Behavioral/Cognitive

Pain-Related Expectation and Prediction Error Signals in the Anterior Insula Are Not Related to Aversiveness

Sepideh Fazeli and [ID]**Christian Büchel**

Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany 20246

The anterior insula has repeatedly been linked to the experience of aversive stimuli, such as pain. Previously, we showed that the anterior insula is involved in the integration of pain intensity and its prior expectation. However, it is unclear whether this integration occurs by a pain-specific expectation or a more general expectation of an aversive event. To dissociate these possibilities, we conducted an experiment using painful stimuli and aversive pictures with three levels of aversiveness on human male volunteers. Stimuli were preceded by a probabilistic, combined modality and intensity cue in a full factorial design. Subjective ratings of pain intensity and skin conductance responses were best explained by a combination of actual pain intensity and expected pain intensity. In addition, using fMRI, we investigated the neuronal implementation of the integration of prior expectation and pain intensity. Similar to subjective ratings and autonomic responses, the dorsal anterior insula represented pain intensity and expectations. The ventral anterior insula additionally represented the absolute difference of the two terms (i.e., the prediction error). The posterior insula only represented pain intensity. Importantly, the pattern observed in the anterior insula was only present if the cued modality was correct (i.e., expect pain); in case of an incorrect modality cue (i.e., expect aversive picture), the ventral anterior insula simply represented pain intensity. The stimulus expectation and prediction error specificity in the ventral anterior insula indicates the integration of expectation with painful stimuli in this area. Importantly, this pattern cannot be explained by aversiveness.

Key words: expectation; human fMRI; pain; predictive coding; somatosensory perception

---

Significance Statement

The anterior insula has been shown to integrate pain intensity and their expectation. However, it is unclear whether this integration is pain-specific or related more generally to an aversive event. To address this, we combined painful stimuli and aversive pictures with three levels of aversiveness. The ventral anterior insula represented pain intensity, expectation, and their absolute difference (i.e., the prediction error). Importantly, this pattern was only observed if the cued modality was correct. In case of an incorrect modality cue, this area simply represented as pain intensity. The stimulus expectation and prediction error specificity in the ventral anterior insula indicates the integration of expectation with painful stimuli in this area. Importantly, this pattern cannot be explained by aversiveness.

Introduction

Accurate and fast perception of pain is crucial for an organism to avoid tissue damage. In many cases, sensory cues precede pain

and allow the organism to generate an expectation or prediction. The influence of such expectations on pain perception has been established in basic cued pain paradigms (Koyama et al., 2005; Keltner et al., 2006; Wiech et al., 2008; Atlas et al., 2010) and in the context of placebo and nocebo effects (Tracey, 2010; Häuser et al., 2012). It has previously been speculated that predictive coding, which integrates predictions and violation of predictions (i.e., prediction errors), could be a suitable mechanism mediating the effect of expectations on pain (Büchel et al., 2014; Geuter et al., 2017). In this framework, expectations (i.e., predictions) and the mismatch between these expectations and the sensory inputs (i.e., prediction errors) are assumed to be combined to minimize

future prediction errors by updating the expectations at multiple levels of the neuronal hierarchy. So far, this framework has mainly been applied to visual and auditory processing (Rao and Ballard, 1999; Sterzer et al., 2008; Egner et al., 2010; Okada et al., 2018) and has only recently been extended to pain perception (Geuter et al., 2017). However, pain is not unimodal but comprises other aspects, such as aversiveness, unpleasantness, and salience, which are shared by negative stimuli in other sensory modalities. Therefore, the observed expectation and prediction error signals in the anterior insula and the intensity representation in the posterior insula (Geuter et al., 2017) might not be related to pain per se but could be related to aversiveness in general. This alternative hypothesis can be tested by adding a second aversive stimulus modality. Consequently, we adapted our cued pain paradigm (Geuter et al., 2017) by adding aversive pictures as a second sensory modality. In both modalities, we presented aversive stimuli at three levels of intensity. Each stimulus was preceded by a combined modality and intensity cue. Cues and stimuli were combined in a full factorial design with factors of cue modality, cue intensity, stimulus modality, and stimulus intensity. This design allowed us to investigate whether expecting an aversive stimulus with a specific intensity only affects the neuronal activity in response to a stimulus in the cued modality or whether this influence also generalizes to a stimulus in the other modality.

Within the framework of predictive coding, the prediction error represents the "surprise" induced by a mismatch between the sensory signals encountered and those predicted. If the previously observed prediction error signal in the anterior insula is related to aversiveness in general, we should observe such a signal if the cued intensity differs from the perceived intensity, regardless of the cued modality: a prediction error for a high-intensity painful stimulus, when a low-intensity aversive picture or a low-intensity painful stimulus was expected. On the contrary, if the anterior insula does not represent aversiveness in general, we should only observe prediction errors if the intensity cue is incongruent with the experienced intensity, but only if the modality cue matches: if a high intensity painful stimulus is perceived, when a low-intensity pain stimulus is cued, but not when a low-intensity aversive picture cue has been presented. Therefore, investigating the modality cue specificity of expectations and prediction errors for pain can potentially shed light on the characteristic of the pain processing system and a possible implementation of predictive coding therein.

Materials and Methods

Participants
Thirty-two healthy human male volunteers (average age: 26 years; range: 19–38 years) successfully participated in the study and were paid as compensation for their participation. Exclusion criteria were neurological, psychiatric, dermatological diseases, pain conditions, current medication, or substance abuse. All volunteers gave their informed consent. The study was approved by the Ethics board of the Hamburg Medical Association.

Stimuli and task
The stimuli were either thermal or visual stimuli. Thermal stimulation was performed using a 30×30 mm^2 Peltier thermode (CHEPS Pathway, Medoc) at three different intensities: warm (42°C), medium painful (46°C), and highly painful (48°C) to the left radial forearm at different skin patches in each experimental block to avoid sensitization. The stimulation pattern was chosen based on the distribution of average human pain thresholds to have three distinguishable levels of thermal stimulation. The visual stimuli were pictures chosen from the International Affective Picture System (IAPS) (Lang et al., 2005) database at three levels

of aversiveness. The classification of IAPS pictures was performed manually to divide the reported valance spectrum into six separate groups, where three of them were assigned to the training set and the other three were assigned to the fMRI set. The pictures in the fMRI set had three levels of aversiveness of which the low aversive category had aversiveness values of 2.02 ± 0.05 (mean $\pm$ SE), the medium aversive category had aversiveness values of 4.06 ± 0.02 (mean $\pm$ SE), and the high aversive category had aversiveness values of 5.23 ± 0.01 (mean $\pm$ SE). Before each picture or heat stimulus, a visual cue was presented. The color of the cue (triangle) indicated the modality of the stimulus (yellow for picture and red for heat). The size of the cue indicated its intensity (small, medium, and large triangle for low, medium, and high intensities, respectively).

At the beginning of each trial (Fig. 1), the cue was presented for 500 ms, indicating the modality and intensity of the following stimulus. The cue validity for the stimulus modality was 70%. Cue validity with respect to intensity was 60% (for details of all cue stimulus combinations and frequency, see Fig. 1F). After a jittered blank period of 300–700 ms, volunteers received the visual or heat stimulus with a duration of 2 s. After stimulus presentation, volunteers had to rate the intensity of the aversiveness of the picture or the heat stimulus within a time window of 2 s on a 4 point rating scale (1 labeled as "neutral" and 4 as "strong"). A fixation cross was presented throughout the trial. In addition, we added 4 catch trials to each block, in which volunteers were asked to report the preceding cue within 8 s. No stimulation was given in these trials. Volunteers were rewarded with 50 cents for a correct answer in each catch trial. Finally, a variable intertrial interval (2.5 ± 0.5 s) followed each trial, during which only a fixation cross was displayed.

Procedure
The experiment consisted of two sessions: a behavioral training session and an fMRI session. Before the training session, volunteers were informed about the procedure and gave their written informed consent. They were informed about the experimental paradigm and the nature of the cues and the contingencies between the cues and the stimuli to avoid learning effects during the main fMRI session. The training session contained up to three blocks and was terminated once volunteers had learned the cue regularities. The termination criterion was defined as detecting >75% catch trials and a significant discrimination (repeated-measures ANOVA [rmANOVA]: $p < 0.05$) of the three levels of pain and aversive picture intensities. The fMRI session consisted of four blocks. Each block consisted of 130 trials (including four catch trials) and lasted ~15 min. The trial order within each block was pseudo-randomized, and the order of blocks was randomized across volunteers. During each block, fMRI signal and skin conductance responses (SCRs) were measured. A high-resolution anatomical image for each volunteer was acquired at the end of the fMRI session. The whole fMRI experiment lasted for ~2 h.

