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: Although there is behavioural evidence that pain is coded adaptively, the neural mechanisms that serve this process are not well understood. This study used functional MRI to provide the first evidence that the left anterior insula, an area associated with aversive learning, responds to pain in a manner consistent with the adaptive coding of pain prediction error. This study aids our understanding of the neural basis of subjective pain representation, and thus can contribute to the advancement of analgesic treatments.

Abstract

Background. An understanding of the mechanisms behind the influence of expectation and context on pain perception is crucial for improving analgesic treatments. Prediction error (PE) signalling is thought to contribute to the influence of expectation on pain perception. It is also thought that the brain engages in ‘adaptive coding’ of pain PE, such that sensitivity to unexpected outcomes is modulated by contextual information. While there is behavioural evidence that pain is coded adaptively, controversy remains regarding the underlying neural mechanism, despite evidence that reward PE signals are coded adaptively.

Methods. A cued-pain task was performed by 19 healthy adults while undergoing FMRI scanning. BOLD responses to the task were tested using an axiomatic approach to identify areas that may code pain PE adaptively.

Results. The left anterior insula demonstrated a pattern of response consistent with adaptively-coded pain PE. Signals from this area were sensitive to both predicted pain magnitudes on the instigation of expectation, and the unexpectedness of pain delivery. Crucially however, the response at pain delivery was consistent with the local context of the pain stimulation, rather than the absolute magnitude of delivered pain, a pattern suggestive of an adaptively coded PE signal.

Conclusions. The study advances our understanding of the neural basis of pain prediction. Alongside existing evidence that the periaqueductal grey codes pain PE and the posterior insula codes pain magnitude, the results highlight a distinct contribution of the left anterior insula in the processing of pain.

Introduction

A major function of the nervous system is the ability to learn the relationship between stimuli, and to identify, and respond accordingly, when expected relationships are confounded. According to the predictive coding framework, prediction error (PE) is a key part of this adaptive process. PEs reflect the difference between the expected and delivered stimuli. A neural signal that corresponds to PE is generated when an expected stimulus is not delivered, and functions as a form of feedback, updating the cognitive schema that governs subsequent predictions about the world.

The PE is typically computed according to reinforcement learning models (Sutton & Barto, 1998). A signal coding this computational PE should adhere to the following three axioms (Caplin & Dean, 2008; Rutledge et al., 2010):

Axiom 1: The signal should differentiate better and worse outcomes, such that one is higher and one is lower than the signal to fully-expected outcomes.

Axiom 2: The signal should be proportional to outcome likelihood.

Axiom 3: The signal should not differentiate between fully-expected outcomes.

Another, related function 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 is thought to engage in ‘adaptive coding’, where the sensitivity of some neurons to value is modulated by contextual information (Seymour & McClure, 2008). A neural signal that satisfies the axioms listed above as regards the magnitude of stimulation, independent of context, might be referred to as an ‘absolute’ PE signal, because it is sensitive to the absolute magnitude of the delivered stimuli. In contrast, an ‘adaptively-coded’ PE signal would produce a signal that is scaled to the range of magnitudes that could possibly be experienced in the local context, independent of the absolute magnitude of the

delivered stimulus. Figure 1 illustrates how the adaptively-coded PE diverges from the absolute computational PE.

For rewarding stimuli, adaptive coding is implemented through dopaminergic neuronal firing (Tobler et al., 2005). Tobler and colleagues demonstrated adaptively coded PE in Macaques. In each trial of their experiment the Macaques observed one of three cues that signalled a 50% chance of receiving reward, with each cue signalling one reward magnitude only (low, medium, or large). Cue onsets triggered a midbrain dopaminergic signal proportional to the magnitude of the reward it predicted, namely, it scaled to the size of the potential reward. On trials where the reward was delivered, higher-than-baseline response in the same neurons reflected reward PE. Crucially for an adaptive signal, this response to the reward outcome no longer distinguished reward magnitudes, remaining the same regardless of actual reward delivered. This pattern of responses reflects an adaptively coded PE, as in each individual trial the delivery of the reward was the better outcome (compared to non-reward), even in trials where the outcome itself was lower than those available in other trials.

