Adaptive coding of pain prediction error in the anterior insula

Hoskin, R[1] & Talmi, D[12]

1 University of Manchester

2 University of Cambridge

Running header: Adaptive coding of pain prediction error

Corresponding author: Deborah Talmi, Department of Psychology, University of Cambridge, Downing site, Cambridge, UK dt492@cam.ac.uk

Original article submitted to the European Journal of Pain research.
Funding source. ESRC ES/I010424/1.

Significance statement: There is behavioural evidence that pain is coded adaptively but controversy as to the neural mechanism that serves this process, and limited understanding of how adaptive coding of pain relates to adaptively-coded pain prediction error. We provide evidence that the left anterior insula complies with axioms that describe adaptive coding of pain prediction error, adding to the understanding of the role it plays in subjective pain representation.

Abstract

Background. To reduce the computational demands of the task of determining values, the brain is thought to engage in ‘adaptive coding’, where the sensitivity of some neurons to value is modulated by contextual information. There is good behavioural evidence that pain is coded adaptively, but controversy regarding the underlying neural mechanism. Additionally, there is evidence that reward prediction errors are coded adaptively, but no parallel evidence regarding pain prediction errors.

Methods. We tested the hypothesis that pain prediction errors are coded adaptively by scanning 19 healthy adults with fMRI while they performed a cued pain task. Our analysis followed an axiomatic approach.

Results. We found that the left anterior insula was the only region which was sensitive both to predicted pain magnitudes and the unexpectedness of pain delivery, but not to the magnitude of delivered pain.

Conclusions. This pattern suggests that the left anterior insula is part of a neural mechanism that serves the adaptive prediction error of pain.

Introduction

One function required of the nervous system is to quickly identify the subjective value of stimuli. A problem with such a task is that the range of stimulus intensities is wide, while the processing range of neurons is limited. To reduce the computational demands of determining values, the brain engages in ‘adaptive coding’, where the sensitivity of some neurons to value is modulated by contextual information (Seymour & McClure, 2008). Adaptive coding of reward is implemented in dopaminergic neuronal firing (Tobler et al., 2005). In each trial of Tobler and colleagues’ seminal experiment, Macaques observed a cue that signalled 50% chance of receiving reward. There were three cues. Each signalled one reward magnitude (low, medium, or large). Following training, each cue triggered a midbrain dopaminergic signal proportional to the magnitude of the reward it predicted. On trials where the reward was delivered, higher-than-baseline response in the same neurons reflected the Prediction Error (PE) of receiving the reward. Crucially, this response no longer distinguished reward magnitudes, remaining the same regardless of actual reward delivered. Thus, once expectations were set via the cue, the dopaminergic signal reflected an adaptively-coded PE, which diverged from the computational PE (Fig1). Subsequent studies obtained substantial evidence for adaptively-coded reward prediction error in human ventral striatum (Park et al., 2012).

FIGURE 1

As with reward, PEs triggered by the prediction or experience of aversive stimuli, such as pain is also coded by dopamine neurons (Lammel et al., 2011; Schultz, 2016). Computational pain PE signal correlates with activation in the human ventral striatum and insula (Geuter et al., 2017; Seymour et al., 2004; Shih et al., 2019), although only in the periaqueductal grey did the signal adhere to all three axioms of computational pain PE (Roy et al., 2014): stronger response to pain compared to no-pain outcomes, unless fully expected; stronger response to unexpected pain; and equivalent signal for fully-expected outcomes (Caplin & Dean, 2008; Rutledge et al., 2010). These studies searched for neural signal that tracked the computational PE signal; they have not tested if the neural response demonstrated adaptive coding.

There is ample evidence that the local context influences behavioural responses to pain so that they scale with expectations (Atlas & Wager, 2012; Tracey, 2010). Behavioural evidence specific for adaptive coding of pain is evident in that participants pay more for relief of moderate pain when the pain intensities they could receive range from low to moderate, compared when they range from moderate to high (Vlaev et al., 2009; Winston et al., 2014). There is controversy about the neural mechanism that enables adaptive coding of pain, with research implicating either the insula and ACC (Leknes et al., 2013), or the orbitofrontal cortex (Winston et al., 2014); one study reported null neural effects (Bauch et al., 2017). These studies have not focused specifically on whether a region represents pain PEs in an adaptively-coded manner. We set out to examine whether neural evidence for adaptive coding of pain PEs can be acquired by following Tobler et al.'s (2005) design more closely.

Methodology

Experimental design

The experiment employed a 2X2 event-related design with pain magnitude (high vs. low) and pain probability (high vs. low) as the factors. Our task resembled one we have previously employed to detect neural signature of pain and reward PEs on the scalp (Talmi et al., 2013). Participants viewed chance cues that predicted electric skin stimulation with a 33% or 66% probability. In each trial it was only possible to receive a stimulation of a single magnitude, with the cue colour being fully predictive of which magnitude this could be (Fig2). Across the experiment two different magnitudes were used: a 'high' level, equivalent to the most intense level of stimulation participants had previously indicated they were willing to tolerate, and a 'low' level, equivalent to a stimulation level that participants had indicated was "just painful". Importantly, an event related design was used such that every block contained each of the 4 trials type (high and low probability and high and low pain). The trials were designed to permit the cue to adaptively scale the participants' expectations, such that when the cue signalled a chance of low stimulation participants expected low (or no) pain, and when the cue signalled a chance of high stimulation they expected high (or no) pain.