Data acquisition
fMRI data were recorded on a Trio 3 tesla system equipped with a 32-channel head coil (Siemens). Forty transversal slices (voxel size $2 \times 2 \times 2$ mm, 1 mm interslice gap) were obtained within each volume using a T2*-sensitive single-shot EPI sequence (TR = 1.24 s, TE = 26 ms, flip angle: 60°, FOV: 220×220 mm) with parallel imaging (GRAPPA, in-plane acceleration factor 2) (Griswold et al., 2002) and simultaneous multislice acquisitions ("multiband," slice acceleration factor 2) (Feinberg et al., 2010; Moeller et al., 2010; Xu et al., 2013) as described previously (Setsompop et al., 2012). The corresponding image reconstruction algorithm was provided by the University of Minnesota Center for Magnetic Resonance Research. Slices were tilted ~30° relative to the AC-PC line to improve coverage in the brainstem. Additionally, T1-weighted structural images ($1 \times 1 \times 1$ mm resolution) were acquired using an MPRAGE sequence (TR = 2300 ms, TE = 2.98 ms, flip-angle = 9°). Stimulus presentation, response recording, and timing relative to MR data acquisition were done using Psychophysics Toolbox 3 (http://www.psychtoolbox.org). Skin conductance was recorded using an EDA100C MRI system (Biopac Systems), amplified using an analog amplifier (MP150, Biopac Systems), and sampled at 100 Hz using a CED 1401 analog-digital converter (Cambridge Electronic Design). SCR electrodes
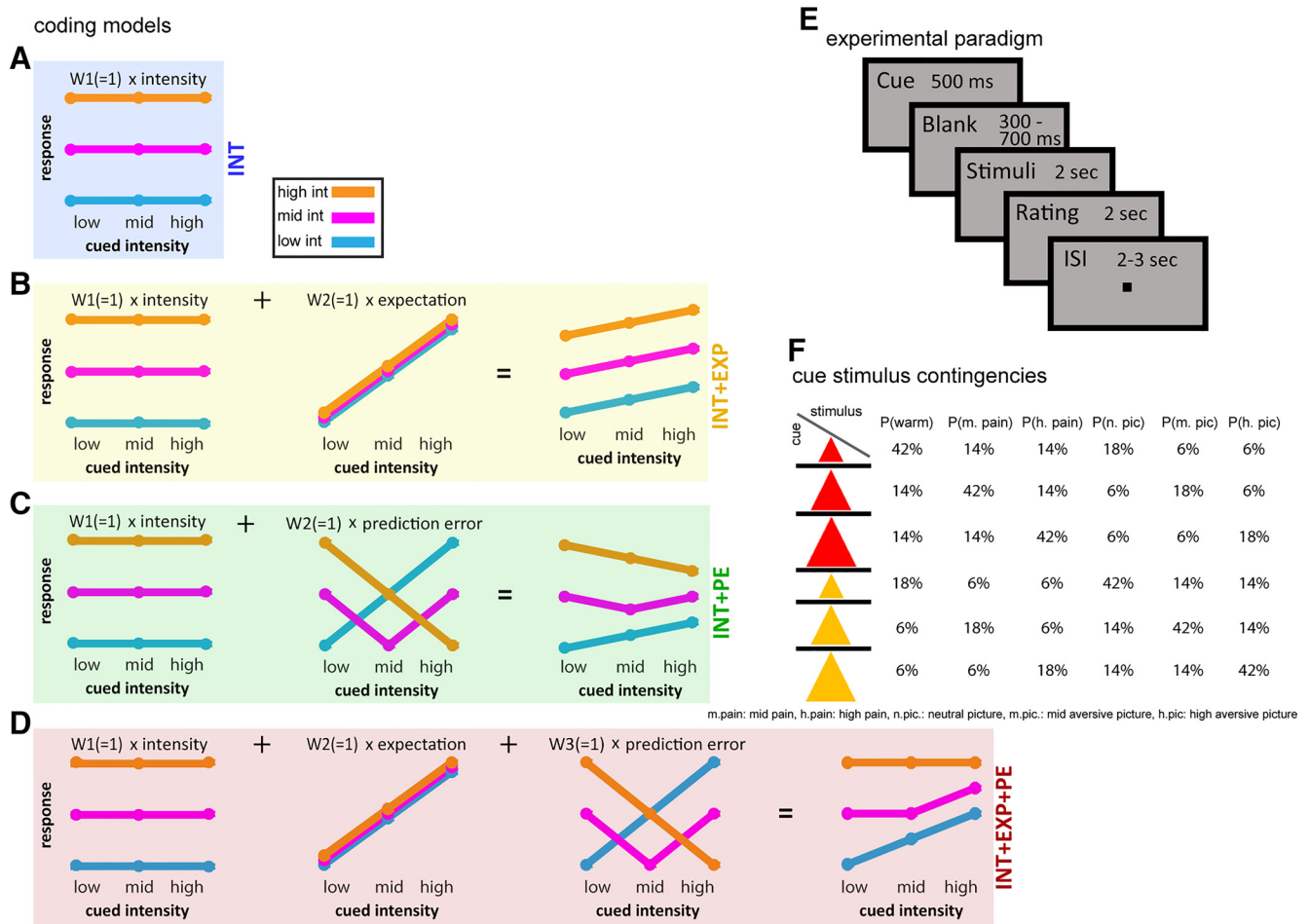
coding models

A W1(=1) x intensity — response / cued intensity (low, mid, high) — INT

high int / mid int / low int

B W1(=1) x intensity + W2(=1) x expectation = INT+EXP

C W1(=1) x intensity + W2(=1) x prediction error = INT+PE

D W1(=1) x intensity + W2(=1) x expectation + W3(=1) x prediction error = INT+EXP+PE

E experimental paradigm

Cue 500 ms
Blank 300 - 700 ms
Stimuli 2 sec
Rating 2 sec
ISI 2-3 sec

F cue stimulus contingencies

stimulus cue	P(warm)	P(m. pain)	P(h. pain)	P(n. pic)	P(m. pic)	P(h. pic)
	42%	14%	14%	18%	6%	6%
	14%	42%	14%	6%	18%	6%
	14%	14%	42%	6%	6%	18%
	18%	6%	6%	42%	14%	14%
	6%	18%	6%	14%	42%	14%
	6%	6%	18%	14%	14%	42%

m.pain: mid pain, h.pain: high pain, n.pic.: neutral picture, m.pic.: mid aversive picture, h.pic: high aversive picture

Figure 1. Experimental design and predicted response patterns for INT, INT+EXP, INT+PE, and INT+EXP+PE models. *A*, In the INT model, the response to the stimulus is only driven by its intensity. *B*, In the INT+EXP model, the stimulus response is the weighted sum of the stimulus intensity (left) and the stimulus expectation (middle). *C*, In the INT+PE model, the stimulus response is the weighted sum of the stimulus intensity (left) and the stimulus prediction error (middle). *D*, For the INT+EXP+PE model, the stimulus response is the weighted sum of the stimulus intensity (left), the stimulus expectation (second from left), and the prediction error (third from left). Right, Examples shown here assume an equal contribution of the expectation and the prediction error signals to the stimulus response. Different contributions of the expectation or prediction error can lead to a different pattern for the stimulus response. *E*, Trial structure: At the beginning of each trial, a central cue indicated the probability of the intensity and the modality of the upcoming stimulus. Then after a blank screen with variable duration, volunteers received the stimulus (2 s), followed by the rating (2 s). The intertrial intervals were randomly jittered between 2 and 3 s. Volunteers were asked to fixate a central-fixation-dot (except for the rating duration). *F*, Table illustrating the cue-stimulus contingencies.

were connected to the palm (thenar and hypothenar eminences) of the left hand. During the study, lights in the MRI room were dimmed, and luminance was kept constant across volunteers.

Data analyses

SCRs. The phasic skin conductance drive locked to the onset of each stimulus was estimated from the raw SCRs in each trial using an established deconvolution method as implemented in ledalab (Benedek and Kaernbach, 2010). The resulting phasic responses were then separately averaged for different conditions within a temporal window from 2 to 5 s after stimulus onset for the painful stimuli and within a window from 0 to 3 s for the aversive pictures, which was selected to cover the peak of the averaged SCR signal in each modality across volunteers.

fMRI data preprocessing. Functional imaging data were analyzed using MATLAB (version 2014b, The MathWorks) and SPM12 (Wellcome Trust Centre for Neuroimaging, London). The first five volumes of each functional block were discarded. The remaining images were spatially realigned to correct for motion artifacts and were then spatially normalized using DARTEL, a high-dimensional warping algorithm implemented in SPM (Ashburner, 2007). The functional images were spatially smoothed using a Gaussian kernel with an FWHM of 6 mm. First-level models included separate regressors using a boxcar with a duration of 2 s for each of the 36 experimental conditions: 6 stimuli (3 levels of pain + 3

levels of aversive images) × 3 intensity cues × 2 modality cues. Additional regressors with a duration of 2 s modeled the rating period. Another regressor (boxcar, duration 9 s) modeled the catch trials. In addition, the model also included six motion parameters estimated during realignment. All regressors were convolved with a canonical hemodynamic response function as implemented in SPM. The first-level contrast estimates for the 18 pain conditions were the basis for our computational models. Model estimation used a linear mixed-effect model as implemented in MATLAB (version 2014b, The MathWorks). Models were estimated for all gray-matter voxels. Gray-matter voxels were determined using the segmentation routine in SMP12 at the single volunteer level, then averaged across volunteers, and finally thresholded at an 80% gray matter probability.

Neurological pain signature (NPS) analyses. The NPS is a multivariate pattern of brain activity with the aim of distinguishing experimental pain from other conditions (Wager et al., 2013; Krishnan et al., 2016). The NPS response mirrors subjective pain and has been suggested as a surrogate marker for heat pain intensity (Wager et al., 2013). We computed NPS expression values for each of the experimental conditions based on the regressors described above separately for each participant (Wager et al., 2011, 2013; Krishnan et al., 2016). The resulting NPS values were then the basis for our model comparison.

ROI masks and analysis. According to the finding reported by Geuter et al. (2017), insular cortex is the area among pain-responsive regions representing expectation and prediction error signals. Based on these findings, we mainly focused our analysis on the insular cortex and its anatomically defined subdivisions as defined by the Destrieux atlas (Fischl et al., 2004; Destrieux et al., 2010). Using insular subdivisions, we tested our four models on the average parameter estimates extracted from voxels within these ROIs (for the detailed model comparison routine, see below). Even though the insular cortex is the focus of this study because of its key role in the integration of pain expectation as the internal model, and sensory signals as the physical input (Ostrowsky et al., 2002; Nieuwenhuys, 2012; Geuter et al., 2017), the processing of nociceptive information is not limited to the insular cortex. We therefore conducted additional ROI-based analyses in several other pain-related regions, including the parietal operculum (SII), postcentral sulcus (SI), frontal medial cortex, anterior cingulate (ACC), posterior cingulate (PCC), amygdala, hippocampus, thalamus, cerebellum, and periaqueductal gray (PAG). Anatomical masks were the same as used by Geuter et la. (2017), except for the PCC, frontal medial cortex, hippocampus, and cerebellum, which were generated by thresholding the probability map of the corresponding areas in the Harvard-Oxford Atlas (Desikan et al., 2006) at 50% probability (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases). Similar to insular subdivisions, we tested our four models on the average parameter estimates from voxels within these masks.

Modeling behavioral, SCR, NPS, and voxelwise fMRI data. Crossing our three levels of stimulus intensity, three levels of intensity cues, and two levels of modality cues resulted in 18 ($3 \times 3 \times 2$) different pain conditions, in which nine stimuli were preceded by a valid modality cue and nine stimuli were preceded by an invalid modality cue. For these valid and invalid cue conditions, subjective ratings, SCR, NPS, and voxelwise fMRI data were fitted separately to four different models. All models included a stimulus intensity term. The first model (intensity model [INT]) only contained the stimulus intensity term as the physical characteristic (intensity) of the stimulus (Fig. 1A). The second model (expectation model [INT+EXP]) additionally comprised an expectation term represented by the intensity cue (Figure 1B). The third model (prediction error model [INT+PE]), comprised two terms: the prediction error (i.e., the absolute difference between intensity cue and stimulus intensity) and stimulus intensity (Fig. 1C). This model was inspired by several studies reporting reduced neuronal activity in response to expected versus unexpected stimuli (Egner et al., 2010; Jiang et al., 2013). The final full model (INT+EXP+PE) contained both the expectation and the prediction error term, as well as the intensity term. Conceptually, this model assumes that the response to a painful stimulus is a weighted sum of its intensity value, the prediction error as well as the expectation (Fig. 1D). This model was inspired by studies showing the existence of both expectation and prediction error signals in different brain regions (Egner et al., 2010; Geuter et al., 2017). Prediction errors were coded in absolute differences, motivated by previous studies in the visual and auditory system (den Ouden et al., 2010; Todorovic et al., 2011; Todorovic and de Lange, 2012; Boll et al., 2013; Geuter et al., 2017). Furthermore, to account for the uneven "perceptual distance" between three levels of intensity, we added a second-order intensity term to all models. Model comparisons were performed on behavioral ratings, autonomous (SCR) data, as well as voxelwise BOLD signals and the NPS score. Model fit and parameter estimation were conducted using the FITLME function in MATLAB (version 2014b, The MathWorks).

Model comparison and statistical analysis. To find the best of four alternative models, we used a nested-model comparison in a backward selection fashion. We started with the INT+EXP+PE model as the most comprehensive model and compared it with the INT+EXP and INT+PE models using likelihood-ratio tests. The likelihood-ratio test constructs a test statistic using the log-likelihood objective function evaluated under the unrestricted model parameter estimates (INT+EXP+PE) and the restricted model parameter estimates (INT+EXP or INT+PE). The test statistic follows a χ^2 distribution with 1 df. Comparing the unrestricted INT+EXP+PE model with the two alternative-restricted INT+EXP and INT+PE models resulted in two *p* values. If both *p* values were below our threshold (e.g., *p* < 0.05), then INT+EXP+PE (unrestricted model) was

selected as the best model. Otherwise, the model associated with the biggest *p* value was discarded, and the fit to the other restricted model was compared with the fit to the INT model as the simplest model using likelihood-ratio test. If the ensuing *p* value was below the threshold, then the intermediate model (in our case either INT+EXP or INT+PE) was the winning model. Otherwise, the INT model as the simplest model was determined as the best model. In summary, model selection proceeded according to the following algorithm:

1: Fit the data to the full model (INT+EXP+PE).
2: Compare the model with the two simpler models (INT+EXP and INT+ PE).
3: Find the largest associated *p* value from the previous tests.
≫ If the largest *p* value is above threshold, go to Step 4.
≫ Otherwise, determine the full unrestricted model as the best model and stop.
4: Compare the model with smallest associated *p* value with the INT model.
≫ If the *p* value is above threshold, determine the INT model as the best model and stop.
≫ Otherwise, determine the intermediate unrestricted model as the best model and stop.

The criteria used to select each model as the winning model are also illustrated in Figure 2. In addition, we used Bayesian model weights to verify the model comparison results of the likelihood-ratio test. The Bayesian model weight w_M for the model M is given by the following:

$$ w_M = \frac{\exp\left(-\frac{1}{2}\Delta\mathrm{BIC_M}\right)}{\sum_i \exp\left(-\frac{1}{2}\Delta\mathrm{BIC_i}\right)} $$

where BIC is the Bayesian Information Criterions computed using FITLME function in MATLAB (version 2014b, The MathWorks). Each w_M is between 0 and 1, and the model with the biggest w_M is the best model. Furthermore, given the computed BIC, the Bayes factor comparing two alternative model was obtained as follows:

$$ \mathrm{BF} = \exp\left(-\frac{1}{2}\Delta\mathrm{BIC}\right) $$

Therefore, the ratio of any two w_M is equal to the corresponding Bayes factor comparing those models. For behavioral, SCR, and NPS data, the *p* value threshold was set at 0.01. At the fMRI voxel level, we used an ROI (insula) based correction for multiple comparisons using false discovery rate (FDR) correction at $p = 0.05$. FDR correction was performed using the FDR_BH function in MATLAB (version 2014b, The MathWorks). Moreover, we also tested for main effects and interactions across stimulus intensity, intensity cue, as well as modality cue using rmANOVA as implemented in MATLAB (version 2014b, The MathWorks). We also conducted paired *t* test, where all tests were two-sided unless otherwise stated. All statistical tests were conducted in MATLAB (version 2014b, The MathWorks).

Results
Behavioral results
To validate our intensity manipulation for heat pain and to verify the discriminability between different levels of pain, we first tested for the main effect of stimulus intensity (Fig. 3A). Results regarding the aversive pictures are beyond the scope of this report but are depicted in Figure 3A for the sake of comparison. An rmANOVA, based on the behavioral data, revealed significant effects for the three levels of pain ($F_{(2,62)} = 736.4$, $p = 6 \times 10^{-44}$, rmANOVA). Similarly, the main effects of intensity cue ($F_{(2,62)} = 30.7$, $p = 5 \times 10^{-5}$, rmANOVA) and modality cue ($F_{(1,31)} = 10.85$, $p = 0.0025$, rmANOVA) were significant. The two-way interaction between stimulus intensity and intensity cue ($F_{(4,124)} = $
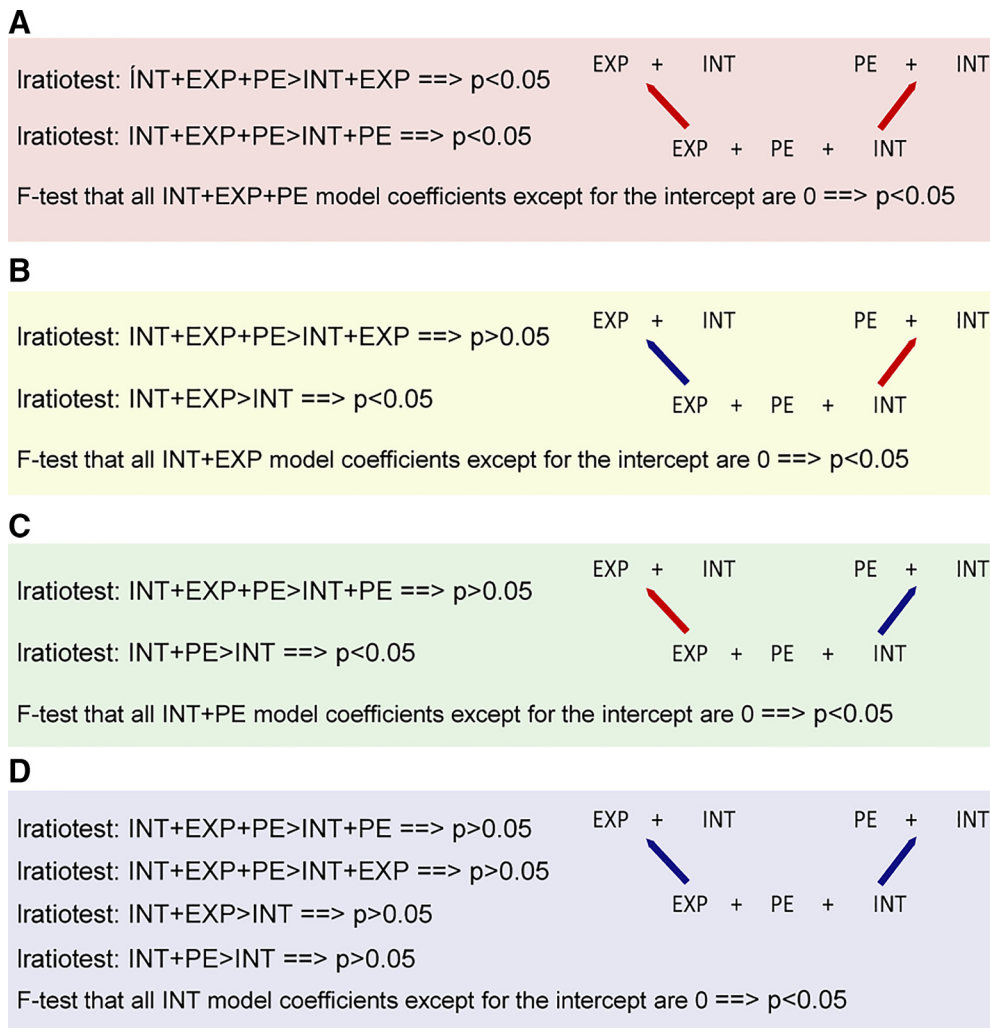
Figure 2. Flowchart of the model selection procedure. **A**, The EXP+PE+INT model will be considered the winning model if the associated p values of the likelihood ratio tests are <0.05. **B**, The EXP+INT is considered the best model, if the p value for the likelihood ratio test of EXP+PE+INT against PE+INT is <0.05, but the p value for EXP+PE+INT versus EXP+INT is >0.05. **C**, The PE+INT model is considered the best model, if the p value for the likelihood ratio test of EXP+PE+INT versus EXP+INT is <0.05, and the p value for EXP+PE+INT versus EXP+INT is >0.05. **D**, If both p values of the likelihood ratio tests are >0.05, the INT model is considered the best model.

7.09, $p = 3 \times 10^{-5}$, rmANOVA) and the interaction between modality cue and intensity cue ($F_{(2,62)} = 11.45$, $p = 0.0001$, rmANOVA) were also significant (Table 1). The rmANOVA results regarding the aversive pictures are summarized in Table 1.

In addition, we fitted the behavioral ratings to our four alternative models (INT, INT+EXP, INT+PE, and INT+EXP+PE). Backward model selection revealed that the INT+EXP model best explained the data in the conditions with a valid modality cue. In contrast, the INT model best explained the data in the invalid modality cue conditions. In addition to the likelihood-ratio test, BIC and Bayesian model weights were computed for each model to validate the consistency of our model selection routine. Comparing the four models in the valid modality cue conditions based on their corresponding BIC, INT+EXP+PE model showed the biggest weight (BIC = 144.2 and BICw = 0.6), even though the Bayes Factor comparing INT+EXP+PE and INT+EXP model was $BF_{PC>EXP} = 1.52$, which indicates their difference is not substantial; and as the expectation model (INT+EXP) has less degree of freedom, it can be determined as the best model describing the behavioral rating data in the valid modality cue conditions (Lewandowsky and Farrell, 2010). For the invalid modality cue conditions, again the INT+EXP+PE

model had the biggest weight (BIC = 144.2 and BICw = 0.29), but there was no substantial difference between INT+EXP+PE model and the other three models ($BF_{PC>EXP} = 1.18$, $BF_{PC>PE} = 1.21$, $BF_{PC>IC} = 1.34$). Therefore, the INT model as the simplest model was identified as the best model describing the behavioral ratings in the invalid modality cue conditions. In summary, both model selection strategies identified the INT model as the best model describing the behavioral ratings in the invalid modality cue conditions and the INT+EXP model as the best model for the data in the conditions with a valid modality cue.

SCR

Analogous to the behavioral data, the main effects of stimulus intensity ($F_{(2,62)} = 53.2$, $p = 3 \times 10^{-14}$, rmANOVA), intensity cue ($F_{(2,62)} = 8.48$, $p = 0.006$, rmANOVA), and modality cue ($F_{(1,31)} = 12.71$, $p = 0.001$, rmANOVA) were significant. Similar to the ratings, the two-way interaction between stimulus intensity and intensity cue ($F_{(4,124)} = 12.33$, $p = 2 \times 10^{-8}$, rmANOVA) as well as the interaction between the modality cue and the intensity cue ($F_{(2,62)} = 8.4$, $p = 0.0006$, rmANOVA) were significant (Table 1; Fig. 3B). The best model describing the SCR data were identified using likelihood-ratio test in both valid and invalid
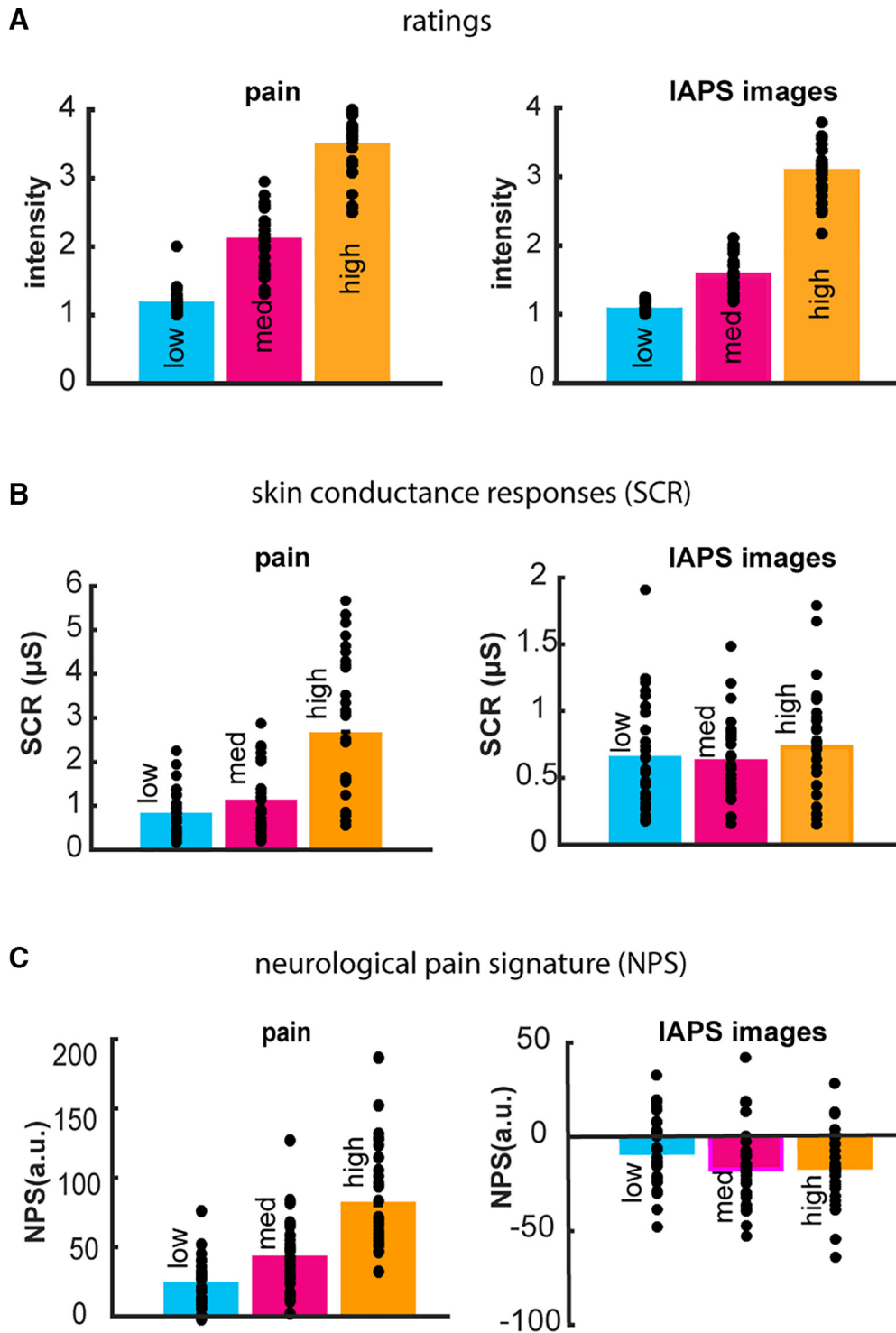
Figure 3. Behavioral, autonomous responses and NPS to thermal pain and aversive pictures. *A*, Color bars represent the subjective ratings of warm (38°), medium-painful (46°), and highly painful (48°) stimuli (left) and subjective ratings for aversiveness of IAPS pictures (right). *B*, SCRs to the pain and aversive pictures. *C*, NPS responses to painful stimuli and aversive pictures. Error bars indicate SEM across volunteers.

modality cue conditions. As in the behavioral ratings, INT+EXP was the winning model in explaining the SCR data in the valid cue conditions. The INT model was the best model in the invalid modality cue conditions. This result was verified using Bayesian model weights for the model comparisons for the valid and invalid modality cue conditions. For valid modality cue conditions, the INT+EXP model had the biggest BICw of 0.78, whereas the BICw of INT+EXP+PE model was 0.21, the

BICw of INT+PE model was 4×10^{-4}, and the INT model had a BICw of 0.002. For the invalid modality cue conditions, INT+PE had the biggest BICw of 0.47, which was only 1.25 times bigger than the BICw of the INT model with the value of 0.37. INT+EXP+PE and INT+EXP models had BICws of 0.08 and 0.06, respectively. Therefore, the INT model can be considered as the best model in the invalid modality cue conditions.

Table 1. Main effect of stimulus intensity, intensity cue, modality cue, as well as their corresponding two-way and three-way interactions: subjective ratings, SCR, and NPS in both pain and picture conditions

|	Modality cue		Intensity cue		Stimulus intensity		Modality cue × Intensity cue		Modality cue × Stimulus intensity		Intensity cue × Stimulus intensity		Modality cue × Intensity cue × Stimulus intensity	
	$F_{(1,31)}$	p	$F_{(2,62)}$	p	$F_{(2,62)}$	p	$F_{(2,62)}$	p	$F_{(2,62)}$	p	$F_{(4,124)}$	p	$F_{(4,124)}$	p
Pain														
Ratings	10.85	2.e-3	30.7	5.e-5	736.4	6.e-44	11.4	1.e-4	4.91	0.01	7.09	3.e-5	1.79	0.13
SCR	12.71	1.e-3	8.48	6.e-4	53.2	3.e-14	8.43	6.e-4	1.4	0.25	12.33	2.e-8	2.44	0.049
NPS	1.88	0.18	2.09	0.13	122.65	3.e-22	5.1	0.008	1.97	0.14	1.68	0.15	2.54	0.04
Pictures														
Ratings	0.57	0.45	6.73	0.002	794.6	6.e-45	0.14	0.86	0.87	0.42	3.37	0.01	0.84	0.49
SCR	75.29	8.e-10	3.89	0.02	1.46	0.23	7.21	0.001	0.34	0.7	8.2	6.e-6	4.32	0.002
NPS	3.41	0.07	0.4	0.67	8.19	7.e-4	2.2	0.11	0.33	0.71	1.44	0.22	1.44	0.22
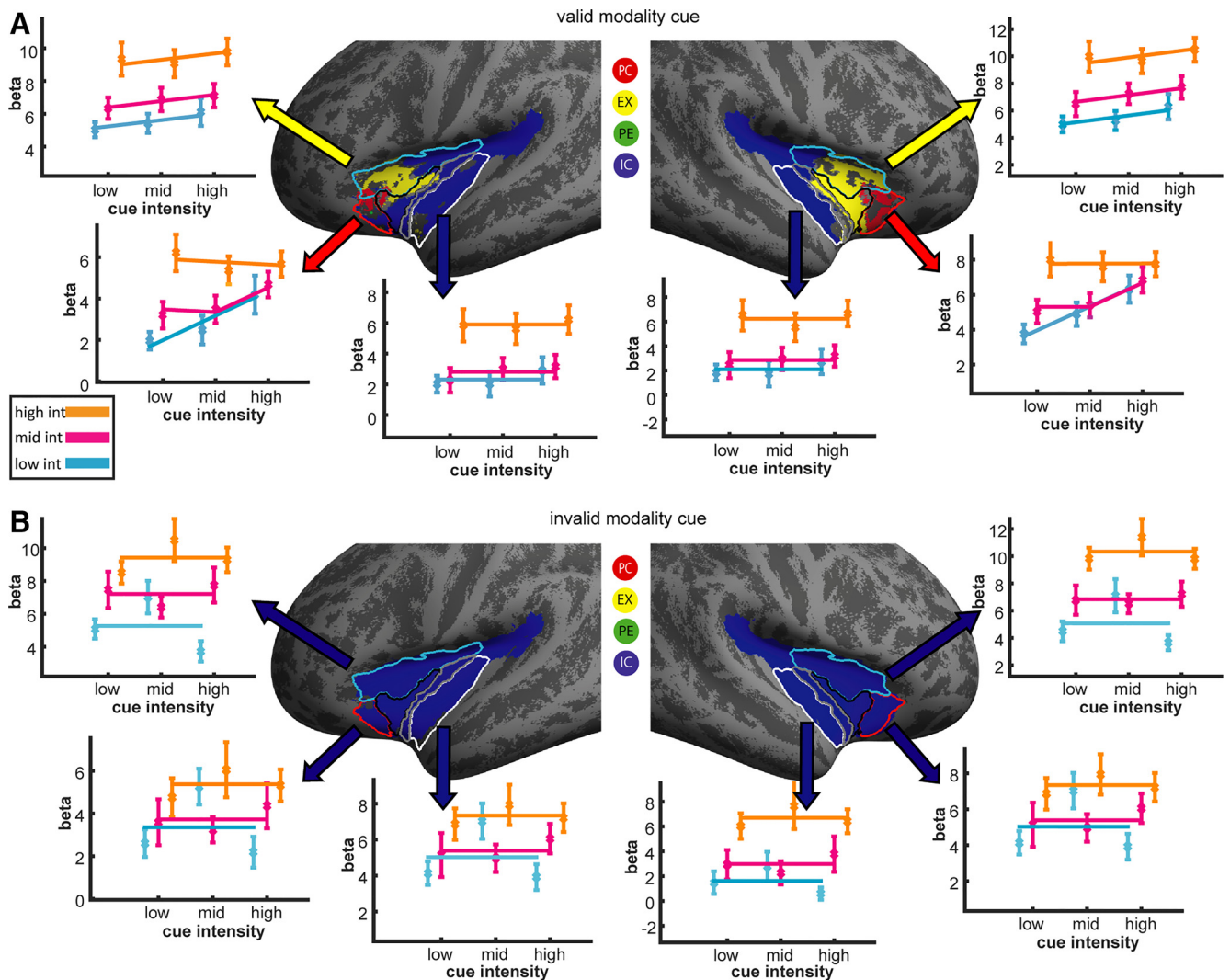
Figure 4. Voxelwise representation of the best model explaining responses to thermal pain within insular cortex. **A**, Contrast estimates were fitted to four alternative models of INT, INT + EXP, INT + PE, and INT + EXP + PE (valid modality cue). The INT + EXP + PE model (red) best explains the data in the anterior division of the circular insula (vAI). Color coding is based on likelihood-ratio test model comparison (FDR-corrected, $q < 0.05$). q values are the adjusted p values found using an optimized FDR approach in voxels within the insula. The INT + EXP model (yellow) best explains the fMRI data in the superior division of the circular insula (dAI). Within the inferior division of circular insula, the INT model (blue) is the winning model. **B**, When the modality cue was invalid, the contrast estimates were best explained by the INT model (blue) in all regions. Error bars indicate SEM across volunteers.

fMRI data

In a first step, we analyzed fMRI data from all conditions using the NPS (Wager et al., 2013). For the NPS data, an rmANOVA showed a significant effect for the three levels of pain ($F_{(2,62)} = 122.65$, $p = 3 \times 10^{-22}$, rmANOVA). Neither the main effect of intensity cue nor modality cue was significant. The two-way interaction between modality cue and intensity cue was significant ($F_{(2,62)} = 5.1$, $p = 0.008$, rmANOVA); however, the other two-way interactions were not significant (Fig. 3C; Table 1). As for the behavioral data, we fitted the NPS-derived scores to all our four
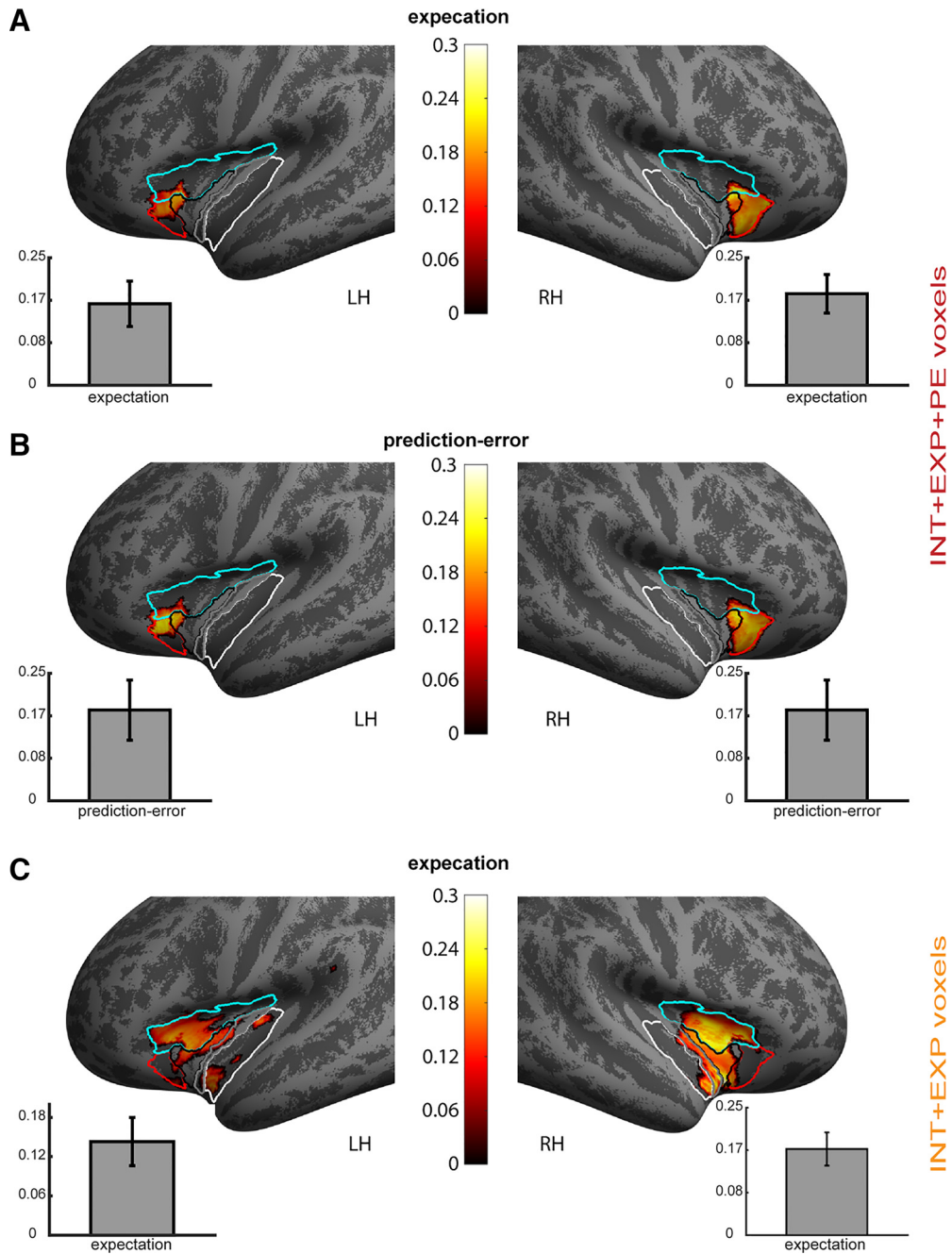
Figure 5. Voxelwise map of the respective weights of the individual components of the INT+EXP+PE and INT+EXP models within insular cortex for the valid modality cue conditions. **A**, Weights for the expectation term are plotted for all INT+EXP+PE representative voxels in insular cortex (i.e., Fig. 4, red voxels). Bar graph represents the average of the expectation weight within voxels best explained by the INT+EXP+PE model. Error bars indicate SEM across volunteers. **B**, Voxelwise map for the weight of the prediction error term as well as its average for all voxels best explained by the INT+EXP+PE model. **C**, Map of the expectation term and its average across voxels best explained by the INT+EXP model (i.e., Fig. 4, yellow voxels).

models. Analogous to the behavioral ratings, the INT+EXP model best explained the data for the valid modality cue conditions using hierarchical model comparison based on the likelihood-ratio test. In the invalid modality cue conditions, the INT model was best in explaining the NPS data based on the nested likelihood-ratio test. This was confirmed by Bayesian model weights in the valid and invalid modality cue conditions. The INT+EXP model had the biggest BICw of 0.88 (with the Bayes Factor of 14.83 compared with the INT+EXP+PE model as the second wining model) for valid modality cue conditions. The Bayesian model weights for INT+EXP+PE, INT+PE, and INT models were 0.06, 0.004, and 0.05, respectively. In the invalid

modality cue conditions, the INT model had the biggest BICw of 0.89 (with the Bayes Factor of 16.15 compared with the INT+PE model as the second wining model). The other alternative models of INT+EXP+PE, INT+EXP, and INT+PE had the BICw of 0.003, 0.053, and 0.055, respectively. As control, we also tested the NPS on picture stimuli (Table 1; Fig. 3C).

We next investigated the response patterns of voxels in subregions of insular cortex according to a surface atlas (Fischl et al., 2004; Destrieux et al., 2010). Regions in this atlas are defined according to sulcal depth (Fischl et al., 2004; Destrieux et al., 2010). To obtain an intuitive visualization of the cortical representation for each model, we color-coded surface vertices accord-
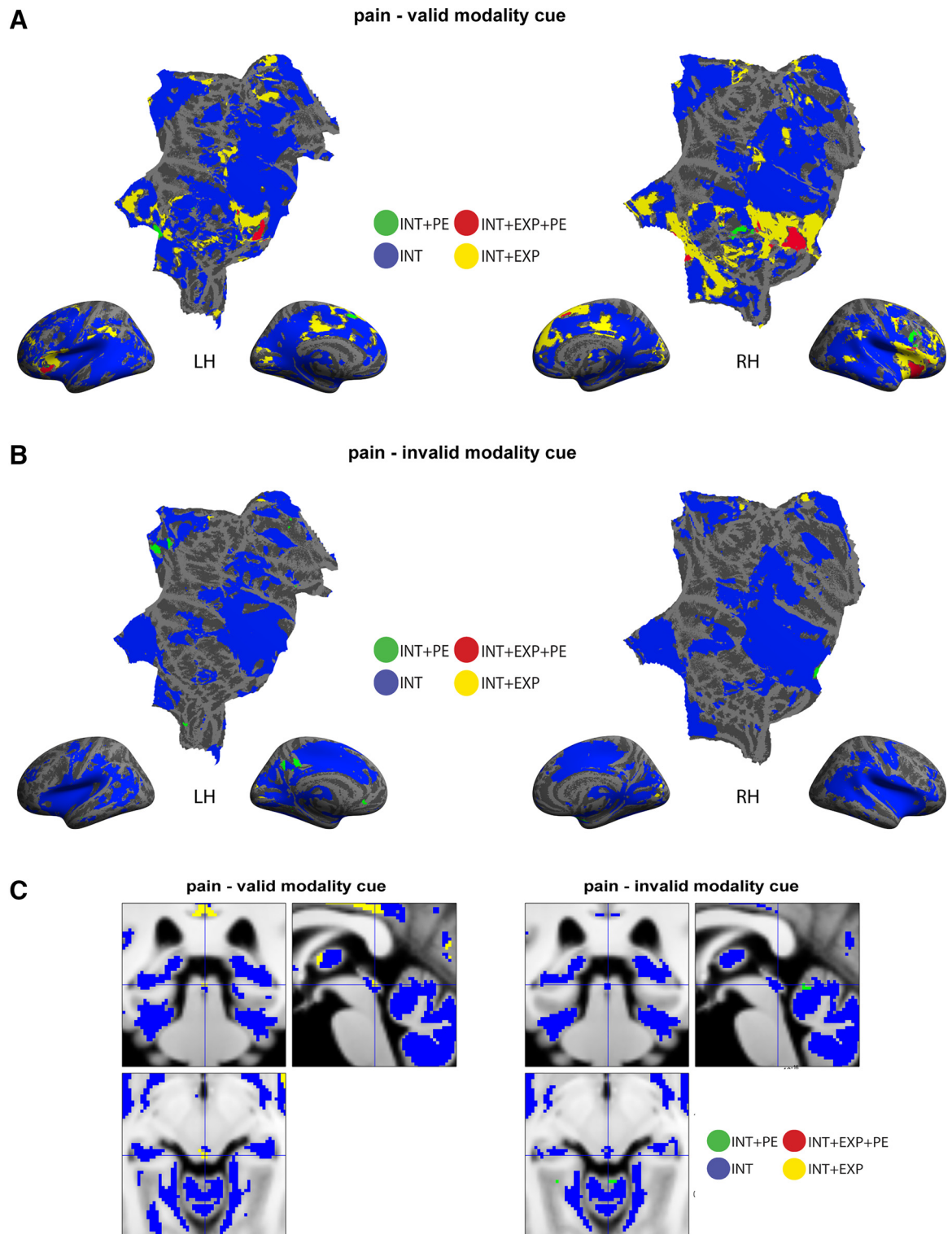
Figure 6. Color map representing the best model for each voxel at an uncorrected threshold of 0.01 projected on flattened cortex and inflated brain surfaces. ***A***, In addition to the insular clusters depicted in Figure 4, an INT+EXP cluster is also evident within the ACC and visual cortex. There is also a small cluster of the INT+PE model within the cingulate cortex (green). ***B***, For the invalid modality cue, the INT model is the best model across all pain-responsive areas. ***C***, Color-coded model selection for subcortical areas. In parts of the PAG, the INT+EXP model best described the data. For the invalid modality cue, the INT model best explains PAG responses.

ing to the best explanatory model (four different colors) on the inflated cortical surface. As shown in Figure 4, the spatial distribution of the insular subregions according to this atlas closely resembled the distribution of the winning models. Activity patterns in the anterior division of the circular insula (ventral part of anterior insula [vAI]) were best explained by the INT+EXP+PE model, activity patterns in the superior division of the circular insula (dorsal part of anterior Insula) were best explained by the INT+EXP model, and the inferior division of the circular insula showed activation patterns compatible with the INT model. To

Table 2. Main effect of stimulus intensity, intensity cue, modality cue, as well as their corresponding two-way and three-way interactions for all ROIs, including the SII, SI, frontal medial cortex, ACC, PCC, amygdala, hippocampus, thalamus, cerebellum, and PAG[a]

Region	Modality cue		Intensity cue		Stimulus intensity		Modality cue × Intensity cue		Modality cue × Stimulus intensity		Intensity cue × Stimulus intensity		Modality cue × Intensity cue × Stimulus intensity	
	$F_{(1,31)}$	p	$F_{(2,62)}$	p	$F_{(2,62)}$	p	$F_{(2,62)}$	p	$F_{(2,62)}$	p	$F_{(4,124)}$	p	$F_{(4,124)}$	p
ACC														
L	0.003	0.95	2.25	0.11	31.32	3.e-10	4.62	0.01	0.01	0.98	2.34	0.05	3.01	0.02
R	0.03	0.85	2.15	0.12	29.14	1.e-9	4.63	0.01	0.02	0.97	2.16	0.07	2.93	0.02
PCC														
L	4.01	0.05	0.56	0.57	13.7	1.e-5	3.2	0.05	0.02	0.98	2.36	0.06	2.42	0.05
R	2.61	0.12	0.32	0.72	12.82	2.e-5	2.55	0.08	0.05	0.95	1.99	0.1	2.13	0.08
Parietal operculum														
L	3.56	0.06	1.39	0.25	26.78	4.e-9	1.36	0.26	1.68	0.19	1.12	0.34	4.15	0.003
R	0.82	0.36	1.09	0.37	47.09	3.e-13	1.39	0.25	1.5	0.23	1.16	0.33	2.24	0.06
Postcentral														
L	0.37	0.54	1.86	0.16	28.97	1.e-9	0.32	0.72	1.15	0.32	2.86	0.02	2.43	0.05
R	1.e-4	0.99	1.37	0.26	24.15	1.e-8	0.58	0.55	0.09	0.9	1.6	0.17	2.06	0.08
Frontal medial cortex														
L	0.02	0.88	1.17	0.31	5.54	0.006	1.94	0.15	0.21	0.81	2.05	0.09	2.4	0.05
R	0.49	0.49	1.22	0.3	7.78	0.001	2.02	0.14	0.05	0.94	1.55	0.19	2.53	0.04
Hippocampus														
L	0.43	0.51	1.71	0.18	4.09	0.02	1.59	0.21	0.43	0.65	1.74	0.14	2.17	0.07
R	0.01	0.92	0.68	0.51	5.08	0.009	1.01	0.37	0.65	0.52	1.59	0.18	2.25	0.06
Amygdala														
L	0.1	0.75	0.39	0.67	6.77	0.002	1.72	0.18	0.52	0.59	0.75	0.55	1.51	0.2
R	0.67	0.41	0.75	0.47	8.98	3.e-4	0.94	0.39	0.32	0.72	0.99	0.14	1.87	0.11
Thalamus														
L	2.5	0.12	1.37	0.26	34.31	9.e-9	1.74	0.18	0.44	0.64	1.34	0.25	4.54	0.001
R	1.13	0.29	2.07	0.13	37.49	2.e-11	1.62	0.2	0.1	0.9	1.39	0.23	4.23	0.003
Cerebellum														
L	4.52	0.51	2.11	0.19	16.73	1.e-6	1.39	0.26	0.25	0.78	2.27	0.06	3.71	0.007
R	1.17	0.29	1.87	0.16	19.11	3.e-7	1.68	0.19	0.06	0.94	2.00	0.1	2.92	0.2
PAG	2.36	0.13	3.88	0.02	14.88	5.e-6	0.24	0.78	0.26	0.78	2.53	0.04	3.82	0.005

[a]The main effect of stimulus intensity was significant in all ROIs. A significant main effect of cue intensity was only observed in the PAG. The only significant two-way interaction between modality cue and intensity cue was identified in the ACC.

verify this, we averaged the contrast estimates in each anatomical subregion of the insula and performed a formal model comparison on this average activation pattern. For the valid modality cue conditions, the INT+EXP+PE model best accounted for the fMRI pattern in the anterior division of circular insula, the INT+EXP model best described the response pattern in the superior division of circular insula, and the INT model was the best model explaining fMRI activation patterns in the inferior division of circular insula (Fig. 4, inserted). For all invalid modality cue conditions, the INT model was the best model across all insular subdivisions. We also compared the individual contribution of the expectation and the prediction error terms in voxels best explained by the INT+EXP+PE model within the insula. For this purpose, we mapped the corresponding parameter estimates within the insular cortex on the inflated cortical surfaces (Fig. 5A,B). This analysis revealed that the prediction error parameter estimate and the expectation parameter estimate are not significantly different in both hemispheres. We also observed that the expectation weight was not significantly different in INT+EXP representative cluster compared with INT+EXP+PE voxels in both hemispheres (Fig. 5C).

In an exploratory fashion, we also performed formal model comparisons with the threshold of 0.01 outside the insular ROI (Fig. 6) on the surface and in all voxels, including subcortical areas. Here we additionally observed groups of voxels representing the INT+EXP model as the best model in the ACC, but only in valid modality cue conditions (Fig. 6A). Among the subcortical areas, the PAG showed activation patterns that were best ex-

plained by the INT+EXP model (Fig. 6C). As in cortical regions, this was only true if the modality cue was valid. Otherwise, intensity coding was identified as the best model (Fig. 6B,C).

In an explorative fashion, we also examined the averaged parameter estimates in other pain-related anatomically defined ROIs for the valid modality cue conditions. Based on rmANOVA, we found a significant main effect of stimulus intensity in all ROIs, including ACC, PCC, parietal operculum, postcentral, frontal medial cortex, hippocampus, amygdala, thalamus, cerebellum, as well as the PAG (for detailed results, see Table 2). Moreover, among all ROIs, a significant main effect of cue intensity was observed in the PAG ($F_{(2,62)} = 3.88$, $p = 0.02$, rmANOVA). The two-way interaction between modality cue and intensity cue was significant in the ACC (LH: $F_{(2,62)} = 4.62$, $p = 0.01$; RH: $F_{(2,62)} = 4.63$, $p = 0.01$, rmANOVA). With regards to model selection, we used condition-specific parameter estimates for each voxel and subjected those parameter estimates to our formal model comparison. As for the behavioral data, the model best explaining the data were identified using hierarchical likelihood-ratio tests. The INT+EXP model was determined as the best model (for detailed results, see Fig. 7; Table 3) in the ACC and PAG. In all other regions, the INT model was the winning model.

Discussion

Our data show a clear discriminability of different levels of pain intensity based on behavioral ratings, autonomic measures, and a neural compound marker (NPS). Furthermore, we observed a
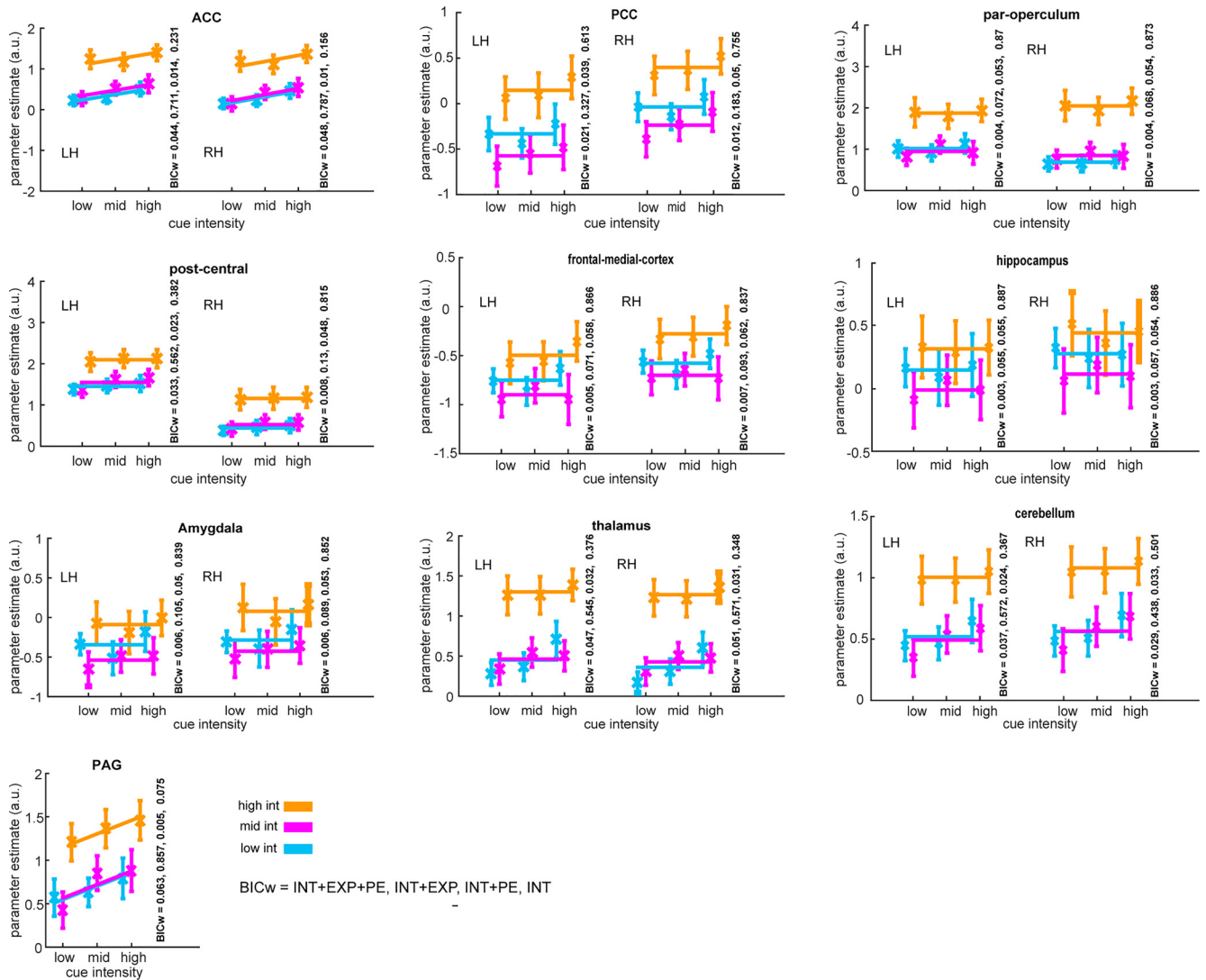
Figure 7. Parameter estimates of ROIs. Mean of the parameter estimates (± SE) at three pain intensities of high, medium, and low in different intensity cue conditions for valid modality cue trials are plotted for left (LH) and right (RH) hemispheres (except for the midline PAG) in each ROI. The fit of the parameter estimates to the best model among the four nested models is also depicted. In the ACC and PAG, the best model was determined as the INT + EXP model. Otherwise, the INT model was the best explanatory model. Bayesian model weights (BICw) for all different models are illustrated in each ROI. Error bars indicate SEM across volunteers.

strong effect of the intensity cue on pain perception, in agreement with previous findings (Koyama et al., 2005; Keltner et al., 2006; Atlas et al., 2010; Atlas and Wager, 2012; Wiech, 2016). Using computational modeling, activity patterns in the posterior insula were best explained by pure intensity coding, whereas activity in the dorsal anterior insula (dAI) was best explained by a model comprising the expected intensity. Finally, activity in the vAI was best explained by a model comprising stimulus intensity, cued intensity, and the prediction error (i.e., the difference between perceived and cued intensity) (Geuter et al., 2017). Finally, we observed a significant two-way interaction between modality cue and intensity cue, indicating a modality specific effect of expectation (Langner et al., 2011). Using formal modeling, we could reveal that the contribution of expectation in dAI and expectation plus prediction errors in vAI were only present for valid modality cues. In cases of an invalid modality cue, insular regions all showed simple intensity coding patterns.

Insular cortex

The insular cortex is crucial for pain processing (Ostrowsky et al., 2002; for review, see Nieuwenhuys (2012). It is ideally suited to

represent pain intensity and related features because of its structural and functional organization (Barrett and Simmons, 2015; Nomi et al., 2017). While the traditional assumption (Cerliani et al., 2012; Cloutman et al., 2012; Jakab et al., 2012; Dennis et al., 2014) refers to an anterior–posterior dichotomy within insular cortex, more recent neuroimaging studies suggest a tripartite organization (Deen et al., 2011; Nomi et al., 2017).

In this tripartite organization, the posterior insula with its direct spinothalamocortical projections (originating in lamina I dorsal horn neurons) (Craig, 2002; Dum et al., 2009) as well as its intense connectivity to the somatosensory cortices (Wiech et al., 2014) makes it an ideal candidate for encoding the intensity of nociceptive thermal stimuli. Such sensory coding within the PI has previously been observed for both tonic and phasic pain (Segerdahl et al., 2015; Geuter et al., 2017) and even innocuous thermal stimuli (Davis et al., 1998). Our observation of a pure intensity coding pattern in the posterior insula confirms these findings.

The dAI is functionally connected to a cognitive control network (Dosenbach et al., 2007) and has stronger connections to

Table 3. Bayesian model weights for all ROIs, including the SII, SI, frontal medial cortex, ACC, PCC, amygdala, hippocampus, thalamus, cerebellum, and PAG[a]

Region		INT + EXP + PE	INT + EXP	INT + PE	INT
ACC	L	0.04	0.71	0.01	0.23
	R	0.04	0.78	0.009	0.15
PCC	L	0.01	0.18	0.05	0.75
	R	0.02	0.32	0.04	0.61
Parietal operculum	L	0.004	0.07	0.05	0.87
	R	0.004	0.06	0.05	0.87
Postcentral	L	0.03	0.56	0.02	0.38
	R	0.007	0.12	0.04	0.81
Frontal medial cortex	L	0.007	0.09	0.06	0.83
	R	0.004	0.07	0.06	0.87
Hippocampus	L	0.003	0.06	0.05	0.89
	R	0.003	0.05	0.05	0.89
Amygdala	L	0.006	0.10	0.04	0.83
	R	0.005	0.08	0.05	0.85
Thalamus	L	0.04	0.54	0.03	0.37
	R	0.05	0.57	0.03	0.34
Cerebellum	L	0.03	0.43	0.03	0.5
	R	0.04	0.57	0.02	0.37
PAG		0.06	0.85	0.005	0.07

[a]In all ROIs except for the ACC and PAG, the INT model was the winning model. In those regions, the best model was determined as the INT + EXP model.

posterior regions of the cingulate cortex (Taylor et al., 2009). Therefore, the dAI is presumably involved in expectation and attentional mechanisms (Deen et al., 2011). Moreover, the dAI as a part of cortical limbic areas with an agranular laminar structure is considered to encode expectation signals (Barrett and Simmons, 2015; Chanes and Barrett, 2016). This is in accordance with our findings showing a representation of stimulus intensity and expectation (INT+EXP model) in the dAI. Previous studies have also identified a role of this area in prestimulus effects and attentional modulation (Ploner et al., 2010; Wiech et al., 2010).

In contrast, the vAI exhibits strong connectivity to the posterior insula, ACC, thalamus, amygdala, and basal ganglia, which are also contributing to pain processing (DuPont et al., 2003; Baliki et al., 2009; Mutschler et al., 2009; Uddin et al., 2014). Such a strong connection of the vAI to these limbic structures makes it an ideal mediator between areas representing pain intensity and expectation to evaluate the expectation against the sensory input and estimating a prediction error (Downar et al., 2002; Seymour et al., 2004; Geuter et al., 2017). Our observed representation of stimulus intensity, expectation, and prediction error (INT+EXP+PE model) in the vAI is in agreement with these findings.

Expectation and prediction error in the valid modality
Predictions and prediction errors are key ingredients for perceptual theories, such as predictive coding (Knill and Pouget, 2004), and theories of brain function, such as the free energy model, including active inference (Knill and Pouget, 2004; Friston, 2005, 2008). In the predictive coding framework, the brain is relying on its internal model of the physical reality but tries to make it an adequate model by updating it based on the incoming data. Consequently, the integration of "prediction errors" into this model will change it over time to better reflect the physical reality. These concepts can account for extra-classical-receptive-field effects (Rao and Ballard, 1999), repetition suppression (Summerfield et al., 2008; Todorovic and de Lange, 2012), illusory contours formation (Dura-Bernal et al., 2011), and mismatch negativity (Ylinen et al., 2016). Such a mechanism was also proposed as a model to account for expectation effects in pain (Geuter et al., 2017) and

placebo hypoalgesia (Büchel et al., 2014; Grahl et al., 2018). According to predictive coding in the case of pain, we hypothesize the presence of sensory specific representations of the expectation and prediction error terms at the hierarchical level where the intensity of pain is encoded similar to other modalities (Rao and Ballard, 1999; Friston, 2005). Crucially, this predicts that these representations are restricted to the case in which the modality cue is valid. As our second stimulus category comprised aversive pictures, this allows to test whether aversiveness, which is shared by pain and picture stimuli, can account for insula activation.

Importantly, our data show that the prediction error signal in the anterior insula is not generally related to aversiveness, which is shared by the pictures and the painful stimuli, but is limited to the pain expectation. This is indicated by the fact that we could only observe the "expectation" effect in the valid modality cue conditions. In case the modality cue was invalid (when volunteers expected an aversive picture but received pain), the intensity cue had no effect on the response patterns in the insula. In other words, an expectation that shares an aversiveness component (i.e., aversive pictures) is not encoded in these areas. This is further corroborated by the absence of a prediction error in the insula in this case. As a consequence, our data suggest a more specific role for the vAI, representing expectations and prediction errors related to pain. This indicates that this area is part of the hierarchy where pain intensity integration occurs and can be seen in analogy to other stimulus modalities, in which feature-selective predictive coding within the fusiform face area in response to faces, and within the parahippocampal place area in response to houses (den Ouden et al., 2010; Egner et al., 2010; Jiang et al., 2013) has been shown. Moreover, the observed functional dissociation between insular subregions resonates with previous accounts of insular subspecialization for different features of aversive stimuli (Baliki et al., 2009; Deen et al., 2011; Nomi et al., 2017).

Signed versus unsigned prediction errors
Prediction errors could either be represented as absolute deviations (i.e., absolute deviation of outcome from expectation) or signed deviations (i.e., more or less pain than expected). Unfortunately, due to the design of our experiment, distinguishing between unsigned and signed prediction errors cannot be dissociated in our model. Our data indicate the representation of unsigned prediction error in the anterior insula. Similar observations of unsigned prediction errors have been made in the anterior insula, amygdala, and visual and auditory cortices (den Ouden et al., 2009, 2010; Todorovic et al., 2011; Todorovic and de Lange, 2012; Boll et al., 2013; Geuter et al., 2017). Unlike the signed prediction error, which can be used to guide approach and avoidance behavior, this type of error (i.e., absolute prediction error) determines the captured attention, which is related to stimulus saliency rather than its actual value (Kahnt et al., 2014).

In conclusion, the vAI exhibits all features necessary for the implementation of a predictive coding type integration of pain. Importantly, the functional organization revealed by our model comparison is remarkably similar to the structurally defined tripartite organization of the insular cortex (Fischl et al., 2004; Destrieux et al., 2010). Our results corroborate the previously observed gradient from stimulus intensity coding to perceptual coding from posterior to anterior insula (Geuter et al., 2017). This might also reflect a gradient from nociception to pain, where somatosensory information from posterior somatosensory regions is more and more integrated with cognitive factors such as

expectation, which finally allows the estimation of prediction errors in the anterior insula.

References

Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38:95–113. CrossRef Medline

Atlas LY, Wager TD (2012) How expectations shape pain. Neurosci Lett 520:140–148. CrossRef Medline

Atlas LY, Bolger N, Lindquist MA, Wager TD (2010) Brain mediators of predictive cue effects on perceived pain. J Neurosci 30:12964–12977. CrossRef Medline

Baliki MN, Geha PY, Apkarian AV (2009) Parsing pain perception between nociceptive representation and magnitude estimation. J Neurophysiol 101:875–887. CrossRef Medline

Barrett LF, Simmons WK (2015) Interoceptive predictions in the brain. Nat Rev Neurosci 16:419–429. CrossRef Medline

Benedek M, Kaernbach C (2010) Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology 47:647–658. CrossRef Medline

Boll S, Gamer M, Gluth S, Finsterbusch J, Büchel C (2013) Separate amygdala subregions signal surprise and predictiveness during associative fear learning in humans. Eur J Neurosci 37:758–767. CrossRef Medline

Büchel C, Geuter S, Sprenger C, Eippert F (2014) Placebo analgesia: a predictive coding perspective. Neuron 81:1223–1239. CrossRef Medline

Cerliani L, Thomas RM, Jbabdi S, Siero JC, Nanetti L, Crippa A, Gazzola V, D'Arceuil H, Keysers C (2012) Probabilistic tractography recovers a rostrocaudal trajectory of connectivity variability in the human insular cortex. Hum Brain Mapp 33:2005–2034. CrossRef Medline

Chanes L, Barrett LF (2016) Redefining the role of limbic areas in cortical processing. Trends Cogn Sci 20:96–106. CrossRef Medline

Cloutman LL, Binney RJ, Drakesmith M, Parker GJ, Lambon Ralph MA (2012) The variation of function across the human insula mirrors its patterns of structural connectivity: evidence from in vivo probabilistic tractography. Neuroimage 59:3514–3521. CrossRef Medline

Craig AD (2002) How do you feel? Interoception: the sense of the physiological condition of the body. Nat Rev Neurosci 3:655–666. CrossRef Medline

Davis KD, Kwan CL, Crawley AP, Mikulis DJ (1998) Functional MRI study of thalamic and cortical activations evoked by cutaneous heat, cold, and tactile stimuli. J Neurophysiol 80:1533–1546. CrossRef Medline

Deen B, Pitskel NB, Pelphrey KA (2011) Three systems of insular functional connectivity identified with cluster analysis. Cereb Cortex 21:1498–1506. CrossRef Medline

den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009) A dual role for prediction error in associative learning. Cereb Cortex 19:1175–1185. CrossRef Medline

den Ouden HE, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010) Striatal prediction error modulates cortical coupling. J Neurosci 30:3210–3219. CrossRef Medline

Dennis EL, Jahanshad N, McMahon KL, de Zubicaray GI, Martin NG, Hickie IB, Toga AW, Wright MJ, Thompson PM (2014) Development of insula connectivity between ages 12 and 30 revealed by high angular resolution diffusion imaging. Hum Brain Mapp 35:1790–1800. CrossRef Medline

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31:968–980. CrossRef Medline

Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53:1–15. CrossRef Medline

Dosenbach NU, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RA, Fox MD, Snyder AZ, Vincent JL, Raichle ME, Schlaggar BL, Petersen SE (2007) Distinct brain networks for adaptive and stable task control in humans. Proc Natl Acad Sci U S A 104:11073–11078. CrossRef Medline

Downar J, Crawley AP, Mikulis DJ, Davis KD (2002) A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. J Neurophysiol 87:615–620. CrossRef Medline

Dum RP, Levinthal DJ, Strick PL (2009) The spinothalamic system targets motor and sensory areas in the cerebral cortex of monkeys. J Neurosci 29:14223–14235. CrossRef Medline

DuPont S, Bouilleret V, Hasboun D, Semah F, Baulac M (2003) Functional anatomy of the insula: new insights from imaging. Surg Radiol Anat 25:113–119. CrossRef Medline

Dura-Bernal S, Wennekers T, Denham SL (2011) The role of feedback in a hierarchical model of object perception. Adv Exp Med Biol 718:165–179. CrossRef Medline

Egner T, Monti JM, Summerfield C (2010) Expectation and surprise determine neural population responses in the ventral visual stream. J Neurosci 30:16601–16608. CrossRef Medline

Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Gunther M, Glasser MF, Miller KL, Ugurbil K, Yacoub E (2010) Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PLoS One 5:e15710. CrossRef Medline

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM (2004) Automatically parcellating the human cerebral cortex. Cereb Cortex 14:11–22. CrossRef Medline

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815–836. CrossRef Medline

Friston K (2008) Hierarchical models in the brain. PLoS Comput Biol 4:e1000211. CrossRef Medline

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:e24770. CrossRef Medline

Grahl A, Onat S, Büchel C (2018) The periaqueductal gray and Bayesian integration in placebo analgesia. eLife 7:e32930. CrossRef Medline

Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A (2002) Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 47:1202–1210. CrossRef Medline

Häuser W, Hansen E, Enck P (2012) Nocebo phenomena in medicine: their relevance in everyday clinical practice. Dtsch Arztebl Int 109:459–465. CrossRef Medline

Jakab A, Molnár PP, Bogner P, Béres M, Berényi EL (2012) Connectivity-based parcellation reveals interhemispheric differences in the insula. Brain Topogr 25:264–271. CrossRef Medline

Jiang J, Summerfield C, Egner T (2013) Attention sharpens the distinction between expected and unexpected percepts in the visual brain. J Neurosci 33:18438–18447. CrossRef Medline

Kahnt T, Park SQ, Haynes JD, Tobler PN (2014) Disentangling neural representations of value and salience in the human brain. Proc Natl Acad Sci U S A 111:5000–5005. CrossRef Medline

Keltner JR, Furst A, Fan C, Redfern R, Inglis B, Fields HL (2006) Isolating the modulatory effect of expectation on pain transmission: a functional magnetic resonance imaging study. J Neurosci 26:4437–4443. CrossRef Medline

Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci 27:712–719. CrossRef Medline

Koyama T, McHaffie JG, Laurienti PJ, Coghill RC (2005) The subjective experience of pain: where expectations become reality. Proc Natl Acad Sci U S A 102:12950–12955. CrossRef Medline

Krishnan A, Woo CW, Chang LJ, Ruzic L, Gu X, López-Solà M, Jackson PL, Pujol J, Fan J, Wager TD (2016) Somatic and vicarious pain are represented by dissociable multivariate brain patterns. eLife 5:e15166. CrossRef Medline

Lang PJ, Bradley MM, Cuthbert BN (2005) International affective picture system (IAPS): affective ratings of pictures and instruction manual. Gainesville, FL: National Institute of Mental Health, Center for the Study of Emotion and Attention.

Langner R, Kellermann T, Boers F, Sturm W, Willmes K, Eickhoff SB (2011) Modality-specific perceptual expectations selectively modulate baseline activity in auditory, somatosensory, and visual cortices. Cereb Cortex 21:2850–2862. CrossRef Medline

Lewandowsky S, Farrell S (2010) