of switching per treatment group, as a function of pain 1-6 trials back. Error bars are standard errors.

There were no treatment effects on subjective state (alertness, calmness or contentment; Figure 1—figure supplement 1) or self-reported heat pain during a separate pain-rating task immediately preceding the avoidance-learning task (Figure 1—figure supplement 2).

Nine participants were excluded from the fMRI (but not the behavioral) analyses because of excessive head movement. Thus, the behavioral analyses included 83 participants (26-29 per treatment group), and the fMRI analyses included 74 participants (24-26 per treatment group).

No drug effects on model-independent measures of task performance

On average, participants received pain on 57.1 of the 144 trials (SD = 8.6). As expected, participants switched to the other choice option more frequently after receiving pain (46.6% of those trials, SD = 21.2) than after avoiding pain (5.4% of those trials, SD = 6.9; $t(82) = 18.6$, $p < .001$). The effect of previous pain outcomes on switching also decayed exponentially over time, in all treatment groups (Figure 1C; $p < 0.001$ for 1 trial back, $p < .002$ for 2 trials back, and $p > .047$ for 3-6 trials back, in all groups).

The three treatment groups did not differ in the number of received pain stimuli ($F(2,80) = 0.56$, $p = 0.57$), frequency of switching following pain outcomes ($F(2,80) = 0.03$, $p = 0.97$), or frequency of switching following no-pain outcomes ($F(2,80) = 1.18$, $p = 0.31$). Thus, our pharmacological manipulations did not affect basic measures of task performance.

Computational modeling

To formalize and quantify the latent learning and decision processes thought to underlie participants' choice behavior, we applied two candidate reinforcement-learning models to the choice data, using a hierarchical Bayesian approach. Group-level parameters were estimated separately for each treatment group. Both models update the expected pain probability for the chosen option on each trial, in proportion to the prediction error (Rescorla and Wagner 1972). The two models differ in that Model 1 uses a single learning rate, α , for all outcomes, whereas Model 2 uses separate learning rates for received and avoided pain: α_{pain} and $\alpha_{no-pain}$, respectively (see Methods for model equations and parameter-estimation details). If Model 2 is better able to explain the choice data than Model 1, this could be taken as initial support for the idea that learning from received and avoided pain is subserved by different learning systems.

Both learning models were combined with a softmax decision function that translates expected pain probabilities into choice probabilities. Inverse-temperature parameter β controls the degree of choice randomness, such that the likelihood that the model chooses the option with the lowest expected pain probability increases as β increases.

Model comparison: evidence for asymmetric learning from received and avoided pain

In all treatment groups, the model with separate learning rates for received and avoided pain (Model 2) outperformed the model with a single learning rate (Model 1), providing initial evidence for the presence of two learning systems. The Watanabe-Akaike information criterion (WAIC) for Model 1 vs. Model 2 was 3109 vs. 2959, 2958 vs. 2941, and 3164 vs. 3106 for the placebo, levodopa and naltrexone groups, respectively.

Parameter estimates: levodopa and naltrexone increase learning rates for avoided pain

We next examined the parameter estimates of the best fitting model. We focus on the hyperparameters governing the means of the group-level distributions, which we denote with overbars (e.g., $\bar{\alpha}_{pain}$ refers to the group-level mean of α_{pain}). Figure 2A shows the posterior distribution of these group-level mean parameters, per treatment group. The corresponding 95% highest density intervals (HDIs) are reported in Figure 2—figure supplement 1. The 95% HDIs of each individual participant's learning rate posteriors are shown in Figure 2—figure supplement 2.

In the placebo group, the posterior distribution of $\bar{\alpha}_{pain}$ (median = 0.72) was considerably higher than the posterior distribution of $\bar{\alpha}_{no-pain}$ (median = 0.32; $\bar{\alpha}_{pain} > \bar{\alpha}_{no-pain}$ for 99.6% of the MCMC samples), indicative of stronger expectation updating when pain was received than avoided. In contrast, in both drug groups, the posterior distributions of $\bar{\alpha}_{pain}$ and $\bar{\alpha}_{no-pain}$ were highly similar, due to a specific increase in $\bar{\alpha}_{no-pain}$ relative to the placebo group. In the levodopa group, the posterior medians of $\bar{\alpha}_{pain}$ and $\bar{\alpha}_{no-pain}$ were, respectively, 0.66 and 0.66 ($\bar{\alpha}_{pain} > \bar{\alpha}_{no-pain}$ for 50% of the MCMC samples). In the naltrexone group, the posterior medians of $\bar{\alpha}_{pain}$ and $\bar{\alpha}_{no-pain}$ were, respectively, 0.72 and 0.76 ($\bar{\alpha}_{pain} > \bar{\alpha}_{no-pain}$ for 38% of the MCMC samples). Note that the best-fitting model for both drug groups contained separate learning rates for received and avoided pain. Combined with the finding that the group-level mean learning-rate parameters for these two outcomes were highly similar, this suggests that some participants in each drug group learned more from received than avoided pain while others showed the opposite bias,

but that there was no systematic learning asymmetry (at the individual level, α_{pain} was higher than $\alpha_{no-pain}$ for 50% of the levodopa and 41% of the naltrexone participants).

Figure 2. Model parameters. **A.** Posterior distributions of the parameters' group-level means for each group (left and right panels). Parameters $\alpha_{no-pain}$ and α_{pain} are learning rates for avoided and received pain outcomes, respectively; parameter β is the inverse-temperature parameter. The middle panels are joint density plots of $\bar{\alpha}_{pain}$ and $\bar{\alpha}_{no-pain}$ (dots are samples from the MCMC), showing that $\bar{\alpha}_{pain}$ is reliably greater than $\bar{\alpha}_{no-pain}$ in the placebo group only. **B.** The difference between the posterior distributions for each drug group vs. the placebo group, showing that $\bar{\alpha}_{no-pain}$ is greater and $\bar{\beta}$ is smaller in both drug groups compared to the placebo group. Red lines indicate 95% HDIs.

Thus, at the group level, both levodopa and naltrexone, as compared to placebo, increased learning rates for avoided pain, while not affecting learning from received pain (Figure 2A, left and middle panels). To test the significance of these group differences, we computed the difference between the posterior distributions of the group-level mean parameters for each drug group vs. the placebo group (Figure 2B). For $\bar{\alpha}_{no-pain}$, 99.7% of the difference distribution for levodopa vs. placebo, and 99.9% of the difference distribution for naltrexone vs. placebo, lay above 0. In contrast, $\bar{\alpha}_{pain}$ did not differ between the drug and placebo groups: 34% and 49% of the difference distributions for levodopa vs. placebo and naltrexone vs. placebo, respectively, lay above 0. Thus, both drugs selectively increased learning rates for avoided pain.

The posterior distribution of inverse-temperature parameter $\bar{\beta}$ was higher for the placebo group (median = 8.8) than the levodopa and naltrexone group (median = 5.3 and 5.7, respectively), as well. Specifically, 98.8% of the $\bar{\beta}$ difference distribution for levodopa vs. placebo, and 98.1% for naltrexone vs. placebo, lay below 0, indicating that $\bar{\beta}$ was reliably lower for both drug groups compared to the placebo group. This suggests that participants in the two drug groups, as compared to the placebo group, were less prone to choose the option with the lowest expected pain probability (i.e., more stochastic choice behavior).

Together, the parameter estimates suggest that (i) untreated (placebo group) participants updated their expectations more rapidly following received than avoided pain, (ii) levodopa and naltrexone negated this learning asymmetry by selectively increasing learning rates for avoided pain, and (iii) levodopa and naltrexone also increased choice stochasticity, possibly reflecting a more exploratory or risky choice strategy.

Replication of learning rate asymmetry in an independent group of untreated participants

To examine the robustness of our finding of asymmetric learning rates in the placebo group, we applied our hierarchical Bayesian modeling approach to the choice data of 23 untreated participants from our previous pain-avoidance learning fMRI study (Roy et al. 2014), and tested whether the higher learning rate for received than avoided pain found in our placebo group was replicated in this previous dataset (Figure 2—figure supplement 3). In this previous dataset, $\bar{\alpha}_{pain}$ (median = 0.62) was indeed higher than $\bar{\alpha}_{no-pain}$ (median = 0.44), resembling the placebo-group results from our current study, although the learning-rate asymmetry was somewhat smaller in the previous dataset ($\bar{\alpha}_{pain} > \bar{\alpha}_{no-pain}$ for 90% of the

MCMC samples). The posterior median of $\bar{\beta}$ in the previous dataset was 7.1 (95% HDI = 4.7-9.9). The finding of a higher learning rate for pain than no-pain outcomes in this independent dataset corroborates the idea that people normally (in the absence of a pharmacological manipulation) update their expectations more rapidly following received than avoided pain.

Parameter recovery

To validate the conclusion that our pharmacological manipulations affected both $\bar{\alpha}_{no-pain}$ and $\bar{\beta}$, we simulated two sets of choice data, using the parameter values found in the placebo and drug groups, and performed parameter-recovery analyses (Appendix 1). In sum, we found that our modeling procedure can distinguish the two patterns of parameter values found in the placebo and drug groups, even though they produce similar model-independent performance measures (number of received pain stimuli, and frequency of switching following pain and no-pain outcomes). Thus, the observed drug effects on both $\bar{\alpha}_{no-pain}$ and $\bar{\beta}$ are unlikely to merely reflect a tradeoff between these parameters. Instead, the parameter-recovery results suggest that levodopa and naltrexone had two computational effects—an increased learning rate for avoided pain, and an increased degree of choice stochasticity—whose combination yielded no significant effects on basic, model-independent, performance measures.

fMRI analyses

Next, we will report two sets of fMRI analyses. First, we performed an axiomatic analysis to identify brain activation encoding *general* aversive prediction errors (i.e., activation encoding the degree to which both pain and no-pain outcomes are relatively worse, or less good, than expected). In our previous study, this analysis revealed a general aversive prediction error signal in the PAG (Roy et al. 2014). Here, we examined whether we could replicate this finding, and whether our pharmacological manipulations affected the brain activation associated with the prediction-error axioms. Second, to address the question whether learning from received and avoided pain is supported by two separate brain systems, we sought to identify brain activation encoding *outcome-specific* prediction error signals. In both analyses, we focused on the first second of the outcome period as this is when prediction errors are triggered.

We modeled drug effects using two second-level regressors (levodopa vs. placebo and naltrexone vs. placebo) in both analyses. All fMRI results are thresholded at $q < 0.05$, false

discovery rate (FDR)-corrected for multiple comparisons across the whole brain (gray matter masked), unless otherwise stated (e.g., for visualization purposes). Unthresholded t maps can be found on <https://neurovault.org/collections/RIVRRMAK/>.

General (outcome-nonspecific) aversive and appetitive prediction error signals

Prediction-error related activation is often examined by regressing fMRI activity at outcome onset on model-derived prediction errors (see Figure 3—figure supplement 1 for the corresponding activation in our study). However, as prediction errors are defined as the outcome minus the expected outcome, a problem with this approach is that the resulting brain activity may predominantly track the outcome (in our task: pain vs. no pain) or the expected outcome (in our task: the expected pain probability), which are intrinsically correlated with the prediction error.