FIGURE 1

In humans, evidence suggests that adaptively-coded reward prediction error is expressed in the ventral striatum (Park et al., 2012). As with reward, PEs triggered by the prediction or experience of aversive stimuli, such as pain, are also coded by dopamine neurons (Lammel et al., 2011; Schultz, 2016). Computational pain PE 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 does the signal adhere to all three axioms (Roy et al., 2014). As regards adaptive pain responses, there is ample evidence that 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 however, 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. The understanding of the neural mechanisms of adaptively coding pain PE is important as it can help us harness the cognitive system to decrease pain, as in cognitive therapy for chronic pain.

This study examines whether neural evidence for adaptive coding of pain PEs can be acquired by closely following Tobler et al.'s design. To this end a paradigm was implemented where four cues predicted either high or low pain intensity with either high or low probability. Crucially, as with the Tobler study, cues were presented at trial onsets, and each cue only predicted the delivery (or omission) of one intensity of stimulation. Thus, in the local context of each trial, the delivery of the pain stimulation at outcome was the worst outcome, regardless of the intensity of the actual stimulation. It was hypothesized, based on Tobler et al., that a brain area that records PE would, at the point during which expectations are set, produce a

response that scales with the intensity of the predicted stimulus,. Thus in the current experiment the activation of an area demonstrating an adaptively-coded pain PE would meet the following four requirements:

1. At outcome, respond more to delivered stimulations versus omitted stimulations, – following Axiom 1 of the computational PE.
2. At outcome, respond more to unexpected (versus expected) stimulations – following Axiom 2 of the computational PE
3. At outcome, the greater response to the unexpected stimulation (requirement 2) does not vary with the actual intensity of the delivered stimulus – a hallmark of adaptive (rather than absolute) PE signal (Figure 1).
4. At cue onset, respond more to the high pain cue than the low pain cue.

Note that with the experimental design utilised, it was not possible to test Axiom 3 of the computational PE. As pain delivery was necessarily probabilistic, no stimulation was ever fully expected.

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. The task resembled one 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 (Figure 2). 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.

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 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 that was to be performed. Participants also 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) as part of a wider study investigating the relationship between anxiety, impulsivity and the neural response to pain.

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’ pain 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 2. 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 indicator’, taking the form of a one-segment pie chart, coloured according to the stimulation that was to be delivered (grey for no stimulation, and coloured as per the chance cue for the low/high pain stimulations). These were included in all trials firstly to equalise the visual stimulation across conditions, and secondly to provide certainty as to the outcome for trials where no stimulation was delivered, thus avoiding introducing

confounds around temporal uncertainty as to the delivery of pain stimulations into trials where they were not delivered. 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 high or 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. Prior to starting the task, participants were informed of the nature of the chance cues and outcome indicators, for example that grey indicated no stimulation, and which colour represented each pain level. Participants were not however informed that the probabilities represented on the

chance cues were in fact accurate, thus retaining the need to pay attention to the cue-outcome relationships during the task.

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 (i.e. the delivery of a stimulation), 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 outcomes (i.e. the presentation of the outcome indicator that signalled that no pain would be delivered), 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 (the 9th regressor). 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 "yes". This level of agreement was true for each individual cue, as well as for all cues together.

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 of the computational PE (and requirement 1 of the adaptive PE) 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. For completion, we also examined the opposite contrast (non-pain outcomes > pain outcomes) a contrast which could involve psychological relief. This analysis did not identify any statistically significant voxels.