A brain region that codes for pain PE should, according to the axiomatic model of computational PE (Caplin & Dean, 2008; Rutledge et al., 2010), respond more strongly in delivered pain than omitted pain trials (axiom 1); and it should respond more strongly when the delivered stimulation was unexpected (33% cue) than when

it was expected (66% pain, axiom 2). In our experimental paradigm, because pain delivery was necessarily probabilistic, we could not assess the third axiom, of equivalent signal for fully-expected outcomes, but we could assess the PE signal associated with the cue. A region that codes for pain PE should respond more to chance cues that predicts high pain than to the chance cue that predict low pain. Crucially, and diverging from the computational PE signal, once the specific utility of pain is set through expectations, the firing of a region that codes for pain PE adaptively should only reflect the deviance from expectation; it should not otherwise scale with the intensity of the delivered stimulus. Put another way, the region should respond equally strongly to unexpected pain (compared to expected pain), regardless of the absolute magnitude of the unexpected pain, because in the local context of each trial, the unexpected pain represents the highest possible pain intensity. By contrast, a region that codes for the computational PE will produce a stronger signal for unexpected pain when stronger stimulation is delivered.

FIGURE 2

Participants

Twenty participants took part in the experiment, although data collection was aborted for one participant as they could not tolerate being in the scanner, leaving a final sample of 19 participants (11 female) between the ages of 18-36 (M=25.16, SD=4.35). This sample size was based on Hoskin et al. (2019) where a similar sample size was sufficient to reveal effects of expectations on pain experience, and

with the sample sizes used by previous studies on adaptive pain PE (Bauch et al., 2017; Leknes et al., 2013; Winston et al., 2014). For two participants data was only collected for 3 (of the 4) blocks due to equipment failure. All participants were screened for any conditions that would prevent MR scanning, and for psychiatric and neurological history. Participants were proficient in English, had normal or corrected-to-normal hearing and vision, and did not take centrally-acting medication. The study received ethical approval from the University of Manchester ethics committee. Sample characteristics are described in Table 1.

Apparatus

Pain stimulation. The electrical stimulations were delivered to the back of the right hand via an in-house built ring electrode (Medical Physics, Salford Royal Hospital) attached to a Digitimer DS5 Isolated Bipolar Constant Current Stimulator (http://www.digitimer.com/). To counter the effect of the magnetic field of the MR scanner, the DS5 stimulator was placed within a custom-built Faraday cage (Medical Physics, Salford Royal Hospital). For reasons of participant safety, this stimulator was limited to delivering a maximum of 5V during the experiment. To ensure adequate conductance between the electrode and the skin, the back of each participant's hand was prepared with Nuprep Skin Preparation Gel and Ten20 Conductive Paste. The SCR electrodes were placed on the inside medial phalange of the second and fourth fingers of the participant's left hand. The inputs to the DS5 machine were controlled via 1401plus data acquisition interface connected to a laptop running the Spike2 software (Cambridge Electronic Designs, Cambridge, UK). Both the laptop and the 1401 machine were kept inside the scan control room during data collection. The 1401 machine was connected to the the faraday cage using a

BAC cable. To avoid interference between the faraday cage and the scanner, the faraday cage was kept a minimum of 2m from the scanner. The experiment was delivered via Cogent2000 on a Matlab platform (www.Mathworks.com).

Neuroimaging. Participants were scanned using a 3T Philips Achieva scanner, fitted with a Philips 32-channel receive-only coil. Whole-brain functional images were collected using a single-shot dual-echo protocol (Halai et al., 2014), with TR=3s, TE=12ms & 35ms, FOV=240,240,132, flip angle=85. In each volume, 33 slices of voxel size 3x3x4mm were collected in ascending order. Volumes were sampled at a 30 degree angle from the AC-PC line. Each functional scan included 145 whole brain volumes. Prior to the functional scans, a whole-brain T1-weighted anatomical scan was acquired from each participant (TR=8.4s TE=3.8s, flip angle=8).

Procedure

Upon arrival participants read a participant information sheet and signed a consent form. Participants were then provided with comprehensive instructions outlining the task-procedure that was to be performed, and filled out the following questionnaires (Table 1): Spielberger state-trait anxiety inventory (Spielberger et al., 1983), BIS and BAS scales (Carver & White, 1994) and the Barratt Impulsivity Scale (Patton et al., 1985).

Prior to engaging in the main task, participants first completed a calibration procedure which was designed to identify the two levels of electrical stimulation to be used during the main task. During this procedure, participants received a succession of 5ms square wave electric stimulations, starting from 0.2V, and incrementing at

levels of 0.2V. Participants rated each stimulation on a scale from 0 – 10 where a score of 0 reflected not being able to feel the stimulation, 4 reflected a stimulation that was on the threshold of being painful, 7 related to a stimulation that was deemed ‘painful but still tolerable’ and 10 related to ‘unbearable pain’. The procedure was terminated once the participant reported the level of pain as being equivalent to ‘7’ on the scale. The procedure was performed twice, both times with a one-step-up method, to allow for initial habituation/sensitisation to the stimulation. The voltage levels rated as ‘4’ and ‘7’ on the second scaling were used for the ‘low’ and ‘high’ stimulation levels during the main task. The calibration procedure ensured that the pain levels were psychologically equivalent across participants.

Participants first completed 1 block of 60 trails outside the scanner to ensure that the understood the task. Inside the scanner participants completed a further 4 blocks of 60 trials. The trial structure is depicted in Figure 1. Each trial began with a 0.5s fixation cross before a ‘chance’ cue appeared in the form of a two-segment pie chart. Each chance cue signalled two attributes of the outcome of the trial. The magnitude of the pain that could be administered in the trial was signalled through the colour of the non-grey area, with one colour signalling high pain and the other low pain (colours used were blue and yellow and their assignment to high and low were counterbalanced across participants). The probability of receiving that pain level was signalled through the portion of the coloured/grey areas. A chart where 66% was coloured signalled a 66% chance of receiving the pain stimulation in the trial (high pain probability); whereas a chart where 33% was coloured signalled a 33% chance of receiving the pain stimulation (low pain probability). After the cue, an interval of between 1 and 5s occurred before the delivery (or not) of the 5ms stimulation, which

was signalled visually by the simultaneous appearance of an ‘outcome’ cue, taking the form of a one-segment pie chart, coloured according to the stimulation that was to be delivered (grey for no stimulation, coloured as per the chance cue for the low/high pain stimulations). A jittered inter-trial interval of between 2.5s and 7.5s then occurred before the start of the next trial. Pain outcomes were assumed to be expected when 66% of the chance cue was coloured, and unexpected when 33% was coloured. Each block contained 15 trials of each of the 4 cues, with the cues being accurate, both across the entire experiment and within each block in terms of the probability of the stimulation being delivered (e.g. the 15 presentations of each 33% cue would results in 5 stimulations and 10 no stimulations within each block). Note that the pain magnitude, when delivered, was always that signalled by the cue (i.e. the high pain magnitude never followed a chance cue predicting low pain, and vice-versa), thus a crucial feature of the task was that participants never needed to consider a possibility that they may get either a high or a low pain. While the outcome was probabilistic, the chance cue clarified that only an outcome of a single magnitude (high in half the trials, low in the other half) or no stimulation at all, was possible. By setting participants’ expectations in this way, the chance cue allowed the pain PE to be adaptively coded.

We did not request trial-by-trial responses from the participants in order to ensure that the neural responses to outcomes were not contaminated by cognitive processes associated with response to questions. At the end of each block, participants were shown each chance cue again (order randomised) and asked whether they thought the cue accurately predicted the actual pain delivered. These questions were included to ensure participants paid attention to the contingencies

presented by the cues; the participants were not informed that the cues were in fact accurate.

Analysis

Pre-processing. MATLAB and SPM12 were used to pre-process the raw scans and complete analysis of the resulting data in the 1st and 2nd level GLM (general linear model). Raw PARREC files of the structural and functional MRI data were converted into file formats suitable for SPM12 (.img/.hdr & .nii files) in order to be pre-processed using MATLAB code. A standard neuroimaging pipeline using a mass-univariate approach was subsequently implemented on these data-files/scans. Realignment/motion-correction was applied first, followed by slice timing correction using the central slice for reference. Spatial normalisation was then applied, with the functional data being normalised into a standard stereotactic MNI (Montreal Neurological Institute) space resampled to 3 × 3 × 4mm³ voxels before being spatially smoothed using a 8-mm [8 8 8] full-width at half-maximum (FWHM) Gaussian kernel in order to optimise sensitivity (Ashburner, 2017). Following pre-processing the resulting data-files were specified in the 1st level GLM. A 128s high-pass filter was used to reduce the effects of MRI scanner drift. Inclusion of temporal derivatives in subject's 1st-level analysis were considered, however, this was ultimately deemed counter-productive due to evidence (Sladky, 2011; Della-Maggiore et al., 2002) that including temporal derivatives in paradigms centred on response latencies > 1s can diminish power, thereby directly subverting the increase in sensitivity provided by slice-timing correction. The literature dedicated to onset latency and delays of the hemodynamic response function (hrf) in response to pain

stimuli are quite limited (Pomares et al., 2013; Cauda et al., 2014) with only Cauda et al. reporting that the canonical hrf may not be well-suited to capturing the BOLD response of pain measured mechanical pain. As this paradigm used transcutaneous electric nerve stimulation, not mechanical pain, no derivatives were included in the GLM, and the canonical hrf was used, as in the majority of published research on pain.

Individual models. Individual GLMs were constructed for each participant, encompassing all completed runs. Each run was modelled with 9 regressors. 4 regressors corresponded to chance cues, crossing the factors pain magnitude (high vs. low) and pain probability (high vs. low). 4 regressors corresponded to pain outcomes, crossing the same factors. Note that when the chance cue indicated that pain was likely, the subsequent delivery of pain was taken to correspond to expected pain, while when the chance cue indicated that pain was unlikely, the subsequent delivery of pain was taken to correspond to unexpected pain. The 9th regressor modelled all non-pain outcome cues, following from previous work where non-pain outcomes did not trigger a pain PE (Geuter et al., 2017). Six motion parameters were also included for each run.

Group analysis. Two second-level ANOVA tests were constructed. The *'pain anticipation'* model analysed the response to chance cues and the *'pain response'* model analysed the response to pain outcomes, each crossing the factors pain magnitude x pain probability. We examined main effects using *t*-contrasts and interactions using f-contrasts, with $p<.05$ for initial voxel selection, and *FWE* <.05 to define statistically-significant voxels. The search volume for all group analyses was constrained to Regions Of Interest (ROIs) that were sensitive to pain. These ROIs were defined functionally through a one-sample *t*-test, which contrasted the

response to pain outcomes (an average across all four pain outcome regressors) to responses to non-pain outcomes. A conservative threshold of *FWE*<.05 was used to define the functional ROI mask. The mask was then used to constrain the search volume in all reported group analyses, unless mentioned otherwise.

Results

The response to the monitoring question at the end of the block showed that participants were more likely to say they thought the cue accurately predicted the probability of pain. The question was binary (participants could say yes or no) and 64% of the responses to the monitoring question were ‘. This was true for each individual cue, as well as for all cues together. The relatively low agreement that cues were accurate could reflect a form of probability matching (Vulkan, 2000).