To address this issue, and identify brain activity that truly integrates actual and expected outcomes into a prediction error signal, a set of conditions has recently been specified (Rutledge et al. 2010, Roy et al. 2014). These conditions, or axioms, for general aversive prediction error signals in our task are: (i) activation at outcome onset should be higher for received than avoided pain, unless pain is fully expected; (ii) when pain is received, activation should be higher when pain was less expected (i.e., negative correlation with expected pain probability); and (iii) when pain is avoided, activation should also be higher when pain was less expected, that is, when avoidance was more expected (Figure 3A, left panels). To identify regions encoding a general aversive prediction-error signal, we tested for activation that fulfilled each of these three axioms, using a whole-brain conjunction analysis. In addition, to search for regions encoding the opposite (i.e., appetitive-like) prediction error signal, we also tested for activation that fulfilled each of the *reverse* axioms. Note that we did not detect activation encoding appetitive-like prediction errors in our previous study (Roy et al. 2014).

A fourth axiom, that applies to both aversive and appetitive prediction errors, is that activation for received and avoided pain should be equivalent if the outcome is fully predicted (i.e., when the prediction error is zero). As outcomes could never be fully predicted in our task, we could not test this axiom.

Figure 3. Axiomatic tests of brain activation encoding general aversive and appetitive prediction errors (N = 74). **A.** Activation associated with the three axioms for aversive prediction errors in our task. Yellow regions showed the effects illustrated in the left panels, and blue regions showed the reverse effects (i.e., the axioms for appetitive prediction errors). Expected P(pain) is expected pain probability. All maps were thresholded at $q < 0.05$, FDR-corrected for multiple comparisons across the whole brain, with higher voxel thresholds superimposed for display. **B.** Conjunction results. Regions activated for each of the above three contrasts, all thresholded at $q < 0.05$ FDR-corrected. Yellow and blue regions showed positive and negative responses for each contrast, respectively, thus encoded general aversive and appetitive prediction errors.

Axiom 1. A large part of the brain fulfilled the first axiom for aversive prediction errors (stronger response to received than avoided pain), including typical pain-processing regions such as the dorsal anterior cingulate cortex (ACC), (pre)motor cortex, insula and thalamus, as well as occipital (visual) cortex (Figure 3A, upper panel). We found the opposite effect (stronger response to avoided than received pain) in regions of the ventromedial

prefrontal cortex (vmPFC), dorsolateral prefrontal cortex (dlPFC), somatosensory cortex, posterior ACC, and later occipital cortex (LOC).

Axiom 2. A test of the second axiom for aversive prediction errors (stronger responses to more unexpected pain) revealed several activation clusters, including regions in the vmPFC, dorsal ACC, insula, amygdala, and a midbrain area covering part of the PAG (Figure 3A, middle panel). In addition, several other regions, including the right dlPFC and bilateral somatosensory cortex, showed the opposite effect (stronger responses to more expected pain).

Axiom 3. The third axiom for aversive prediction errors (stronger responses to more expected pain avoidance) was fulfilled by a few regions in the vmPFC and rostral ACC (rACC), as well as part of the PAG. We also found the opposite effect (stronger responses to more unexpected pain avoidance) in several regions, including the dorsal ACC, sensorimotor cortex, thalamus, putamen and insula.

Conjunction. A conjunction analysis of the three contrasts reported above (Figure 3B) revealed two brain regions that satisfied all three axioms for aversive prediction errors: A midbrain region including part of the PAG (16 voxels) and an area in the rostral ACC (24 voxels). Importantly, we also identified activation that showed a *negative* effect for all three axioms—thus encoding appetitive-like prediction errors—in bilateral somatosensory cortex (433 and 65 voxels in the left and right hemisphere, respectively), left frontopolar cortex (47 voxels), right dlPFC (middle frontal gyrus; 27 voxels), and right LOC (253 voxels).

Together, these results replicate our previous finding that the PAG encodes general aversive prediction errors (Roy et al. 2014), and suggest a role for the rostral ACC in encoding aversive prediction errors as well. Furthermore, the identification of an additional neural circuit encoding appetitive-like prediction errors provides an important extension to our previous results, possibly owing to the larger number of participants and hence higher power in the present study.

No drug effects on prediction-error related brain activation

We found no differences between the levodopa and placebo group, nor between the naltrexone and placebo group, for any of the three prediction-error contrasts reported above (whole-brain FDR-corrected). Drug effects were virtually absent at lower, uncorrected, thresholds as well (see unthresholded *t* maps on <https://neurovault.org/collections/RIVRRMAK/>). More specific tests for drug effects in the clusters identified by our conjunction analysis provided no evidence for effects of our

pharmacological manipulations on prediction-error related brain activation either (Figure 3—figure supplement 2).

Outcome-specific prediction error signals

The previous analysis identified regions that encoded the degree to which outcomes were relatively worse- or better-than-expected in the *same* direction when pain was received and avoided, but did not *dissociate* learning from received and avoided pain. To address this issue, we next examined whether different brain regions encode the unexpectedness, or surprise (which drives learning in reinforcement-learning models), evoked by received and avoided pain.

Regions encoding the surprise evoked by received pain should respond stronger to pain outcomes when pain was less expected (i.e., a negative correlation with expected pain probability on pain trials). In contrast, regions encoding the surprise evoked by avoided pain should respond stronger to no-pain outcomes when pain was more expected (i.e., a positive correlation with expected pain probability on no-pain trials; Figure 4A, left plots). To identify regions that show a stronger surprise response for received than avoided pain, we thus specified the following contrast: negative correlation with expected pain probability on pain trials > positive correlation with expected pain probability on no-pain trials. This contrast revealed activation in the vmPFC and rACC, posterior cingulate cortex, insula extending into the left parahippocampal gyrus, cerebellum and a brainstem region encompassing part of the PAG (Figure 4A, yellow regions). We also found extensive clusters that showed the opposite effect—i.e. positive correlation with expected pain probability on no-pain trials > negative correlation with expected pain probability on pain trials—in sensorimotor cortex, parietal and occipital cortex, and the bilateral frontal poles (Figure 4A, blue regions), suggesting that these regions show a stronger surprise response for avoided than received pain.

Figure 4. Outcome-specific prediction-error signals (N = 74). **A.** Activation tracking surprise more for received than avoided pain (yellow) and vice versa (blue). Note that this includes activation that tracks expected pain probability across both outcomes. **B.** Activation tracking surprise for both received and avoided pain (i.e., absolute prediction error). Activation maps in A and B are thresholded at $q < 0.05$, FDR-corrected for multiple comparisons across the whole brain, with adjacent areas thresholded at $p < 0.01$ and $p < 0.05$ (uncorrected) for display. **C.** Regions encoding surprise more for received than avoided pain, which cannot be explained by a general sensitivity to expected pain probability. These regions showed positive activation for both the first (A) and second (B) contrast, each thresholded at $q < 0.05$, FDR-corrected. **D.** Regions encoding surprise more for avoided than received pain, which cannot be explained by a general sensitivity to expected pain probability. These regions showed negative activation for the first, and positive activation for the second contrast, each thresholded at $q < 0.05$, FDR-corrected. The line plots show the mean activity extracted from the brainstem and right amygdala (C) and left dlPFC and parietal (D) clusters per quartile of

expected pain probability, illustrating the encoding of outcome-specific prediction errors in these regions.

We also sought to identify activation encoding the surprise elicited by both received and avoided pain, that is, activation encoding absolute prediction errors. To this end, we specified a second contrast that tested for a negative correlation with expected pain probability on pain trials *and* a positive correlation with expected pain probability on no-pain trials (Figure 4B). This contrast revealed extensive activation clusters in the dorsal ACC extending into the supplementary motor cortex, insula, sensorimotor cortex, thalamus, part of the brainstem and cerebellum, suggesting that these regions encoded absolute prediction error (Figure 4B, yellow regions). In addition, a few smaller clusters in left sensorimotor cortex, right dlPFC, and left frontopolar cortex showed the opposite effect, suggesting that these regions encoded the overall *expectedness* of outcomes (Figure 4B, blue regions).

Finally, note that a caveat of the first contrast reported above (Figure 4A) is that it also identifies activation that is stronger when the expected pain probability is lower (i.e., activation encoding expected safety), regardless of the outcome. This likely explains the vmPFC activation, which has been found to represent positive affective value in many domains (Roy et al. 2012, Bartra et al. 2013, Rich and Wallis 2016). The second contrast (for absolute prediction errors; Figure 4B), on the other hand, will *not* detect activation encoding expected safety regardless of the outcome, as the correlation with expected pain probability is specified in opposite directions for pain and no-pain outcomes. Thus, we reasoned that regions identified by both of the contrasts reported in Figure 4A and 4B encode outcome-specific prediction errors (Figure 4A) unconfounded by outcome-nonspecific expected pain probability (Figure 4B). Therefore, we next examined the conjunction of these two contrasts. Specifically, we masked the activation identified by the first contrast (separately for the positive and negative activation) by the positive activation identified by the second contrast. Positive activation for both contrasts was found in a set of mostly subcortical and limbic regions, including a large cluster in the brainstem (not covering the PAG), bilateral insula extending into the amygdala, rostral ACC, posterior cingulate cortex, bilateral supramarginal gyrus, and cerebellum (Figure 4C). These regions thus encoded surprise more for received than avoided pain (pain-specific prediction errors), which could not be explained by a general sensitivity to expected pain probability. Negative activation for the first contrast and positive activation for the second contrast, on the other hand, was found in several cortical regions, including the supplementary motor cortex, left parietal cortex (supramarginal gyrus), bilateral

somatosensory cortex (postcentral gyrus), and left dlPFC (middle and superior frontal gyrus) (Figure 4D). These regions thus encoded surprise more for avoided than received pain (no-pain specific prediction errors), which could not be explained by a general sensitivity to expected pain probability. Together, these results provide evidence that prediction errors evoked by pain and no-pain outcomes are encoded in largely distinct brain regions.

No drug effects on surprise-related brain activation

Because levodopa and naltrexone specifically increased learning rates for no-pain outcomes, we expected these drugs to increase prediction-error related brain activation for no-pain outcomes as well. However, we found no differences between the levodopa and placebo group, or between the naltrexone and placebo group, for any of the contrasts reported above (whole-brain FDR-corrected). Drug effects were virtually absent at lower, uncorrected, thresholds as well (see unthresholded t maps on <https://neurovault.org/collections/RIVRRMAK/>).

Discussion

Our results provide novel evidence that unexpectedly received and avoided pain—signaling threat and safety, respectively—drive human pain-avoidance learning via different learning systems. First, computational modeling suggested that participants' choices were best explained by a model with separate learning rates for received and avoided pain, and that untreated participants learned more from received than avoided pain. Second, levodopa and naltrexone selectively increased learning rates for avoided pain, suggesting a role for the dopamine and endogenous-opioid systems in safety, but not threat, learning. Third, our fMRI analyses revealed that different brain circuits encode prediction errors elicited by received and avoided pain, providing evidence for two dissociable learning systems at the neural level as well. Somewhat surprisingly, however, we found no drug effects on fMRI activity for any of the contrasts we examined. We discuss each of these findings below.

Learning rates for received vs. successfully avoided pain

The higher learning rate for pain than no-pain outcomes in the placebo group suggests that people normally (in the absence of a pharmacological manipulation) update their expectations more following received than avoided pain. A similar learning-rate asymmetry was found in a recent aversive reversal-learning study (Wise et al. 2019). Interestingly,

however, reward-learning studies using secondary outcomes (e.g., monetary gains and losses) have provided evidence for the *opposite* asymmetry: higher learning rates for favorable than unfavorable outcomes. This has been attributed to an optimistic learning bias (Sharot and Garrett 2016, Lefebvre et al. 2017) and a tendency to learn preferentially from information that confirms one's current action (Palminteri et al. 2017). The opposite learning bias in pain-avoidance learning tasks may be due to the intrinsically aversive nature of pain, arguably rendering unexpected pain a more salient teaching signal than unexpected pain absence. Relatedly, the experience of pain may trigger a reflexive tendency to change one's course of action (Huys et al. 2012), expressed in elevated learning rates for pain outcomes. Thus, higher learning rates for received than avoided pain may reflect a Pavlovian influence on choice which operates in parallel to the instrumental learning system. Alternatively, the seemingly opposite learning asymmetries in reward-learning and pain-avoidance learning tasks may also reflect a cognitive process related to the framing of the task. That is, participants who are instructed to maximize reward vs. minimize pain may pay most attention to—and hence learn most from—reward vs. pain outcomes, respectively.

The presence and direction of learning asymmetries may also depend on the specific task demands. For example, a previous fMRI study that used a more complex pain-avoidance learning task (pain probabilities of three different options were learned in parallel, in an indirect manner), and included a risk-taking component, found no systematic difference in learning rates for received and avoided pain (Eldar et al. 2016). Interestingly, in contrast to our present and previous (Roy et al. 2014) fMRI findings, the PAG in that study *positively* encoded expected pain probability on no-pain trials (one of the axioms for appetitive prediction errors), and did not encode expected pain probability on pain trials. These findings suggest that the task used in that previous study (Eldar et al. 2016) evoked different learning processes than our simpler task. Indeed, learning rates in that previous study were much lower than those found in our task as well. Further examination of the degree to which avoidance-learning processes and their neural implementation generalize across different learning tasks is an important objective for future work (Yarkoni 2020).

Effects of levodopa and naltrexone on learning parameters

Levodopa and naltrexone selectively increased learning rates for successfully avoided pain, consistent with a role for dopamine and endogenous opioids in safety learning. Previous studies have associated levodopa-induced increases in phasic dopamine activity with enhanced learning from (secondary) rewards (Frank et al. 2004, Pessiglione et al. 2006).

Combined with these previous findings, our levodopa results suggest that phasic dopamine activity may signal the degree to which outcomes are ‘better than expected’ across both reward and punishment domains. We are aware of one previous study that provided correlational evidence for a role of dopamine in human safety learning in a Pavlovian fear-conditioning task (Raczka et al. 2011). That study found that individual differences in fear-extinction learning rates were associated with genetic variation in the dopamine transporter gene, which presumably affects phasic striatal dopamine release. Our levodopa results are consistent with this result, and provide the first *causal* evidence for a role of dopamine in human safety learning in an instrumental-learning task.

It has been proposed that the dopamine system signals aversive prediction errors as well, via a subset of midbrain dopaminergic neurons that is responsive to aversive outcomes (Brooks and Berns 2013). Consistent with this hypothesis, a subset of putative dopamine neurons in non-human primates (Matsumoto and Hikosaka 2009) and rodents (Brischoux et al. 2009, Lammel et al. 2011) have been found to increase their activity in response to aversive stimuli. However, it is yet unclear whether these neurons signal aversiveness, salience, or relief from a threat period (Mirenowicz and Schultz 1996, Schultz 2019), and whether these neurons are really dopaminergic (Ungless et al. 2004). In our study, levodopa did not affect learning rates for received pain; hence our results do not support the idea that dopamine supports learning from aversive outcomes. According to another account—which is largely based on studies using secondary reinforcers in Parkinson’s disease patients—punishment learning (e.g., learning from monetary losses) is mediated by phasic *decreases* (dips) or pauses in dopamine activity, via an effect on D2 receptors (Frank 2005, Maia and Frank 2011). Although levodopa is thought to primarily increase phasic dopamine bursts (Floel et al. 2008, Black et al. 2015), it may increase tonic dopamine levels as well, which could in turn interfere with phasic dopamine dips (Breitenstein et al. 2006, Moustafa et al. 2013, Poletti and Bonuccelli 2013). Thus, if dopamine dips following aversive prediction errors support pain-avoidance learning, we would expect levodopa to *impair* learning from received pain. Our finding that levodopa did not affect learning rates for received pain does not support this hypothesis either. Instead, our results suggest a selective role for the human dopamine system in learning from successfully avoided pain.

Regarding the endogenous opioid system, we expected that if pain-avoidance learning relies on μ -opioid activity, naltrexone—which blocks this activity—would *suppress* learning from received and/or avoided pain. However, we found the opposite effect for avoided-pain outcomes: like levodopa, naltrexone increased learning rates for avoided pain. This finding

counterintuitively suggests that μ -opioid activity normally suppresses learning from avoided pain and that naltrexone countered this effect, which seems to contradict findings that μ -opioid receptor antagonists impair fear-extinction learning in rats (McNally and Westbrook 2003, McNally et al. 2004, McNally et al. 2005, Kim and Richardson 2009, Parsons et al. 2010). Obviously, these animal studies differed from our study in several ways—such as the nature of the learning task (Pavlovian vs. instrumental), outcome measure (freezing behavior vs. learning-rate estimates), opioid-receptor antagonist (naloxone vs. naltrexone) and drug administration (injection into the PAG vs. oral administration)—each of which may explain the apparently contradictory results. Thus, our results show that untreated participants learn more rapidly from received than avoided pain (replicated in our previous study) and that both levodopa and naltrexone negate this learning asymmetry, but the neurobiological mechanisms underlying the naltrexone effect remain to be elucidated. One informative approach for future studies would be to directly compare effects of opioid-receptor agonists and antagonists on pain-avoidance learning parameters.

The levodopa and naltrexone groups also showed a higher degree of choice stochasticity than the placebo group, suggesting that participants in both drug groups were less prone to choose the option with the lowest expected pain probability. Importantly, our parameter-recovery analysis indicated that the drug effects on learning rate for avoided pain and choice stochasticity could be independently and correctly retrieved. The increased choice stochasticity in the levodopa group may reflect a positive association between dopamine activity and risk preference (Voon et al. 2006, Gallagher et al. 2007, St Onge and Floresco 2009, Chew et al. 2019) or exploration (Beeler et al. 2010, Kayser et al. 2015, Gershman and Tzovaras 2018). It is also broadly consistent with recent evidence that levodopa reduces the impact of valence on information-seeking (Vellani et al. 2020). We are not aware of previous studies that associated the endogenous opioid system with choice stochasticity, risk-taking or exploration. It is possible that the similar effects of levodopa and naltrexone on choice stochasticity were due to common general side effects of both drugs on participants' attention or motivation, which disrupted the decision-making process. However, we believe this is unlikely because (i) the drugs did not affect subjective state (alertness, calmness or contentment) and (ii) general side effects cannot easily explain the specific increase in learning rates for avoided pain.

Interestingly, the drug effects on learning rate for avoided pain and choice stochasticity were not accompanied by drug effects on basic performance measures (number of received pain outcomes, pain-switch or avoid-switch behavior). In a similar vein, previous

studies have reported effects of pharmacological manipulations and dopaminergic genotype on model parameters despite a lack of significant effects on behavioral measures (Raczka et al. 2011, Chakroun et al. 2020), illustrating the added value of computational models. Our simulation results suggested that the drug effects on learning rate and choice stochasticity in our study cancelled each other out. Specifically, symmetric learning rates for received and avoided pain (found in the drug groups) result in more accurate pain-probability estimates than asymmetric learning rates (found in the placebo group). This beneficial effect of a symmetric learning process in the drug groups was, however, counteracted by the detrimental effect of a more stochastic choice process, resulting in no net performance difference between the placebo and drug groups.

Separate brain circuits support learning from received and avoided pain

Our fMRI results provided evidence for two dissociable learning systems at the brain level as well. Pain-specific prediction errors were predominantly encoded in subcortical (brainstem, cerebellum) and limbic (insula, amygdala, rostral ACC) regions that are typically associated with emotional and affective processes, including fear conditioning (Phillips and LeDoux 1992) and affective responses to errors (Bush et al. 2000). In contrast, no-pain specific prediction errors were encoded in frontal and parietal cortical areas typically associated with higher-order cognitive processing (Ptak et al. 2017).

Prediction errors for no-pain outcomes were not represented in the ventral striatum, which is typically found for reward prediction errors (O'Doherty et al. 2003, Rutledge et al. 2010). Instead, prediction errors for no-pain outcomes were associated with frontoparietal activity. This activity is unlikely to reflect a reward signal, but possibly reflected increased attention on trials in which pain was expected but not received. That is, the unexpected absence of pain may have prompted participants to carefully monitor the thermode's temperature in order to verify whether pain was really avoided or still to come, as reflected in increased frontoparietal activity. The absence of a striatal 'reward-like' prediction-error response for no-pain outcomes suggests that, in terms of neural processing, avoiding pain is not comparable to gaining a reward. However, the lack of a detectable reward-like prediction-error response may also be related to our task design. Specifically, pain outcomes in our task involved a change in sensory input (a rise in temperature) whereas no-pain outcomes did not (maintenance of the baseline temperature), which may have caused a more prominent neural prediction-error response for the pain outcomes. One way to examine this issue would be to use a task in which choices result in either an increase or a decrease in painful stimulus

intensity from a tonic pain level, such that aversive and appetitive-like outcomes are associated with similar changes in sensory input (Seymour et al. 2005). Such a task would examine pain-relief rather than pain-avoidance learning. The current task design, however, more closely resembles the situation of a patient recovering from injury or surgery, who is pain-free as long as he or she is resting but expects that physical activity may result in pain. It is an interesting speculation that, in such situations, the stronger subcortical and limbic (‘emotional’) responses for pain than for no-pain prediction errors, as found in our study, may foster a behavioral state that favors inactivity and rest, which could promote tissue healing and recovery.

No effects of levodopa and naltrexone on fMRI activation

Unexpectedly, we found no effects of levodopa or naltrexone on any of our fMRI prediction-error contrasts. This may indicate that the dopamine and opioid systems are not involved in pain-avoidance learning, although this is inconsistent with the drug effects on learning rates for avoided pain. It is also possible that our pharmacological manipulations *did* affect prediction-error related dopamine and/or opioid activity, but that we did not have enough power to detect these effects due to our moderate sample size and between-subject design (24-26 participants per treatment group). Relatedly, drug effects on prediction-error related dopamine and/or opioid activity may not always produce corresponding changes in the BOLD signal (Knutson and Gibbs 2007, Brocka et al. 2018). This last possibility is clearly discouraging for pharmacological fMRI studies, but is conceivable given the absence of consistent levodopa effects on reward prediction-error signals in previous fMRI studies: One study reported an increased reward prediction-error signal in an levodopa group compared to a haloperidol group (although neither of these groups differed from a placebo group) (Pessiglione et al. 2006), but two recent studies using within-subject designs found no levodopa effects on neural reward prediction-error signals (Kroemer et al. 2019, Chakroun et al. 2020). To better understand the degree to which drug-induced changes in human neuromodulatory (e.g., dopamine) activity are detectable using fMRI, future research could directly compare drug effects on local changes in neuromodulator activity—e.g., using molecular imaging procedures such as PET—with effects of these same drugs on the BOLD signal.