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 (Figure 3) 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 (Table 2 and Figure 4B), 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 computational model of PE and requirements 1 and 2 of the adaptive PE, . No area responded more to expected pain than unexpected pain. Figure 4-C depicts the parameter estimate for the two main effects from the peak insula voxel. Figure 4-D shows the parameter estimates for the effect of probability from the peak insula voxel for each pain magnitude, illustrating that at outcome the effect of probability was in the same direction (low probability > high probability) for each pain magnitude, and that the 90% confidence interval excluded zero for both. We verified this result by constructing two new flexible factorial models, one for each pain magnitude (high or low), and each with two vectors, one for high and one for low probability. Analysis was limited to the significant left insula cluster. Paired t-tests suggested that the probability effect was significant there for high pain (peak X=-30, Y=23, Z=2) at an uncorrected p<.05 (FWE<.1), and for low pain (peak X=-30, Y=29, Z=6) at FWE<.05.

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, thus suggesting that the left anterior insula also satisfies requirement 3 of the adaptive PE. Consider that a signal that corresponds to the absolute PE pain signal should take the form of stronger response to unexpected pain when it is high, compared to when it is low (Figure 1). Figure 4-D shows that in the left anterior insula peak, the pain magnitude contrast at outcome was (numerically) in the opposite direction. This suggests that the failure to find an interaction between the effects of pain magnitude and probability at outcome delivery was unlikely to be due to a lack of power.

Third, we examined the effects of pain probability and pain magnitude on the response to chance cues (Table 2), 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 chance cues that signalled high pain, compared to those that signalled low pain (Figure 4A and 4-C). Thus this area also satisfied requirement 4 of the adaptive PE. Neither the main effect of pain probability nor the interaction influenced pain anticipation significantly within the functional ROI mask.

Figure 4C summarises the pattern of main effects at anticipation and outcome, showing their parameter estimates for the peak insula voxel at both cue and outcome onsets. It illustrates, firstly, that the pain magnitude contrast was significant

only during anticipation, and secondly, that the pain expectancy contrast was only significant during pain delivery.

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 adaptively-coded prediction error (PE) of pain, defined according to an axiomatic approach such that it should satisfy the 4 requirements defined in the Introduction. 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 (i.e. unexpected) pain compared to high-probability pain. Crucially, despite its sensitivity to pain magnitude when it was predicted, and pain probability when it was actually delivered, this region was insensitive to the magnitude of pain during delivery (requirement 4). The pain response signal 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) 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, better-than-expected outcomes are referred to as causing positive PE, and lesser-than-expected outcomes, 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 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 absolute pain PE, which corresponds to computational reinforcement learning models, where it is insensitive to local context.

This 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 instead 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 on the experience of pain (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 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, individual blocks only involved two levels of stimuli (to create the local context for the PE), 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 produced a significant adaptively-coded pain response. Here we used a design that was closer to that employed by Tobler et al. (2005), with a focus on adaptive coding of pain prediction errors, rather than on the scaling of pain value by expectations more broadly. This perhaps accounts for why our results differ from those of Winston and colleagues. Importantly an event-related methodology was used in the current study, meaning that confidence can be taken that the results do not reflect generalisation effects. The human neuroimaging literature reports a number of brain regions that correlate with ‘absolute’ pain PE, computed 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 in these regions 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). As in the current study, Geuter et al found activation consistent with Pain PE signals in the anterior insula, using heat rather than electrical pain stimulations. In that study the PE could not be adaptively coded because participants were informed that they could receive the least favourable, highest stimulation intensity in each trial, with the result that in the local context of each trial was always the same. This was avoided in the current study because while the cue signalled that the high stimulation intensity was the least favoured option in half the trials, in the other half, it signalled that the low pain stimulation was the least favoured option. The cue therefore created a local context, which ensured the ensuing PE signal could be adaptive scaled. Our results suggest that the PE found in the anterior insula by Geuter et al, may be subject to adaptive scaling.