As described in the Methods, we first identified regions that were sensitive to pain outcomes compared to non-pain outcomes. This contrast identified regions that adhere to axiom 1 in the anterior and posterior left insula, anterior right insula, middle and anterior cingulate cortex, as well as activations in the supramarginal gyrus, angular gyrus, supplementary motor cortex, and inferior frontal gyrus.

Next, as a manipulation check, we used the *‘pain response’* model to examine the main effects of pain magnitude on the response to pain outcomes. The search volume was restricted to regions that were sensitive to pain, using a functional ROI mask (see Methods). The analysis identified peaks in the posterior insula bilaterally (Fig3) corresponding to greater activation to high vs low pain. No areas showed greater activation for the delivery of the low pain stimulus.

FIGURE 3

The key analysis proceeded as follows. First, we examined the main effect of pain probability on pain responses, using the *'pain response'* model, and only in the functional ROI mask of pain-sensitive regions (see Methods). The analysis identified a single significant peak, in the left anterior insula (Fig4 B), where response to unexpected pain was greater than response to expected pain, suggesting that this region adheres to both axioms 1 and 2 of the axiomatic model of PE. No area responded more to expected pain than unexpected pain. Second, the interaction between probability and magnitude at outcome was assessed, to identify whether the main effect of pain probability was qualified by pain magnitude. No areas exhibited a statistically significant interaction effect at outcome between pain magnitude and probability. Figure 4D shows the simple effect parameter estimates for this interaction from the peak insula voxel, illustrating that at outcome the effect of probability was both significant (uncorrected) and in the same direction (low probability > high probability) for each pain magnitude. Third, we examined the effects of pain probability and pain magnitude on the response to chance cues (Fig4 C), using the '*pain anticipation'* model, and constraining the search to the functional ROIs that were sensitive to pain (see Method). The only significant response to the main effect of pain magnitude was again observed in the left anterior insula, where activation was stronger for high pain chance cue compared to low pain chance cue (Fig4 A). Neither the main effect of pain probability nor the interaction influenced pain anticipation significantly within the functional ROI mask. Figure 4C shows the

parameter estimates for the main effects for the peak insula voxel at both cue and outcome, illustrating firstly that the pain magnitude contrast was significant only during anticipation, and secondly that the pain magnitude contrast was actually in the opposite direction at outcome, suggesting that the failure to find an effect equivalent to that seen during anticipation was unlikely to be due to a lack of power.

FIGURE 4

For completion, we examined whether pain magnitude or probability in response to chance cues, or the pain outcomes, activated any additional regions outside of the functional ROI mask. For this purpose, we used a conservative *FWE*<.05 for whole-brain search. Anticipating high (vs low) pain activated the right hippocampus and the right inferior frontal gyrus. The whole-brain analysis of pain outcomes did not uncover any additional activations.

Discussion

Signal in the left anterior insula represented adaptive prediction error of pain, defined according to an axiomatic approach. As expected from a region sensitive to pain PE, upon cue presentation, the left anterior insula responded more strongly to the anticipation of high rather than low pain. Additionally, and again as to be expected from a region sensitive to pain PE, upon pain delivery, the signal in this region was stronger for low-probability pain compared to high-probability pain. Crucially, despite its sensitivity to pain magnitude when it was only predicted, and pain probability

when it was actually delivered, this region was insensitive to the magnitude of pain during delivery. The signal level at delivery was similar when either low or high pain were delivered, but significantly greater when the pain was unexpected. This result suggests that the chance cue scaled the subsequent response to pain PE, such that the anterior insula responded similarly to the highest-possible pain in the local context of the trial. Taken together, this pattern corresponds to the pattern reported by Tobler et al. (2005) and which established adaptive scaling of prediction error for reward. Here we show, for the first time, a similar pattern for pain PEs.

While the anterior insula signalled adaptively-coded pain PEs, the posterior insula was sensitive to delivered pain intensity. Together, both results complement elegant results reported by Geuter and colleagues, where the signal in the anterior insula reflected the sum of cued pain expectations and pain PEs, while the posterior insula and parietal operculum coded for pain stimulation intensity (Geuter et al., 2017). PE is a signed quantity; for both reward, and pain, an outcome of greater intensity or magnitude than expected is referred to as positive PE, and an outcome of lesser intensity or magnitude than expected, negative PE. In both our study and that of Geuter and colleagues, the focus was on positively-signed pain PE. Shih et al (2018) compared negative and positive PE for aversive stimuli and presented findings that suggest separate neural substrates for each. They found no areas that coded both negative and positive PE, but the anterior insula exhibited a higher BOLD response for positive PE, as in our study, and the anterior cingulate for negative PE. Interestingly, although they did not observe a regions that expressed both positive and negative computational PE signal, Shih et al. found that the connectivity of both the insula and the ACC with the PAG increased with respect to the PE regardless of

its sign. This result aligns nicely with Roy et al.'s findings that the PAG expressed the (computational) pain PE.

Our study advances understanding of the neural mechanism that serves the adaptive coding of pain, which is less comprehensive than our understanding of adaptive coding of reward (Vlaev et al., 2009). Despite evidence that reward is coded adaptively, and for shared neural mechanism for adaptively coding monetary gain and loss (Nieuwenhuis et al., 2005), it is, in principle, possible that the brain does not need to represent pain in an adaptively-coded manner and can represent all biologically-feasible levels of pain, either because they span a more limited range than all possible rewards, or because the evolutionary significance of pain, compared to reward, has caused the brain to represent pain more accurately. Nevertheless, adaptive coding of pain reflects an influence of pain expectations, and there is ample evidence that pain perception is influenced by expectations (Atlas & Wager, 2012; Tracey, 2010), that expectations are clinically relevant (Buchel et al., 2014), that they may exert a stronger impact on pain perception than the noxious stimulation itself (Lim et al., 2020) and that many of its detailed characteristics have unique influence (Hoskin et al., 2019; Watkinson et al., 2013). There is also elegant behavioural evidence for adaptive coding of pain, reviewed in the introduction (Vlaev et al., 2009; Winston et al., 2014). The first study to examine the neural mechanism of adaptive coding of pain was conducted by Leknes et al. (2013). They presented participants with two cues, one that predicted a 50% probability of either high or moderate thermal pain, and one that predicted a 50% probability of either low or moderate pain. A comparison of the response to the moderate pain stimulation when it followed each cue revealed that activity in the insula and anterior cingulate cortex (ACC) was lower when the stimulation had been preceded by the cue predicting high

pain, suggesting that these regions represented adaptively-coded pain response. However, Leknes et al presented the two cues in separate blocks, contrasting the response to the moderate pain across the blocks. It is not therefore clear to what extent the activation might found might reflect a generalised response to the block context, rather than a prediction error signal specific to the presentation of the moderate stimuli. Using a similar procedure, Winston et al (2014) found that activity in the lateral OFC, but not the Insula or ACC, reflected an adaptively-coded pain response. Although each individual trial in the Winston study involved a single pain level, blocks only involved two levels of stimuli, again potentially introducing 'block effects' to the contrasts. Interestingly, when Bauch and colleagues utilised an event-related variant of Leknes (2013) block design, they were unable to find any area which produce a significant adaptively-coded pain response. Here we used a design that was closer to that employed by Tobler et al. (2004), with a focus on adaptive coding of pain prediction errors, rather than on the scaling of pain value by expectations more broadly, perhaps accounting for why our results differ from those of Winston and colleagues. Importantly, used an event-related methodology, and can therefore be confident that our findings do not reflect generalisation effects.

The human neuroimaging literature reports a number of brain regions that correlate with computational pain PE, , according to reinforcement learning definitions (Sutton & Barto, 2015), including the ventral striatum, anterior insula, and the cingulate cortex (Geuter et al., 2017; Lim et al., 2020; Roy et al., 2014; Seymour et al., 2004; Shih et al., 2019). These studies have not used an axiomatic approach (Caplin & Dean, 2008), so it is not known whether signal there corresponds with all three axioms (Roy et al., 2014), nor whether it expresses quantities that are correlated with

PE, such as the expected value of pain, shown to activate the ventral striatum and involve the anterior insula (Brooks et al., 2010; Palminteri et al., 2012). The left anterior insula has been long thought to be important for the representation of aversive learning. It was found to respond to level of deviation from expectation (Fouragnan et al., 2018), and correlate with PE for emotionally/motivationally neutral perceptual stimuli (Nazimek et al., 2013). Using patients with lesions in the anterior insula, Palminteri and colleagues (Palminteri et al., 2012) were able to establish that it is involved in updating the value of loss-predicting cues during learning, suggesting that it is unlikely to represent a computational PE signal.

Our results complement those mentioned above. They suggest that while the left anterior insula is sensitive to many aspects of aversive learning, and to quantities associated with the computational PE signal for pain, it may better represent an adaptively-coded pain prediction error.

Acknowledgements: We thank Daniel Wilde for help with data analysis, and E. J. Hird for helpful comments.

Author contribution: DT and RH designed the study. RH collected the data. DT and RH wrote the manuscript. All authors discussed the results and commented on the manuscript.

References

Atlas, L. Y., & Wager, T. D. (2012). How expectations shape pain. In *Neuroscience Letters* (Vol. 520, Issue 2, pp. 140–148). https://doi.org/10.1016/j.neulet.2012.03.039

Bauch, E. M., Andreou, C., Rausch, V. H., & Bunzeck, N. (2017). Neural habituation to painful stimuli is modulated by dopamine: Evidence from a pharmacological fMRI study. *Frontiers in Human Neuroscience*, *11*(December), 1–11. https://doi.org/10.3389/fnhum.2017.00630

Brooks, A. M., Pammi, V. S. C., Noussair, C., Capra, C. M., Engelmann, J. B., & Berns, G. S. (2010). From Bad to Worse: Striatal Coding of the Relative Value of Painful Decisions. *Frontiers in Neuroscience*, *0*(OCT), 176. https://doi.org/10.3389/FNINS.2010.00176

Buchel, C., Geuter, S., Sprenger, C., & Eippert, F. (2014). Perspective Placebo Analgesia□: A Predictive Coding Perspective. *Neuron*, *81*(6), 1223–1239. https://doi.org/10.1016/j.neuron.2014.02.042

Caplin, A., & Dean, M. (2008). Axiomatic methods, dopamine and reward prediction error. *Current Opinion in Neurobiology*, *18*(2), 197–202. https://doi.org/10.1016/J.CONB.2008.07.007

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319

Fouragnan, E., Retzler, C., & Philiastides, M. (2018). Separate neural representations of prediction error valence and surprise: Evidence from an fMRI meta-analysis. *Human Brain Mapping*, *39*(7), 2887–2906. https://doi.org/10.1002/HBM.24047

Geuter, S., Boll, S., Eippert, F., & Büchel, C. (2017). Functional dissociation of stimulus intensity encoding and predictive coding of pain in the insula. *ELife*, *6*. https://doi.org/10.7554/ELIFE.24770

Halai, A. D., Welbourne, S. R., Embleton, K., & Parkes, L. M. (2014). A comparison of dual gradient-echo and spin-echo fMRI of the inferior temporal lobe. *Human Brain Mapping*. https://doi.org/10.1002/hbm.22463

Hoskin, R., Berzuini, C., Acosta-Kane, D., El-Deredy, W., Guo, H., & Talmi, D. (2019). Sensitivity to pain expectations: A Bayesian model of individual differences. *Cognition*, *182*, 127–139. https://doi.org/10.1016/J.COGNITION.2018.08.022

Lammel, S., Ion, D., Roeper, J., & Malenka, R. (2011). Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. *Neuron*, *70*(5), 855–862. https://doi.org/10.1016/J.NEURON.2011.03.025

Leknes, S., Berna, C., Lee, M. C., Snyder, G. D., Biele, G., & Tracey, I. (2013). The importance of context: When relative relief renders pain pleasant. *Pain*, *154*(3), 402. https://doi.org/10.1016/J.PAIN.2012.11.018

Lim, M., O'Grady, C., Cane, D., Goyal, A., Lynch, M., Beyea, S., & Hashmi, J. A. (2020). Threat Prediction from Schemas as a Source of Bias in Pain Perception. *The Journal of Neuroscience*, *40*(7), 1538. https://doi.org/10.1523/JNEUROSCI.2104-19.2019

Nazimek, J., Hunter, M., Hoskin, R., Wilkinson, I., & Woodruff, P. (2013). Neural basis of auditory expectation within temporal cortex. *Neuropsychologia*, *51*(11), 2245–2250. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2013.07.019

Nieuwenhuis, S., Heslenfeld, D. J., von Geusau, N. J. A., Mars, R. B., Holroyd, C. B., & Yeung, N. (2005). Activity in human reward-sensitive brain areas is strongly context dependent. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2004.12.043

Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Czernecki, V., Karachi, C., Capelle, L., Durr, A., & Pessiglione, M. (2012). Critical Roles for Anterior Insula and Dorsal Striatum in Punishment-Based Avoidance Learning. *Neuron*, *76*(5), 998–1009. https://doi.org/10.1016/J.NEURON.2012.10.017

Park, S. Q., Kahnt, T., Talmi, D., Rieskamp, J., Dolan, R. J., & Heekeren, H. R. (2012). Adaptive coding of reward prediction errors is gated by striatal coupling. *Proc Natl Acad Sci U S A*, *109*(11), 4285–4289. https://doi.org/10.1073/pnas.1119969109

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1985). *Factor Structure of the Barratt Impulsiveness Scale*. *1 979*, 768–774. https://doi.org/10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G. E., & Wager, T. D. (2014). Representation of aversive prediction errors in the human periaqueductal gray. *Nature Neuroscience*, *17*(11), 1607–1612. https://doi.org/10.1038/nn.3832

Rutledge, R. B., Dean, M., Caplin, A., & Glimcher, P. W. (2010). Testing the Reward Prediction Error Hypothesis with an Axiomatic Model. *The Journal of Neuroscience*, *30*(40), 13525. https://doi.org/10.1523/JNEUROSCI.1747-10.2010

Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. *Nature Reviews Neuroscience*, *17*(3), 183–195. https://doi.org/10.1038/nrn.2015.26

Seymour, B., & McClure, S. M. (2008). Anchors, scales and the relative coding of value in the brain. *Current Opinion in Neurobiology*, *18*(2), 173–178. https://doi.org/10.1016/j.conb.2008.07.010

Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., Friston, K. J., & Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning in humans. *Nature*, *429*(6992), 664–667. https://doi.org/10.1038/nature02581

Shih, Y.-W., Tsai, H.-Y., Lin, F.-S., Lin, Y.-H., Chiang, C.-Y., Lu, Z.-L., & Tseng, M.-T. (2019). Effects of Positive and Negative Expectations on Human Pain Perception Engage Separate But Interrelated and Dependently Regulated Cerebral Mechanisms. *The Journal of Neuroscience*, *39*(7), 1261. https://doi.org/10.1523/JNEUROSCI.2154-18.2018

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Inventory STAI.* Consulting Psychologists Press.

Sutton, R. S., & Barto, A. G. (2015). *Reinforcement Learning: An*

Introduction (2nd ed.). MIT Press.

Talmi, D., Atkinson, R., & El-Deredy, W. (2013). The Feedback-Related Negativity Signals Salience Prediction Errors, Not Reward Prediction Errors. *The Journal of Neuroscience*, *33*(19), 8264–8269. https://doi.org/10.1523/JNEUROSCI.5695-12.2013

Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005). *Huber et al. - 2005 - Science.pdf*. *203*(March), 1642–1645.

Tracey, I. (2010). Getting the pain you expect: Mechanisms of placebo, nocebo and reappraisal effects in humans. In *Nature Medicine* (Vol. 16, Issue 11, pp. 1277–1283). https://doi.org/10.1038/nm.2229

Vlaev, I., Seymour, B., Dolan, R. J., & Chater, N. (2009). The price of pain and the value of suffering. *Psychological Science*, *20*(3), 309–317. https://doi.org/10.1111/j.1467-9280.2009.02304.x

Vulkan, N. (2000). An Economist's Perspective on Probability Matching. *Journal of Economic Surveys*, *14*(1), 101–118. https://doi.org/10.1111/1467-6419.00106

Watkinson, P., Wood, A. M., Lloyd, D. M., & Brown, G. D. A. (2013). Pain ratings reflect cognitive context: A range frequency model of pain perception. *Pain*, *5*(154), 743–749. https://doi.org/10.1016/J.PAIN.2013.01.016

Winston, J. S., Vlaev, I., Seymour, B., Chater, N., & Dolan, R. J. (2014). *Behavioral/Cognitive Relative Valuation of Pain in Human Orbitofrontal Cortex*. https://doi.org/10.1523/JNEUROSCI.1706-14.2014

Legend

Figure 1. Theoretical signals for absolute and adaptively-coded pain PE. According to reinforcement learning models of PE (Sutton & Barto, 2015)*, upon pain delivery, pain PE=R_t-V_{t-1}, with R corresponding to delivered pain magnitude and V corresponding to expected pain in the previous time step. V is computed on presentation of the chance cue such that V_{t-1}=p(R_{t-1}) where R corresponds to the magnitude of an outcome and p is the probability of outcome delivery. When an outcome is predicted 50% of the time, PEs will have a higher value when pain is high than when it is low, all else held equal. This can be seen in the dark line that plots the computed pain PE according to the equations above as a function of pain magnitude. If cues scale expectations such that either high or low pain are recoded as the top of the pain magnitude range, a signal that expresses adaptively-coded PE will not be sensitive to pain magnitude, plotted in the grey line.*

Figure 2. Graphical display of a single trial. The non-grey portion of the chance cue signalled which pain magnitude could be delivered in this trial and with what probability (33% or 66%). Two non-grey colours (blue and yellow) were used to signal the high and low pain magnitudes with which colour reflected which magnitude being counterbalanced between participants.

Figure 3. Response to pain magnitude, high > low pain.

Figure 4. Effects of pain probability on pain responses, and pain magnitude on pain anticipation, in the left anterior insula. the left anterior insula exhibited a pattern consistent with adaptive coding of pain PEs. It was responsive to the magnitude of pain when no expectations were set, at the time the chance cue was presented (high pain cue > low pain cue), but not to the magnitude of experienced pain at outcome

once expectations were set through the cue (null main effect of pain level), while during pain delivery, being responsive to pain expectancy (unexpected pain > expected pain). *The bar graphs depict responses in a single voxel – the peak voxel of the response to unexpected pain delivery (identified through the main effect of probability on pain responses). ** FWE<.05 *Uncorrected p value <.05*

Figures

Figure 1. Theoretical signals for absolute and adaptively-coded pain PE. ***According to reinforcement learning models of PE*** (Sutton & Barto, 2015)***, upon pain delivery, pain PE=R_t-V_{t-1}, with R corresponding to delivered pain magnitude and V corresponding to expected pain in the previous time step. V is computed on presentation of the chance cue such that V_{t-1}=p(R_{t-1}) where R corresponds to the magnitude of an outcome and p is the probability of outcome delivery. When an outcome is predicted 50% of the time, PEs will have a higher value when pain is high than when it is low, all else held equal. This can be seen in the dark line that plots the computed pain PE according to the equations above as a function of pain magnitude. If cues scale expectations such that either high or low pain are recoded as the top of the pain magnitude range, a signal that expresses adaptively-coded PE will not be sensitive to pain magnitude, plotted in the grey line.***

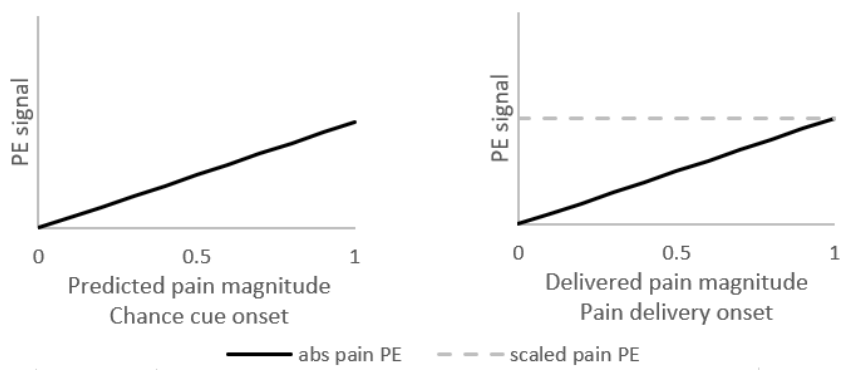


Figure 2. Graphical display of a single trial.

The non-grey portion of the chance cue signalled which pain magnitude could be delivered in this trial and with what probability (33% or 66%). Two non-grey colours (blue and yellow) were used to signal the high and low pain magnitudes with which colour reflected which magnitude being counterbalanced between participants.

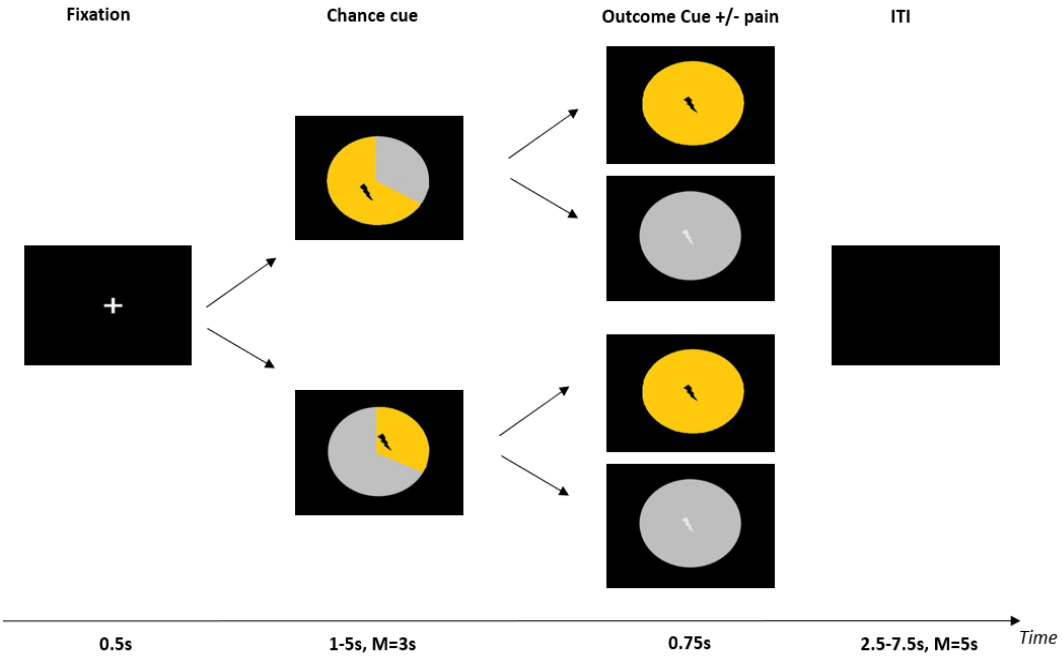


Figure 3. Response to pain magnitude, high > low pain.

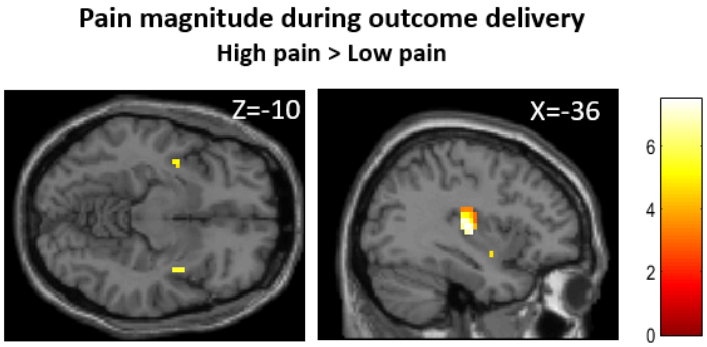
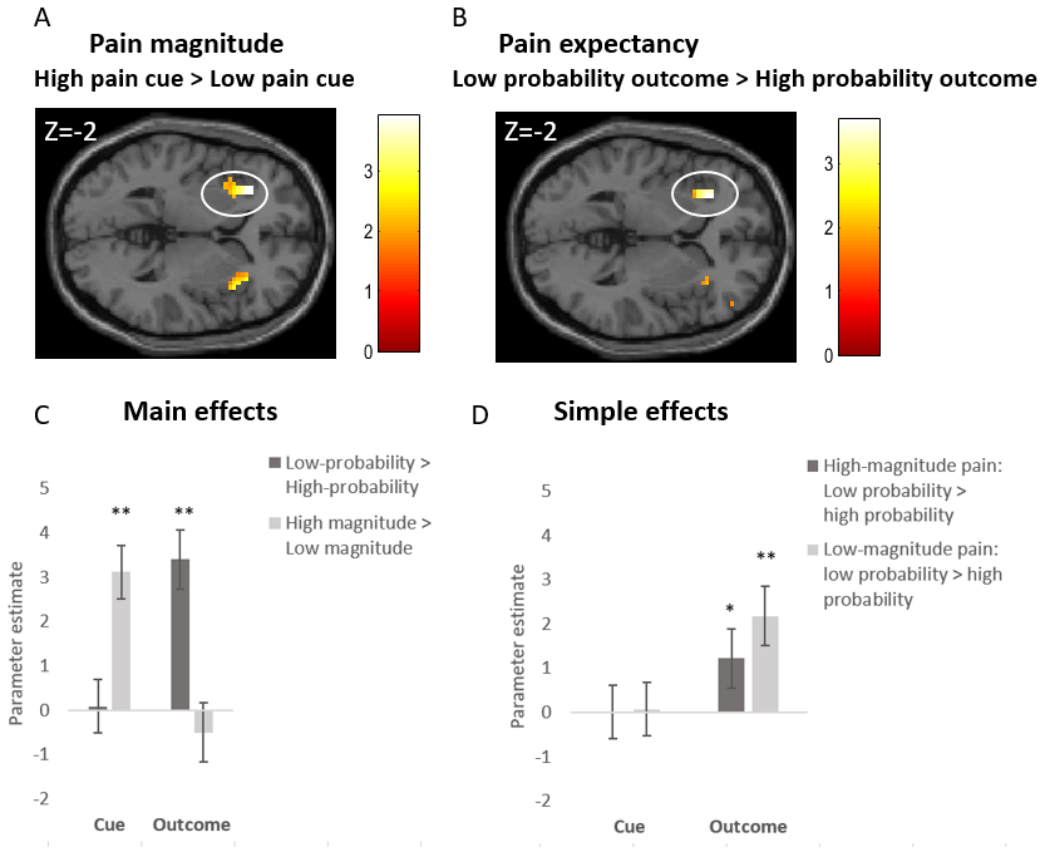


Figure 4. Effects of pain probability and magnitude on pain anticipation and pain response in the left anterior insula.

A. Pain anticipation. Signal in the left anterior insula was higher when the pain cue signalled high pain, compared to when it signalled low pain.

B. Pain response. Signal in left anterior insula was higher when delivered pain was unexpected (33% likely), compared to when it was expected (66% likely).

C. Parameter estimates of the main effects of pain probability (dark bars) and pain magnitude (light bars) in response to the cue (left) and pain delivery (right) in the peak left insula voxel. This region was responsive to the magnitude of pain when no expectations were set, at the time the chance cue was presented (high pain cue > low pain cue), but not to the magnitude of experienced pain at outcome once expectations were set through the cue (null main effect of pain level), while during pain delivery, it was responsive to

pain expectancy (unexpected pain > expected pain).*** FWE<.05. Error bars represent the standard error.*

D. *Simple effects of pain probability* in the peak left insula voxel *when the pain magnitude was high (dark bars) or low (light bars). ** FWE<.05 *Uncorrected p value <.05. Error bars represent the standard error.*