Limitations and directions for future research

As mentioned above, a lack of statistical power due to our moderate sample size and between-subject design may have prevented the detection of drug effects in our fMRI analyses. In addition, our fMRI data suffered from signal dropout in inferior parts of the prefrontal cortex (including the orbitofrontal cortex), which is a common problem in fMRI studies using echo-planar imaging (Ojemann et al. 1997, Deichmann et al. 2003). Therefore, our results are agnostic with respect to the contribution of ventral prefrontal areas to pain-avoidance learning, and their potential modulation by levodopa or naltrexone. Finally, regarding the role of neuromodulators, we focused on dopamine and endogenous opioids, but other neuromodulators are almost certainly involved in pain-avoidance learning as well. In particular, future work may focus on the serotonergic system, which has traditionally been associated with aversive processing, behavioral inhibition, and “fight or flight” responses in rodents (Deakin 1983, Soubrie 1986, Deakin and Graeff 1991). More recent pharmacological and genetic studies, which mostly used tasks with secondary reinforcers, have provided evidence for a role of the human serotonergic system in various aspects of aversive learning (Chamberlain et al. 2006, Cools et al. 2008, Crockett et al. 2012, Hindi Attar et al. 2012, Robinson et al. 2012, den Ouden et al. 2013). Based on these findings, it has been proposed that serotonin acts as an opponent to dopamine by mediating behavioral inhibition in response to punishment (Daw et al. 2002, Dayan and Huys 2009, Boureau and Dayan 2011, Cools et al. 2011). When generalizing these findings to the domain of pain-avoidance learning, we may expect a specific role for the serotonergic system learning from received pain (threat learning). However, other studies have found effects of serotonin manipulations on both reward and punishment learning (Palminteri et al. 2012, Guitart-Masip et al. 2014), and a study in which participants simultaneously learned probabilities of monetary rewards and painful shocks suggested that serotonin selectively modulates reward processing (Seymour et al. 2012). Taken together, previous work suggests that the serotonergic system has highly intricate and multifaceted roles in affective learning, likely due to its large number of receptor types, widespread projections, and interactions with other neuromodulators (Dayan and Huys 2009). Future studies are required to further clarify the functional role(s) of serotonin in pain-avoidance learning.

Conclusion

In sum, our results suggest that received and avoided pain drive human pain-avoidance learning via two different learning systems, in terms of both learning rates and the neural encoding of prediction errors. In addition, our computational-modeling results provide

evidence for a causal role of the dopamine and endogenous opioid systems in learning from avoided, but not received, pain. Future studies are needed to elucidate the neural mechanisms via which our dopamine and opioid manipulations affected learning rates for avoided pain, and to reveal the potential role of other neuromodulators in pain-avoidance learning.

Methods

Participants

Ninety-one healthy students (18-26 years old; 71% female; all right-handed) took part in the study. Participants reported no history of psychiatric, neurological, or pain disorders, and no current pain. Participants were instructed to abstain from using alcohol or recreational drugs 24 hours prior to testing, and to not eat or drink (except for water) 2 hours prior to testing. The study was approved by the medical ethics committee of the Leiden University Medical Center, and all participants provided written informed consent. Participants received a fixed amount of €60 plus a variable bonus of maximally €5 related to their performance on an additional task.

Six participants were excluded from all analyses because of thermode failure, and two additional participants because of poor task performance (see ‘Pain-avoidance learning task’ section below). In addition, nine participants were excluded from the fMRI, but not the behavioral, analyses because of excessive head movement (> 3 mm in any direction). Thus, the final behavioral analyses included 83 participants (placebo group: $N = 28$, mean age = 21.2, 74% female; levodopa group: $N = 26$, mean age = 20.8, 73% female; naltrexone group: $N = 29$, mean age = 20.8, 68% female). The final fMRI analyses included 74 participants (placebo group: $N = 24$, mean age = 21.3, 71% female; levodopa group: $N = 24$, mean age = 20.9, 71% female; naltrexone group: $N = 26$, mean age = 20.8, 69% female). Sample size was based on previous studies that detected effects of dopamine (levodopa) or opioid (naloxone) manipulations on behavioral and/or fMRI measures using between-subject designs (Pessiglione et al. 2006, Eippert et al. 2008, Guitart-Masip et al. 2012, Oei et al. 2012, Beierholm et al. 2013, Bunzeck et al. 2014, Guitart-Masip et al. 2014, Wittmann and D’Esposito 2015). These previous studies used sample sizes ranging from 13 to 30 participants per treatment group. We aimed at a sample size at the higher end of that range (25-30 participants per treatment group).

General procedure

Two to fourteen days prior to the fMRI session, we assessed participants' eligibility using a general health questionnaire and an fMRI safety screening form. During this screening session, participants also practiced the pain-avoidance learning task.

Each eligible participant took part in one fMRI session. On the day of the fMRI session, participants received a single oral dose of either 100 mg levodopa, 50 mg naltrexone, or placebo, according to a double-blind, randomized, between-subject design. Levodopa was combined with 25 mg carbidopa—a decarboxylase inhibitor that does not cross the blood brain barrier—to inhibit the conversion of levodopa to dopamine in the periphery.

Approximately 30 minutes after drug administration, participants were positioned in the MRI scanner, after which we acquired a high-resolution structural scan. Approximately 53 minutes after drug administration, participants completed a 5-minute pain-rating task during which they received a series of (unavoidable) heat stimuli of varying temperatures and rated their experienced pain following each stimulus (Figure 1—figure supplement 2). This task was included to test for drug effects on subjective pain responses, and to select a painful yet tolerable temperature for each participant in the pain-avoidance learning task. Sixty minutes after drug administration, roughly corresponding with peak plasma concentrations of levodopa and naltrexone, participants started the pain-avoidance learning task (described below), which lasted approximately 45 minutes. Following the pain-avoidance learning task, participants performed an 8-minute probabilistic reward-learning task (not reported here).

We measured participants' subjective state at the beginning (before drug administration) and end (two hours after drug administration) of the test session, by means of visual analogue scales measuring alertness, calmness and contentment (Bond, 1974) (Figure 1—figure supplement 1). Both subjective state measures were collected outside the scanner.

Pain-avoidance learning task

This instrumental pain-avoidance learning task contained 144 trials, divided in 4 runs of 36 trials. On each trial, participants made a choice between two options (a diamond and a circle). Choosing each option was associated with a specific probability of receiving a painful heat stimulation. The probabilities of receiving pain when choosing each option drifted across trials according to two independent random walks (Figure 1A). We used three different pairs of random walks (each pair crossed at least once); each pair was administered to approximately one third of the participants in each treatment group.

Each trial started with the presentation of the two choice options, randomly displayed at the left and right side of the screen for 1800 ms (Figure 1B). During this period,

participants had to select one option by pressing a left or right button of the response unit, using their right index or middle finger, respectively. If participants did not respond in time (1.2 % of trials), the computer randomly selected an option for them. The chosen option was highlighted for 200 ms, followed by an anticipation period of 3, 5, or 7 seconds during which a white asterisk (*) was presented in the center of the screen. Then the outcome—a painful heat stimulus applied to participant’s leg for 1.9 s (see Thermal stimulation section below) or no stimulus—was presented. Outcome onset was accompanied by a change of the central asterisk to a colored plus sign (+). The plus sign was red or green during the first 200 ms of each pain and no-pain outcome, respectively, after which it turned white for the remainder of the outcome period. The color change was meant to prevent outcome uncertainty during the initial phase of the outcome period. Each trial ended with an inter-trial interval of 6, 8, or 10 seconds during which an asterisk was presented. Except for the outcome probabilities, participants were fully informed about the task structure and procedure.

During the fMRI session, one participant switched choices more frequently following the absence of pain than following pain, and one other participant did deliberately not make a choice on 17% of the trials, due to the use of irrelevant strategies. We excluded these participants from further analysis.

Thermal stimulation

Heat stimuli (ramp rate = 40°C/s; 1 second at target temperature) were applied to the inner side of participants’ left lower leg using a Contact Heat-Evoked Potential Stimulator (CHEPS; 27-mm diameter Peltier thermode; Medoc Ltd., Israel). After the initial pain-rating task, 16% of the participants (4 in the placebo, 4 in the levodopa, and 7 in the naltrexone group) indicated that they would not tolerate repeated stimulation at the highest temperature they had received so far (50°C). For those participants, we used a temperature of 49°C in the pain-avoidance learning task. For the remaining participants, we used a temperature of 50°C. Between stimulations the thermode maintained a baseline temperature of 32°C. The total duration of each stimulation was 1850 ms (425 ms ramp-up and ramp-down periods, 1 second at target temperature) for 49°C stimuli, and 1900 ms (450 ms ramp-up and ramp-down periods, 1 second at target temperature) for 50°C stimuli. After each scan run, we moved the thermode to a new site on the participant's leg. To reduce the impact of potential site-specific habituation (Jepma et al. 2014), we administered one initial heat stimulus before starting the first trial on a new skin site.

Behavioral analyses

We tested whether the total number of received pain stimuli, the proportion of pain trials followed by a switch to the other choice option, and the proportion of no-pain trials followed by a switch to the other choice option differed between the three treatment groups, using one-way ANOVAs. In addition, for each treatment group, we used logistic regression to analyze the probability of switching choices as a function of outcome (pain vs. no-pain) over the six previous trials.

Computational modeling

We fitted two reinforcement learning (Q learning) models to participants' choice data: one model with a single learning rate (Model 1), and one model with separate learning rates for pain and no-pain outcomes (Model 2). On each trial t , both models update the expected probability of pain for the selected stimulus s , Q_s , in response to the experienced outcome O (pain or no pain), according to:

$$Q_{s,t+1} = Q_{s,t} + \alpha(O_t - Q_{s,t})$$

where $(O_t - Q_{s,t})$ reflects the prediction error on each trial. We coded O as -1 and 0 for pain and no-pain outcomes, respectively, and initialized the Q values of both stimuli to -0.5. The value of the unchosen stimulus is not updated. Learning rate parameter α controls *how much* Q values are updated in response to each new outcome, such that higher values of α result in faster updating. The only difference between our two models is that Model 1 uses a single learning rate α for all outcomes, whereas Model 2 uses separate learning rates for pain and no-pain outcomes: α_{pain} and $\alpha_{no-pain}$, respectively.

Both models were combined with a softmax decision function, which computes the probability of choosing stimulus s on trial t ($P_{s,t}$) as:

$$P_{s,t} = \frac{e^{Q_{s,t} * \beta}}{\sum_{s'=1}^2 e^{Q_{s',t} * \beta}}$$

Inverse-temperature parameter β controls the sensitivity of choice probabilities to differences in Q values. If β is 0, both stimuli are equally likely to be chosen, irrespective of their expected pain probabilities. As β increases, the probability that the model chooses the stimulus with the lower expected pain probability increases. Thus, Model 1 has two free parameters (α and β) and Model 2 has three free parameters (α_{pain} , $\alpha_{no-pain}$ and β).

Parameter estimation

We estimated the model parameters with a hierarchical Bayesian approach, using the hBayesDM package (Ahn et al. 2017). The hierarchical Bayesian approach assumes that

every participant has a different set of model parameters, which are drawn from group-level prior distributions (Gelman 2014). The parameters governing the group-level prior distributions (hyperparameters) are also assigned prior distributions (hyperpriors). We estimated separate group-level parameters for the placebo, levodopa, and naltrexone groups. To test for treatment effects, we compared the posterior distributions of the group-level means (i.e., the hyperparameters governing the means of the group-level distributions) for each drug group vs. the placebo group.

Prior distributions. The group-level distributions for all individual parameters were assumed to be normal distributions. As hyperpriors for the mean and standard deviation (SD) hyperparameters of these group-level distributions we used, respectively, a normal distribution with mean 0 and SD 1 and a positive half-Cauchy distribution with location 0 and scale 5. We transformed the parameters' unconstrained values to a [0,1] range using the inverse probit transformation (Ahn et al. 2017). In addition, we transformed β to a [0,20] range by multiplying its inverse-probit transformed values by 20.

MCMC sampling. The hBayesDM package performs hierarchical Bayesian parameter estimation with a Markov chain Monte Carlo (MCMC) sampling algorithm called Hamiltonian Monte Carlo (HMC) implemented in Stan and its R instantiation RStan. We ran 4 independent MCMC chains with different starting values, which each generated 5,000 posterior samples. We discarded the first 1,000 iterations of each chain as burn-in. In addition, we only used every 5th iteration to remove autocorrelation. Consequently, we obtained 3,200 representative samples (800 per chain) per parameter per model fit. All chains converged (Rhat values < 1.1).

Model comparison

We compared the fit of our two models using the Watanabe-Akaike information criterion (WAIC) (Watanabe 2010). The WAIC provides an estimate of a model's out-of-sample predictive accuracy, adjusted for the number of free parameters, in a fully Bayesian way. Lower WAIC values indicate better out-of-sample predictive accuracy. WAIC is reported on the deviance scale (Gelman et al. 2014) hence a difference in WAIC value of 2 to 6 is considered positive evidence, a difference of 6 to 10 strong evidence, and a difference >10 very strong evidence for one model over another (Kass and Raftery 1995).

Computation of expected pain probability for use in fMRI analyses

We computed trial-specific expected pain probabilities—to be used as parametric modulator regressors in the fMRI analyses—by applying our winning model to each participant's sequence of choices and outcomes. In line with previous studies (Pine et al.

2010, Voon et al. 2010, Seymour et al. 2012, Hauser et al. 2015, Kroemer et al. 2019), we used the exact same model to generate fMRI regressors for all participants by instantiating the model with the mean learning-rate parameters (the individual-level posterior medians, averaged across all participants).

fMRI acquisition and analysis

Imaging acquisition

We acquired fMRI data on a 3T Philips Achieva MRI system (Best, The Netherlands) at the Leiden University Medical Center, using a standard whole-head coil. Stimulus presentation and data acquisition were controlled using E-Prime software (Psychology Software Tools). Visual stimuli were presented via a mirror attached to the head coil, and participants responded with their right hand via an MRI-compatible response unit. The pain-avoidance learning task was divided across four scan runs. Each run lasted 581 s (264 TRs), after discarding the first five TRs which served as dummy scans. Functional images were acquired with a T2*-weighted whole-brain echo-planar imaging sequence (TR = 2.2 s; TE = 30 ms, flip angle = 80°, 38 transverse slices oriented parallel to the anterior commissure-posterior commissure line, voxel size = 2.75 × 2.75 × 2.75 mm +10% interslice gap). In addition, we acquired a high-resolution T1-weighted scan (TR = 9.8 ms; TE = 4.6 ms, flip angle = 8°, 140 slices, 1.17 × 1.17 × 1.2 mm, FOV = 224×177×168), at the beginning of the scan session.

Preprocessing

Prior to preprocessing, global outlier time points (i.e. “spikes” in the BOLD signal) were identified by computing both the mean and the standard deviation (across voxels) of values for each image for all slices. Mahalanobis distances for the matrix of slice-wise mean and standard deviation values (concatenated) x 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 methods) were considered outliers. On average 3.7% of images were outliers (SD = 1.9). 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 coregistered to the first functional image for each subject using an iterative procedure of automated registration using mutual information coregistration 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 $2 \times 2 \times 2$ mm voxels, and smoothed using a 6mm full-width at half maximum Gaussian kernel.

Axiomatic analysis of general aversive and appetitive prediction error signals

For the first-level analysis, we create a general linear model (GLM) for each participant, concatenated over the four pain-avoidance learning blocks, in SPM8. We modeled periods of decision time (cue onset until response, mean response time = 758 ms), outcome anticipation (3-7 s), onsets of pain outcomes (1 s), and onsets of no-pain outcomes (1 s), using boxcar regressors convolved with the canonical hemodynamic response function. As in our previous study (Roy et al. 2014) we only modeled the first second of the outcome periods as this is when prediction errors are triggered. We added the model-derived expected pain probability as a parametric modulator on the outcome-anticipation and outcome-onset regressors. To control for potential effects of outcome-anticipation duration, we also included anticipation duration as a first parametric modulator on the outcome-onset regressors (using serial orthogonalization, such that any shared variance between expected pain probability and anticipation duration is assigned to the anticipation-duration effect). Other regressors of non-interest (nuisance variables) were i) “dummy” regressors coding for each run (intercept for each but the last run); ii) linear drift across time within each run; iii) the 6 estimated head movement parameters (x, y, z, roll, pitch, and yaw), their mean-zeroed squares, their derivatives, and squared derivatives for each run (total 24 columns per run); iv) indicator vectors for outlier time points identified based on their multivariate distance from the other images in the sample (see above); v) indicator vectors for the first two images in each run. Low-frequency noise was removed by employing a high-pass filter of 180 seconds.

For each participant, we created the following three contrast maps, corresponding to the axioms for general aversive prediction errors in our task: (i) pain onset > no-pain onset; (ii) negative correlation with expected pain probability at pain onset; and (iii) negative correlation with expected pain probability at no-pain onset. We performed a second-level (group) analysis on each of these contrasts using robust regression (Wager et al. 2005). We tested for drug effects by including two second-level regressors coding for levodopa vs. placebo (weights [-1 1 0] for the treatment groups [P L N]) and naltrexone vs. placebo (weights [-1 0 1] for the treatment groups [P L N]). All maps were thresholded at FDR $q < 0.05$ corrected for multiple comparisons across the whole brain (gray matter masked), with higher (more conservative) voxel thresholds superimposed for display in Figure 3A.

To identify brain regions that fulfilled all three axioms for general aversive prediction errors, we examined the conjunction between the three second-level contrast maps described above (pain > no pain, negative correlation with expected pain probability on pain trials, and negative correlation with expected pain probability on no-pain trials), each of them thresholded at FDR $q < 0.05$. To identify brain regions that fulfilled all three axioms for general appetitive prediction errors, we examined the conjunction between the three opposite contrast maps (no pain > pain, positive correlation with expected pain probability on pain trials, and positive correlation with expected pain probability on no-pain trials).

Analysis of outcome-specific prediction-error signals

To examine outcome-specific prediction-error signals, we used the same first-level GLM as in the previous analysis, but created different contrast maps. First, to identify activation that tracks surprise more for received than avoided pain (Figure 4A), we used the following contrast: ‘negative correlation with expected pain probability at pain onset’ > ‘positive correlation with expected pain probability at no-pain onset’. Second, to identify activation tracking absolute prediction error (Figure 4B), we used the following contrast: ‘negative correlation with expected pain probability at pain onset’ > ‘negative correlation with expected pain probability at no-pain onset’. Note that this contrast is identical to: ‘positive correlation with expected pain probability at no-pain onset’ > ‘positive correlation with expected pain probability at pain onset’. It is also identical to a contrast with weights [1 1] for the ‘negative correlation with expected pain probability at pain onset’ and positive correlation with expected pain probability at no-pain onset’ regressors.

We performed a second-level analysis on each of these two contrasts using robust regression, again including two second-level regressors coding for levodopa vs. placebo and naltrexone vs. placebo. Maps were again thresholded at FDR $q < 0.05$ corrected for multiple comparisons across the whole brain. Adjacent areas thresholded at $p < 0.01$ and $p < 0.05$ (uncorrected) were added for display in Figure 4A and 4B.

Finally, to identify regions encoding outcome-specific prediction errors which cannot be explained by a general sensitivity to expected pain probability (also see the Results section) we examined the conjunction between the two second-level contrast maps described above, each thresholded at FDR $q < 0.05$ (Figure 4C).

Conventional prediction error analysis

We also conducted a conventional (not axiomatic) analysis to identify aversive prediction-error related brain activation. In this analysis, we created a single regressor modeling all outcome onsets (pain and no pain) and added aversive prediction error

(computed as the outcome [pain = 1, no pain = 0] minus the expected pain probability) as a parametric modulator on this outcome-onset regressor. The other regressors were identical to those in the previous analyses. We created first-level contrast images for the prediction error effect and conducted a second-level analysis on these contrast images, again thresholded at FDR $q < 0.05$ (Figure 3—figure supplement 1A). We also repeated this analysis while adding outcome (pain = 1, no pain = -1) as an additional parametric modulator on the outcome-onset regressor (without serial orthogonalization, such that only the unique variance explained by the prediction-error and outcome variables is assigned to their respective effects; Figure 3—figure supplement 1B).

Acknowledgements

We thank Stephanie Bauduin, Laila Franke, Nikki Nibbering, Iliana Samara, and Iris Spruit for help with data collection.

Competing interests

The authors declare that no competing interests exist.

Data availability

Single-trial behavioral data and model code are available from the OSF database:

<https://osf.io/rqc6g/>. Unthresholded t maps for all reported fMRI contrasts are available on Neurovault: <https://neurovault.org/collections/RIVRRMAK/> (<https://identifiers.org/neurovault.collection:6016>).

References

- 1 Ahn, W. Y., Haines, N. and Zhang, L. (2017). Revealing Neurocomputational Mechanisms of
2 Reinforcement Learning and Decision-Making With the hBayesDM Package. Comput
3 Psychiatr **1**: 24-57. doi: 10.1162/CPSY_a_00002.
- 4 Bartra, O., McGuire, J. T. and Kable, J. W. (2013). The valuation system: a coordinate-based
5 meta-analysis of BOLD fMRI experiments examining neural correlates of subjective
6 value. Neuroimage **76**: 412-427. doi: 10.1016/j.neuroimage.2013.02.063.
- 7 Beeler, J. A., Daw, N., Frazier, C. R. and Zhuang, X. (2010). Tonic dopamine modulates
8 exploitation of reward learning. Front Behav Neurosci **4**: 170. doi:
9 10.3389/fnbeh.2010.00170.
- 10 Beierholm, U., Guitart-Masip, M., Economides, M., Chowdhury, R., Duzel, E., Dolan, R. and
11 Dayan, P. (2013). Dopamine modulates reward-related vigor.
12 Neuropsychopharmacology **38**(8): 1495-1503. doi: 10.1038/npp.2013.48.
- 13 Black, K. J., Piccirillo, M. L., Koller, J. M., Hsieh, T., Wang, L. and Mintun, M. A. (2015).
14 Levodopa effects on [(11)C]raclopride binding in the resting human brain. F1000Res
15 **4**: 23. doi: 10.12688/f1000research.5672.1.
- 16 Boureau, Y. L. and Dayan, P. (2011). Opponency revisited: competition and cooperation
17 between dopamine and serotonin. Neuropsychopharmacology **36**(1): 74-97. doi:
18 10.1038/npp.2010.151.
- 19 Breitenstein, C., Korsukewitz, C., Floel, A., Kretschmar, T., Diederich, K. and Knecht, S.
20 (2006). Tonic dopaminergic stimulation impairs associative learning in healthy
21 subjects. Neuropsychopharmacology **31**(11): 2552-2564. doi:
22 10.1038/sj.npp.1301167.
- 23 Brischoux, F., Chakraborty, S., Brierley, D. I. and Ungless, M. A. (2009). Phasic excitation
24 of dopamine neurons in ventral VTA by noxious stimuli. Proc Natl Acad Sci U S A
25 **106**(12): 4894-4899. doi: 10.1073/pnas.0811507106.
- 26 Brocka, M., Helbing, C., Vincenz, D., Scherf, T., Montag, D., Goldschmidt, J., . . . Lippert,
27 M. (2018). Contributions of dopaminergic and non-dopaminergic neurons to VTA-

- 1 stimulation induced neurovascular responses in brain reward circuits. Neuroimage
- 2 **177**: 88-97. doi: 10.1016/j.neuroimage.2018.04.059.
- 3 Brooks, A. M. and Berns, G. S. (2013). Aversive stimuli and loss in the mesocorticolimbic
- 4 dopamine system. Trends Cogn Sci **17**(6): 281-286. doi: 10.1016/j.tics.2013.04.001.
- 5 Bunzeck, N., Guitart-Masip, M., Dolan, R. J. and Duzel, E. (2014). Pharmacological
- 6 dissociation of novelty responses in the human brain. Cereb Cortex **24**(5): 1351-1360.
- 7 doi: 10.1093/cercor/bhs420.
- 8 Bush, G., Luu, P. and Posner, M. I. (2000). Cognitive and emotional influences in anterior
- 9 cingulate cortex. Trends Cogn Sci **4**(6): 215-222. doi: 10.1016/s1364-6613(00)01483-
- 10 2.
- 11 Chakroun, K., Mathar, D., Wiehler, A., Ganzer, F. and Peters, J. (2020). Dopaminergic
- 12 modulation of the exploration/exploitation trade-off in human decision-making. Elife
- 13 **9**. doi: 10.7554/eLife.51260.
- 14 Chamberlain, S. R., Muller, U., Blackwell, A. D., Clark, L., Robbins, T. W. and Sahakian, B.
- 15 J. (2006). Neurochemical modulation of response inhibition and probabilistic learning
- 16 in humans. Science **311**(5762): 861-863. doi: 10.1126/science.1121218.
- 17 Chew, B., Hauser, T. U., Papoutsis, M., Magerkurth, J., Dolan, R. J. and Rutledge, R. B.
- 18 (2019). Endogenous fluctuations in the dopaminergic midbrain drive behavioral
- 19 choice variability. Proc Natl Acad Sci U S A **116**(37): 18732-18737. doi:
- 20 10.1073/pnas.1900872116.
- 21 Cole, S. and McNally, G. P. (2007). Opioid receptors mediate direct predictive fear learning:
- 22 evidence from one-trial blocking. Learn Mem **14**(4): 229-235. doi:
- 23 10.1101/lm.489507.
- 24 Cole, S. and McNally, G. P. (2009). Complementary roles for amygdala and periaqueductal
- 25 gray in temporal-difference fear learning. Learn Mem **16**(1): 1-7. doi:
- 26 10.1101/lm.1120509.
- 27 Cools, R., Nakamura, K. and Daw, N. D. (2011). Serotonin and dopamine: unifying affective,
- 28 activational, and decision functions. Neuropsychopharmacology **36**(1): 98-113. doi:
- 29 10.1038/npp.2010.121.
- 30 Cools, R., Robinson, O. J. and Sahakian, B. (2008). Acute tryptophan depletion in healthy
- 31 volunteers enhances punishment prediction but does not affect reward prediction.
- 32 Neuropsychopharmacology **33**(9): 2291-2299. doi: 10.1038/sj.npp.1301598.
- 33 Crockett, M. J., Clark, L., Roiser, J. P., Robinson, O. J., Cools, R., Chase, H. W., . . .
- 34 Robbins, T. W. (2012). Converging evidence for central 5-HT effects in acute
- 35 tryptophan depletion. Mol Psychiatry **17**(2): 121-123. doi: 10.1038/mp.2011.106.
- 36 Daw, N. D., Kakade, S. and Dayan, P. (2002). Opponent interactions between serotonin and
- 37 dopamine. Neural Netw **15**(4-6): 603-616. doi: 10.1016/s0893-6080(02)00052-7.
- 38 Dayan, P. and Huys, Q. J. (2009). Serotonin in affective control. Annu Rev Neurosci **32**: 95-
- 39 126. doi: 10.1146/annurev.neuro.051508.135607.
- 40 Deakin, J. F. and Graeff, F. G. (1991). 5-HT and mechanisms of defence. J Psychopharmacol
- 41 **5**(4): 305-315. doi: 10.1177/026988119100500414.
- 42 Deakin, J. F. W. (1983). Roles of serotonergic systems in escape, avoidance and other
- 43 behaviours. Theory in Psychopharmacology. S. J. Cooper. London, Academic Press.
- 44 **2**: 149-193.
- 45 Deichmann, R., Gottfried, J. A., Hutton, C. and Turner, R. (2003). Optimized EPI for fMRI
- 46 studies of the orbitofrontal cortex. Neuroimage **19**(2 Pt 1): 430-441. doi:
- 47 10.1016/s1053-8119(03)00073-9.
- 48 den Ouden, H. E., Daw, N. D., Fernandez, G., Elshout, J. A., Rijpkema, M., Hoogman, M., . . .
- 49 . Cools, R. (2013). Dissociable effects of dopamine and serotonin on reversal
- 50 learning. Neuron **80**(4): 1090-1100. doi: 10.1016/j.neuron.2013.08.030.

- 1 Dinsmoor, J. A. (2001). Stimuli inevitably generated by behavior that avoids electric shock
2 are inherently reinforcing. *J Exp Anal Behav* **75**(3): 311-333. doi:
3 10.1901/jeab.2001.75-311.
- 4 Dombrowski, P. A., Maia, T. V., Boschen, S. L., Bortolanza, M., Wendler, E., Schwarting, R.
5 K., . . . Da Cunha, C. (2013). Evidence that conditioned avoidance responses are
6 reinforced by positive prediction errors signaled by tonic striatal dopamine. *Behav*
7 *Brain Res* **241**: 112-119. doi: 10.1016/j.bbr.2012.06.031.
- 8 Eippert, F., Bingel, U., Schoell, E., Yacubian, J. and Buchel, C. (2008). Blockade of
9 endogenous opioid neurotransmission enhances acquisition of conditioned fear in
10 humans. *J Neurosci* **28**(21): 5465-5472. doi: 10.1523/JNEUROSCI.5336-07.2008.
- 11 Eldar, E., Hauser, T. U., Dayan, P. and Dolan, R. J. (2016). Striatal structure and function
12 predict individual biases in learning to avoid pain. *Proc Natl Acad Sci U S A* **113**(17):
13 4812-4817. doi: 10.1073/pnas.1519829113.
- 14 Fernando, A. B., Urcelay, G. P., Mar, A. C., Dickinson, A. and Robbins, T. W. (2013).
15 Comparison of the conditioned reinforcing properties of a safety signal and appetitive
16 stimulus: effects of d-amphetamine and anxiolytics. *Psychopharmacology (Berl)*
17 **227**(2): 195-208. doi: 10.1007/s00213-012-2952-1.
- 18 Floel, A., Garraux, G., Xu, B., Breitenstein, C., Knecht, S., Herscovitch, P. and Cohen, L. G.
19 (2008). Levodopa increases memory encoding and dopamine release in the striatum in
20 the elderly. *Neurobiol Aging* **29**(2): 267-279. doi:
21 10.1016/j.neurobiolaging.2006.10.009.
- 22 Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: a
23 neurocomputational account of cognitive deficits in medicated and nonmedicated
24 Parkinsonism. *J Cogn Neurosci* **17**(1): 51-72. doi: 10.1162/0898929052880093.
- 25 Frank, M. J., Seeberger, L. C. and O'Reilly R, C. (2004). By carrot or by stick: cognitive
26 reinforcement learning in parkinsonism. *Science* **306**(5703): 1940-1943. doi:
27 10.1126/science.1102941.
- 28 Gallagher, D. A., O'Sullivan, S. S., Evans, A. H., Lees, A. J. and Schrag, A. (2007).
29 Pathological gambling in Parkinson's disease: risk factors and differences from
30 dopamine dysregulation. An analysis of published case series. *Mov Disord* **22**(12):
31 1757-1763. doi: 10.1002/mds.21611.
- 32 Gelman, A. (2014). *Bayesian data analysis. Third edition*. Boca Raton, CRC Press.
- 33 Gelman, A., Hwang, J. and Vehtari, A. (2014). Understanding predictive information criteria
34 for Bayesian models. *Statistics and Computing* **24**: 997-1016.
- 35 Gershman, S. J. and Tzovaras, B. G. (2018). Dopaminergic genes are associated with both
36 directed and random exploration. *Neuropsychologia* **120**: 97-104. doi:
37 10.1016/j.neuropsychologia.2018.10.009.
- 38 Guitart-Masip, M., Chowdhury, R., Sharot, T., Dayan, P., Duzel, E. and Dolan, R. J. (2012).
39 Action controls dopaminergic enhancement of reward representations. *Proc Natl Acad*
40 *Sci U S A* **109**(19): 7511-7516. doi: 10.1073/pnas.1202229109.
- 41 Guitart-Masip, M., Economides, M., Huys, Q. J., Frank, M. J., Chowdhury, R., Duzel, E., . . .
42 Dolan, R. J. (2014). Differential, but not opponent, effects of L-DOPA and
43 citalopram on action learning with reward and punishment. *Psychopharmacology*
44 *(Berl)* **231**(5): 955-966. doi: 10.1007/s00213-013-3313-4.
- 45 Hauser, T. U., Iannaccone, R., Walitza, S., Brandeis, D. and Brem, S. (2015). Cognitive
46 flexibility in adolescence: neural and behavioral mechanisms of reward prediction
47 error processing in adaptive decision making during development. *Neuroimage* **104**:
48 347-354. doi: 10.1016/j.neuroimage.2014.09.018.
- 49 Hindi Attar, C., Finckh, B. and Buchel, C. (2012). The influence of serotonin on fear
50 learning. *PLoS One* **7**(8): e42397. doi: 10.1371/journal.pone.0042397.

- 1 Huys, Q. J., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P. and Roiser, J. P. (2012). Bonsai
2 trees in your head: how the pavlovian system sculpts goal-directed choices by pruning
3 decision trees. *PLoS Comput Biol* **8**(3): e1002410. doi:
4 10.1371/journal.pcbi.1002410.
- 5 Jepma, M., Jones, M. and Wager, T. D. (2014). The dynamics of pain: evidence for
6 simultaneous site-specific habituation and site-nonspecific sensitization in thermal
7 pain. *J Pain* **15**(7): 734-746. doi: 10.1016/j.jpain.2014.02.010.
- 8 Josselyn, S. A., Falls, W. A., Gewirtz, J. C., Pistell, P. and Davis, M. (2005). The nucleus
9 accumbens is not critically involved in mediating the effects of a safety signal on
10 behavior. *Neuropsychopharmacology* **30**(1): 17-26. doi: 10.1038/sj.npp.1300530.
- 11 Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical*
12 *Association* **90**: 773-795.
- 13 Kayser, A. S., Mitchell, J. M., Weinstein, D. and Frank, M. J. (2015). Dopamine, locus of
14 control, and the exploration-exploitation tradeoff. *Neuropsychopharmacology* **40**(2):
15 454-462. doi: 10.1038/npp.2014.193.
- 16 Kim, J. H. and Richardson, R. (2009). The effect of the mu-opioid receptor antagonist
17 naloxone on extinction of conditioned fear in the developing rat. *Learn Mem* **16**(3):
18 161-166. doi: 10.1101/lm.1282309.
- 19 Knutson, B. and Gibbs, S. E. (2007). Linking nucleus accumbens dopamine and blood
20 oxygenation. *Psychopharmacology (Berl)* **191**(3): 813-822. doi: 10.1007/s00213-006-
21 0686-7.
- 22 Kroemer, N. B., Lee, Y., Pooeh, S., Eppinger, B., Goschke, T. and Smolka, M. N. (2019). L-
23 DOPA reduces model-free control of behavior by attenuating the transfer of value to
24 action. *Neuroimage* **186**: 113-125. doi: 10.1016/j.neuroimage.2018.10.075.
- 25 Lammel, S., Ion, D. I., Roeper, J. and Malenka, R. C. (2011). Projection-specific modulation
26 of dopamine neuron synapses by aversive and rewarding stimuli. *Neuron* **70**(5): 855-
27 862. doi: 10.1016/j.neuron.2011.03.025.
- 28 Lee, M. C., Wagner, H. N., Jr., Tanada, S., Frost, J. J., Bice, A. N. and Dannals, R. F. (1988).
29 Duration of occupancy of opiate receptors by naltrexone. *J Nucl Med* **29**(7): 1207-
30 1211.
- 31 Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S. and Palminteri, S. (2017).
32 Behavioural and neural characterization of optimistic reinforcement learning. *Nature*
33 *Human Behaviour* **1**.
- 34 Maia, T. V. and Frank, M. J. (2011). From reinforcement learning models to psychiatric and
35 neurological disorders. *Nat Neurosci* **14**(2): 154-162. doi: 10.1038/nn.2723.
- 36 Matsumoto, M. and Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey
37 positive and negative motivational signals. *Nature* **459**(7248): 837-841. doi:
38 10.1038/nature08028.
- 39 McNally, G. P. and Cole, S. (2006). Opioid receptors in the midbrain periaqueductal gray
40 regulate prediction errors during pavlovian fear conditioning. *Behav Neurosci* **120**(2):
41 313-323. doi: 10.1037/0735-7044.120.2.313.
- 42 McNally, G. P., Johansen, J. P. and Blair, H. T. (2011). Placing prediction into the fear
43 circuit. *Trends Neurosci* **34**(6): 283-292. doi: 10.1016/j.tins.2011.03.005.
- 44 McNally, G. P., Lee, B. W., Chiem, J. Y. and Choi, E. A. (2005). The midbrain
45 periaqueductal gray and fear extinction: opioid receptor subtype and roles of cyclic
46 AMP, protein kinase A, and mitogen-activated protein kinase. *Behav Neurosci*
47 **119**(4): 1023-1033. doi: 10.1037/0735-7044.119.4.1023.
- 48 McNally, G. P., Pigg, M. and Weidemann, G. (2004). Opioid receptors in the midbrain
49 periaqueductal gray regulate extinction of pavlovian fear conditioning. *J Neurosci*
50 **24**(31): 6912-6919. doi: 10.1523/JNEUROSCI.1828-04.2004.

- 1 McNally, G. P. and Westbrook, R. F. (2003). Opioid receptors regulate the extinction of
2 Pavlovian fear conditioning. Behav Neurosci **117**(6): 1292-1301. doi: 10.1037/0735-
3 7044.117.6.1292.
- 4 Mirenowicz, J. and Schultz, W. (1996). Preferential activation of midbrain dopamine neurons
5 by appetitive rather than aversive stimuli. Nature **379**(6564): 449-451. doi:
6 10.1038/379449a0.
- 7 Moustafa, A. A., Herzallah, M. M. and Gluck, M. A. (2013). Dissociating the cognitive
8 effects of levodopa versus dopamine agonists in a neurocomputational model of
9 learning in Parkinson's disease. Neurodegener Dis **11**(2): 102-111. doi:
10 10.1159/000341999.
- 11 Moutoussis, M., Bentall, R. P., Williams, J. and Dayan, P. (2008). A temporal difference
12 account of avoidance learning. Network **19**(2): 137-160. doi:
13 10.1080/09548980802192784.
- 14 Mowrer, O. H. (1951). Two-factor learning theory: summary and comment. Psychol Rev
15 **58**(5): 350-354. doi: 10.1037/h0058956.
- 16 Mowrer, O. H. (1956). Two-factor learning theory reconsidered, with special reference to
17 secondary reinforcement and the concept of habit. Psychol Rev **63**(2): 114-128. doi:
18 10.1037/h0040613.
- 19 O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H. and Dolan, R. J. (2003). Temporal
20 difference models and reward-related learning in the human brain. Neuron **38**(2): 329-
21 337. doi: 10.1016/s0896-6273(03)00169-7.
- 22 Oei, N. Y., Rombouts, S. A., Soeter, R. P., van Gerven, J. M. and Both, S. (2012). Dopamine
23 modulates reward system activity during subconscious processing of sexual stimuli.
24 Neuropsychopharmacology **37**(7): 1729-1737. doi: 10.1038/npp.2012.19.
- 25 Ojemann, J. G., Akbudak, E., Snyder, A. Z., McKinstry, R. C., Raichle, M. E. and Conturo,
26 T. E. (1997). Anatomic localization and quantitative analysis of gradient refocused
27 echo-planar fMRI susceptibility artifacts. Neuroimage **6**(3): 156-167. doi:
28 10.1006/nimg.1997.0289.
- 29 Oleson, E. B., Gentry, R. N., Chioma, V. C. and Cheer, J. F. (2012). Subsecond dopamine
30 release in the nucleus accumbens predicts conditioned punishment and its successful
31 avoidance. J Neurosci **32**(42): 14804-14808. doi: 10.1523/JNEUROSCI.3087-
32 12.2012.
- 33 Palminteri, S., Clair, A. H., Mallet, L. and Pessiglione, M. (2012). Similar improvement of
34 reward and punishment learning by serotonin reuptake inhibitors in obsessive-
35 compulsive disorder. Biol Psychiatry **72**(3): 244-250. doi:
36 10.1016/j.biopsych.2011.12.028.
- 37 Palminteri, S., Lefebvre, G., Kilford, E. J. and Blakemore, S. J. (2017). Confirmation bias in
38 human reinforcement learning: Evidence from counterfactual feedback processing.
39 PLoS Comput Biol **13**(8): e1005684. doi: 10.1371/journal.pcbi.1005684.
- 40 Palminteri, S. and Pessiglione, M. (2017). Opponent brain systems for reward and
41 punishment learning: causal evidence from drug and lesion studies in humans.
42 Decision Neuroscience: An Integrative Perspective. J. C. Dreher and L. Tremblay,
43 Academic Press.
- 44 Parsons, R. G., Gafford, G. M. and Helmstetter, F. J. (2010). Regulation of extinction-related
45 plasticity by opioid receptors in the ventrolateral periaqueductal gray matter. Front
46 Behav Neurosci **4**. doi: 10.3389/fnbeh.2010.00044.
- 47 Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J. and Frith, C. D. (2006). Dopamine-
48 dependent prediction errors underpin reward-seeking behaviour in humans. Nature
49 **442**(7106): 1042-1045. doi: 10.1038/nature05051.

- 1 Phillips, R. G. and LeDoux, J. E. (1992). Differential contribution of amygdala and
2 hippocampus to cued and contextual fear conditioning. Behav Neurosci **106**(2): 274-
3 285. doi: 10.1037//0735-7044.106.2.274.
- 4 Pine, A., Shiner, T., Seymour, B. and Dolan, R. J. (2010). Dopamine, time, and impulsivity in
5 humans. J Neurosci **30**(26): 8888-8896. doi: 10.1523/JNEUROSCI.6028-09.2010.
- 6 Ploghaus, A., Tracey, I., Clare, S., Gati, J. S., Rawlins, J. N. and Matthews, P. M. (2000).
7 Learning about pain: the neural substrate of the prediction error for aversive events.
8 Proc Natl Acad Sci U S A **97**(16): 9281-9286. doi: 10.1073/pnas.160266497.
- 9 Poletti, M. and Bonuccelli, U. (2013). Acute and chronic cognitive effects of levodopa and
10 dopamine agonists on patients with Parkinson's disease: a review. Ther Adv
11 Psychopharmacol **3**(2): 101-113. doi: 10.1177/2045125312470130.
- 12 Preston, K. L. and Bigelow, G. E. (1993). Differential naltrexone antagonism of
13 hydromorphone and pentazocine effects in human volunteers. J Pharmacol Exp Ther
14 **264**(2): 813-823.
- 15 Ptak, R., Schnider, A. and Fellrath, J. (2017). The Dorsal Frontoparietal Network: A Core
16 System for Emulated Action. Trends Cogn Sci **21**(8): 589-599. doi:
17 10.1016/j.tics.2017.05.002.
- 18 Raczka, K. A., Mechias, M. L., Gartmann, N., Reif, A., Deckert, J., Pessiglione, M. and
19 Kalisch, R. (2011). Empirical support for an involvement of the mesostriatal
20 dopamine system in human fear extinction. Transl Psychiatry **1**: e12. doi:
21 10.1038/tp.2011.10.
- 22 Rescorla, R. A. and Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in
23 the effectiveness of reinforcement and nonreinforcement. Classical conditioning II:
24 current research and theory. A. H. Black and W. F. Prokasy. New York, Appleton-
25 Century-Crofts: 64-99.
- 26 Rich, E. L. and Wallis, J. D. (2016). Decoding subjective decisions from orbitofrontal cortex.
27 Nat Neurosci **19**(7): 973-980. doi: 10.1038/nn.4320.
- 28 Robinson, O. J., Cools, R. and Sahakian, B. J. (2012). Tryptophan depletion disinhibits
29 punishment but not reward prediction: implications for resilience.
30 Psychopharmacology (Berl) **219**(2): 599-605. doi: 10.1007/s00213-011-2410-5.
- 31 Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G. E. and Wager, T. D. (2014).
32 Representation of aversive prediction errors in the human periaqueductal gray. Nat
33 Neurosci **17**(11): 1607-1612. doi: 10.1038/nn.3832.
- 34 Roy, M., Shohamy, D. and Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems
35 and the generation of affective meaning. Trends Cogn Sci **16**(3): 147-156. doi:
36 10.1016/j.tics.2012.01.005.
- 37 Rutledge, R. B., Dean, M., Caplin, A. and Glimcher, P. W. (2010). Testing the reward
38 prediction error hypothesis with an axiomatic model. J Neurosci **30**(40): 13525-
39 13536. doi: 10.1523/JNEUROSCI.1747-10.2010.
- 40 Salinas-Hernandez, X. I., Vogel, P., Betz, S., Kalisch, R., Sigurdsson, T. and Duvarci, S.
41 (2018). Dopamine neurons drive fear extinction learning by signaling the omission of
42 expected aversive outcomes. Elife **7**. doi: 10.7554/eLife.38818.
- 43 Schuh, K. J., Walsh, S. L. and Stitzer, M. L. (1999). Onset, magnitude and duration of opioid
44 blockade produced by buprenorphine and naltrexone in humans. Psychopharmacology
45 (Berl) **145**(2): 162-174. doi: 10.1007/s002130051045.
- 46 Schultz, W. (2019). Recent advances in understanding the role of phasic dopamine activity.
47 F1000Res **8**. doi: 10.12688/f1000research.19793.1.
- 48 Schultz, W., Dayan, P. and Montague, P. R. (1997). A neural substrate of prediction and
49 reward. Science **275**(5306): 1593-1599.

- 1 Seymour, B., Daw, N. D., Roiser, J. P., Dayan, P. and Dolan, R. (2012). Serotonin selectively
2 modulates reward value in human decision-making. J Neurosci **32**(17): 5833-5842.
3 doi: 10.1523/JNEUROSCI.0053-12.2012.
- 4 Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., . . .
5 Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning
6 in humans. Nature **429**(6992): 664-667. doi: 10.1038/nature02581.
- 7 Seymour, B., O'Doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K. and
8 Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive
9 learning of pain relief. Nat Neurosci **8**(9): 1234-1240. doi: 10.1038/nn1527.
- 10 Sharot, T. and Garrett, N. (2016). Forming Beliefs: Why Valence Matters. Trends Cogn Sci
11 **20**(1): 25-33. doi: 10.1016/j.tics.2015.11.002.
- 12 Soubrie, P. (1986). Reconciling the role of central serotonin neurons in human and animal
13 behavior. Behavioral and Brain Sciences **9**(2): 319-335.
- 14 St Onge, J. R. and Floresco, S. B. (2009). Dopaminergic modulation of risk-based decision
15 making. Neuropsychopharmacology **34**(3): 681-697. doi: 10.1038/npp.2008.121.
- 16 Stelly, C. E., Haug, G. C., Fonzi, K. M., Garcia, M. A., Tritley, S. C., Magnon, A. P., . . .
17 Wanat, M. J. (2019). Pattern of dopamine signaling during aversive events predicts
18 active avoidance learning. Proc Natl Acad Sci U S A **116**(27): 13641-13650. doi:
19 10.1073/pnas.1904249116.
- 20 Sutton, R. S. and Barto, A. G. (1998). Reinforcement learning : an introduction. Cambridge,
21 Mass., MIT Press.
- 22 Ungless, M. A., Magill, P. J. and Bolam, J. P. (2004). Uniform inhibition of dopamine
23 neurons in the ventral tegmental area by aversive stimuli. Science **303**(5666): 2040-
24 2042. doi: 10.1126/science.1093360.
- 25 Vellani, V., de Vries, L. P., Gaule, A. and Sharot, T. (2020). A selective effect of dopamine
26 on information-seeking. Elife **9**. doi: 10.7554/eLife.59152.
- 27 Voon, V., Hassan, K., Zurowski, M., Duff-Canning, S., de Souza, M., Fox, S., . . . Miyasaki,
28 J. (2006). Prospective prevalence of pathologic gambling and medication association
29 in Parkinson disease. Neurology **66**(11): 1750-1752. doi:
30 10.1212/01.wnl.0000218206.20920.4d.
- 31 Voon, V., Pessiglione, M., Brezing, C., Gallea, C., Fernandez, H. H., Dolan, R. J. and Hallett,
32 M. (2010). Mechanisms underlying dopamine-mediated reward bias in compulsive
33 behaviors. Neuron **65**(1): 135-142. doi: 10.1016/j.neuron.2009.12.027.
- 34 Wager, T. D., Keller, M. C., Lacey, S. C. and Jonides, J. (2005). Increased sensitivity in
35 neuroimaging analyses using robust regression. Neuroimage **26**(1): 99-113. doi:
36 10.1016/j.neuroimage.2005.01.011.
- 37 Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely
38 applicable information criterion in singular learning theory. Journal of Machine
39 Learning Research **11**: 3571-3594.
- 40 Weerts, E. M., McCaul, M. E., Kuwabara, H., Yang, X., Xu, X., Dannals, R. F., . . . Wand, G.
41 S. (2013). Influence of OPRM1 Asn40Asp variant (A118G) on [11C]carfentanil
42 binding potential: preliminary findings in human subjects. Int J
43 Neuropsychopharmacol **16**(1): 47-53. doi: 10.1017/S146114571200017X.
- 44 Wenzel, J. M., Oleson, E. B., Gove, W. N., Cole, A. B., Gyawali, U., Dantrassy, H. M., . . .
45 Cheer, J. F. (2018). Phasic Dopamine Signals in the Nucleus Accumbens that Cause
46 Active Avoidance Require Endocannabinoid Mobilization in the Midbrain. Curr Biol
47 **28**(9): 1392-1404 e1395. doi: 10.1016/j.cub.2018.03.037.
- 48 Wietzikoski, E. C., Boschen, S. L., Miyoshi, E., Bortolanza, M., Dos Santos, L. M., Frank,
49 M., . . . Da Cunha, C. (2012). Roles of D1-like dopamine receptors in the nucleus

- 1   accumbens and dorsolateral striatum in conditioned avoidance responses.
- 2   Psychopharmacology (Berl) **219**(1): 159-169. doi: 10.1007/s00213-011-2384-3.
- 3   Wise, T., Michely, J., Dayan, P. and Dolan, R. J. (2019). A computational account of threat-
- 4   related attentional bias. PLoS Comput Biol **15**(10): e1007341. doi:
- 5   10.1371/journal.pcbi.1007341.
- 6   Wittmann, B. C. and D'Esposito, M. (2015). Levodopa administration modulates striatal
- 7   processing of punishment-associated items in healthy participants.
- 8   Psychopharmacology (Berl) **232**(1): 135-144. doi: 10.1007/s00213-014-3646-7.
- 9   Yarkoni, T. (2020). The generalizability crisis. Behav Brain Sci: 1-37. doi:
- 10   10.1017/S0140525X20001685.
- 11   Zhang, S., Mano, H., Lee, M., Yoshida, W., Kawato, M., Robbins, T. W. and Seymour, B.
- 12   (2018). The control of tonic pain by active relief learning. Elife **7**. doi:
- 13   10.7554/eLife.31949.

Figure 1 – Figure supplement 1. Pre- and post-treatment ratings of alertness, calmness, and contentment in each treatment group. Error bars indicate standard errors. The participants assigned to the three treatment groups did not differ in their pre-treatment ratings of alertness, calmness or contentment (all p 's > 0.3). To assess potential treatment effects on subjective state we conducted analyses of covariance (ANCOVAs) on the post-treatment ratings of alertness, calmness and contentment (made two hours after drug intake), with treatment as a between-subject factor and the pre-treatment ratings as covariate. There was no effect of treatment on any of these ratings (all p 's > 0.08), suggesting that the drugs did not affect subjective state.

Figure 1 – Figure supplement 2. Pain ratings during a pain-rating task that preceded the pain-avoidance learning task, as a function of stimulus temperature and treatment group. Error bars indicate standard errors. Participants received five 47°C, five 49°C, and five 50°C heat stimuli, in random order, to their left lower leg (ramp rate = 40°C/second; 1 second at peak temperature, stimulus onset asynchrony = 17-25 seconds). Following each heat stimulus participants rated their experienced pain on a 100-unit visual analog scale with anchors of “no pain” and “worst-imaginable pain”, respectively. We conducted a mixed ANOVA on participants’ pain ratings with stimulus temperature (47°C, 49°C and 50°C coded as -1, 0 and 1, respectively) as within-subject factor and treatment as between-subject factor. Pain ratings increased as a function of stimulus temperature ($F(1,84) = 424, p < 0.001$). However, there was no main effect of treatment ($F(2,84) = .21, p = .81$) and no temperature x treatment interaction ($F(2,84) = .76, p = .47$). Thus, the drugs did not affect the subjective pain experience evoked by heat-pain stimuli. Note that four participants who were included in this analysis were excluded from the fMRI analyses of the pain-avoidance learning task, because of excessive head movement.

Figure 2 – Figure supplement 1. 95% highest density intervals of the posterior distributions shown in Figure 2A.

	placebo	levodopa	naltrexone
$\bar{\alpha}_{pain}$	0.53-0.90	0.47-0.85	0.55-0.87
$\bar{\alpha}_{no-pain}$	0.20-0.47	0.49-0.84	0.58-0.94
$\bar{\beta}$	6.3-11.7	3.9-7.2	4.2-7.3

Figure 2 – Figure supplement 2. 95% HDIs of the posterior distributions of each participant's learning rate for pain (α_{pain}) and no-pain ($\alpha_{no-pain}$) outcomes. Participants are sorted according to the difference between their two learning rates ($\alpha_{pain} - \alpha_{no-pain}$). Note that α_{pain} and $\alpha_{no-pain}$ were positively correlated in the levodopa group ($r = .43, p = .03$), but were not correlated in the placebo ($r = -.16, p = .4$) and naltrexone ($r = .31, p = .10$) group.

Figure 2 – Figure supplement 3. Modeling results from an independent sample of untreated participants from a previous study (N = 23), replicating the asymmetric learning rates ($\bar{\alpha}_{pain} > \bar{\alpha}_{no-pain}$) found in our placebo group.

Figure 3 – Figure supplement 2. Activation for each of our three prediction-error contrasts, extracted from the clusters identified by our conjunction analysis, per treatment group. The three bars in each plot correspond to the placebo, levodopa and naltrexone group, respectively. There was only one significant treatment effect: The negative correlation with expected pain probability on no-pain trials (axiom 3) in the PAG cluster was stronger in the placebo group than in the drug groups ($p = .04$). As this result did not survive correction for multiple tests, and there were no significant treatment effects for the other prediction-error contrasts in any of the clusters, we conclude that our pharmacological manipulations did not affect prediction-error related brain activation. Error bars indicate standard errors. * $p < .05$ (uncorrected for multiple tests).