The left anterior insula has been long thought to be important for the representation of learning. It has been found to express levels of deviation from expectation (Fouragnan et al., 2018), and correlate with PE for emotionally/motivationally neutral perceptual stimuli (Nazimek et al., 2013). In terms of aversive learning, Palminteri and colleagues (Palminteri et al., 2012) were able to establish, using patients with lesions in the anterior insula, that the area is involved in updating the value of loss-predicting cues. The anterior insula is also associated with signalling the behavioural relevance of information, for example providing early responses to facial stimuli according to the relevance of the emotion expressed upon it (Frot et al., 2022). Since unexpected stimuli are likely to be more behaviourally relevant than expected ones, there is a natural overlap between signalling PE and behavioural relevance. Nevertheless, in the current study the response of the anterior insula was no different when unexpected pain was of low compared to high

magnitude (requirement 3). Because higher pain is more behaviourally relevant, this suggests that the results of the current study cannot solely be attributed to the insula's function in signalling behavioural relevance.

One potential shortcoming of the experimental design is that it did not include a behavioural task that would allow responses to be recorded on each trial. This prevented assessing a behavioural impact of PE. As noted in the procedure section this decision was made to ensure that the neural responses to outcomes were not contaminated by cognitive processes associated with response selection. The prediction error signal is a computational quantity and its relevance to brain function is evidenced in a large number of electrophysiological and imaging studies (Rutledge et al., 2010; Schultz, 2016). Although PEs are essential for learning, whether they have any immediate behavioural impact is unclear. Indeed recent research has demonstrated that it is pain expectations, rather than PE, that influence pain perception during probabilistic processing (Nickel et al., 2022).

At the end of each block participants were required to signal whether they thought the probabilities reflected on the chance cues corresponded accurately to the frequency of pain that was actually delivered. This task was included to ensure participants paid attention to the cue>outcome relationships, and thus engaged with the probabilistic nature of the paradigm. Only a modest majority of these reports confirmed belief that the cues were accurate. Participants may have responded in this way because they believed that the frequency of outcome did not conform exactly with the probabilities shown, even though in reality, it did. Future studies would benefit from using a more focussed task, for example asking participants to estimate the frequency with which each cue was followed by a stimulation. It would

be interesting to see whether participants' responses reveal systematic bias, for example a belief that the frequency of unexpected pain was higher than 25%.

A shortcoming of the experimental design was the use of the outcome period for the no-pain conditions as a baseline for the identification of the pain-sensitive regions. Although this contrast satisfies the first axiom of PE, it is potentially contaminated with relief that might have been experienced on the trials where no stimulation was delivered. However, the opposite contrast (no pain >pain) did not reveal any significant results, suggesting that relief was not strongly experienced in the regions of interest. Furthermore, the other contrasts performed in the analyses which were not associated with relief, such as the main effects of pain probability or magnitude. Therefore, we do not believe this shortcoming impacts the results presented. The region identified here as important for adaptive pain PE, the anterior insula, is widely considered to be involved in pain processing (Garcia-Larrea & Peyron, 2013).

In summary the results of the current research 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

Frot, M., Mauguière, F., & Garcia-Larrea, L. (2022). Insular Dichotomy in the Implicit Detection of Emotions in Human Faces. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhab477

Garcia-Larrea, L., & Peyron, R. (2013). Pain matrices and neuropathic pain matrices: A review. *Pain*, *154*(SUPPL. 1). https://doi.org/10.1016/j.pain.2013.09.001

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

Nickel, M. M., Tiemann, L., Hohn, V. D., May, E. S., Avila, C. G., Eippert, F., & Ploner, M. (2022). Temporal–spectral signaling of sensory information and expectations in the cerebral processing of pain. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(1). https://doi.org/10.1073/pnas.2116616119

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. (1998). Reinforcement Learning: An Introduction. *Advances in Cancer Research*, *104*, 322. https://doi.org/10.1016/S0065-230X(09)04001-9

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). Adaptive Coding of Reward Value by Dopamine Neurons. *Science*, *307(5715)*(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

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*

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.***

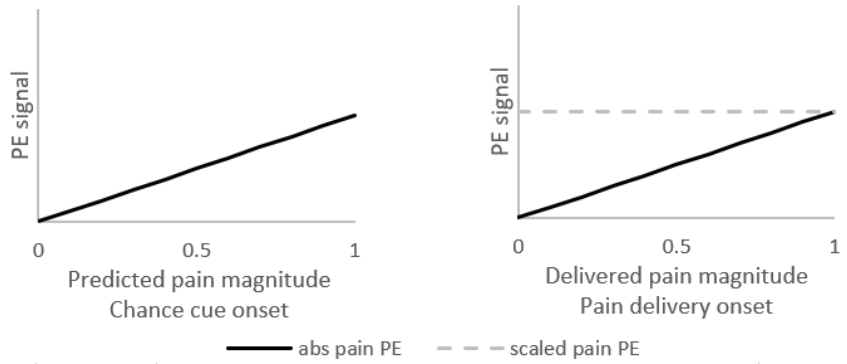


Figure 2. Graphical display of a single trial.

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.

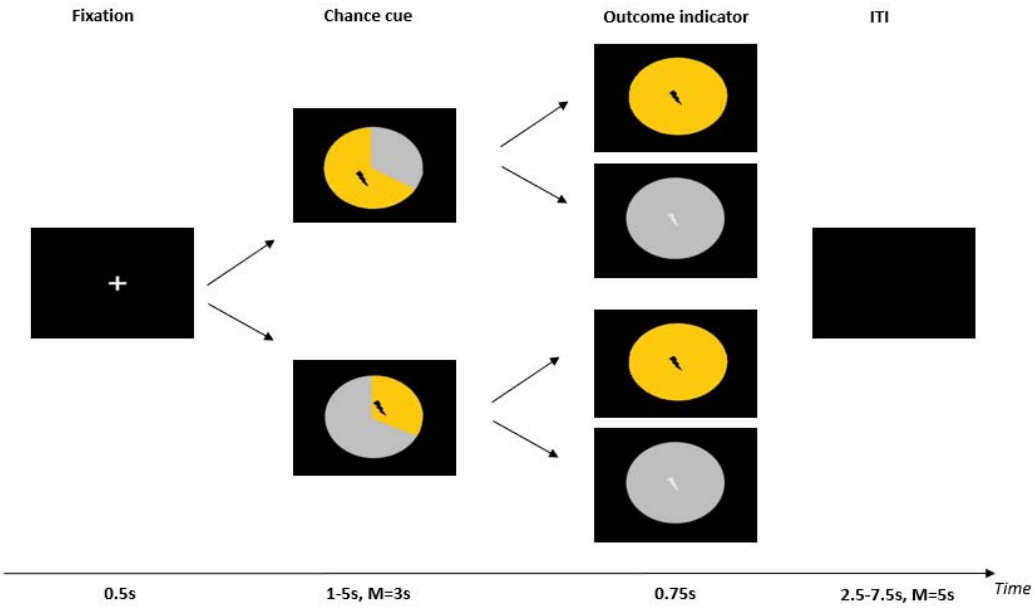


Figure 3. Response to pain magnitude, high > low pain.

The axial (Z=-10) and sagittal (X=-10) slices show signal in the left anterior insula, which was higher when the pain delivered was high, compared to when it was low, *FWE<.05.*

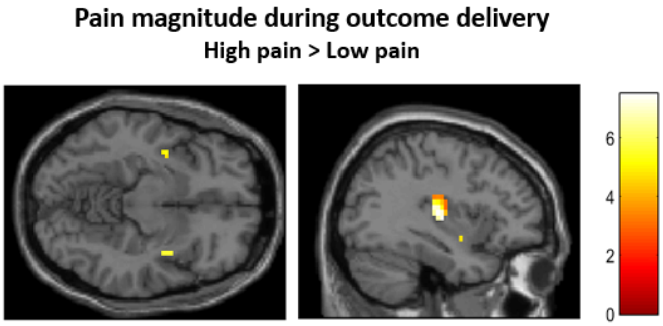


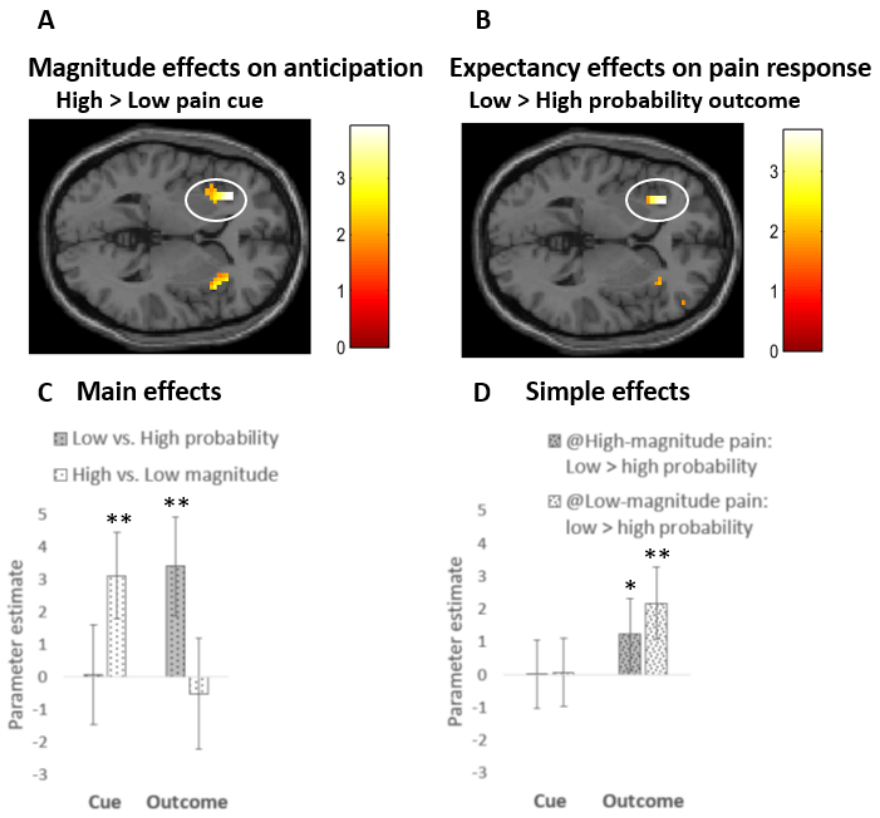
Figure 4. Effects of pain probability and magnitude on pain anticipation and pain response in the left anterior insula.

A. Effect of pain magnitude on pain anticipation. Analysis of the pain anticipation model shows that signal in the left anterior insula (X=-30, Y=29, Z = 2) was higher when the pain cue signalled high pain, compared to when it signalled low pain. *FWE<.05.*

B. Effect of pain expectancy on pain response. Analysis of the pain response model shows that signal in left anterior insula (X=-30, Y=26, Z =2) was higher when delivered pain was unexpected (33% likely), compared to when it was expected (66% likely). *FWE<.05.*

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 left insula (X=-30, Y=26, Z =2), FEW<.05. 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). *Error bars represent the 90% confidence interval.*

D. *Simple effects of pain probability* in the left insula (X=-30, Y=26, Z =2) *when the pain magnitude was high (dark bars) or low (light bars). Error bars represent the 90% confidence interval. * p<.05, ** FEW<.05 in a paired t-test (see text).*


A
Magnitude effects on anticipation
High > Low pain cue
B
Expectancy effects on pain response
Low > High probability outcome
C Main effects
Low vs. High probability
High vs. Low magnitude
Parameter estimate
Cue
Outcome
D Simple effects
@High-magnitude pain:
Low > high probability
@Low-magnitude pain:
low > high probability
Parameter estimate
Cue
Outcome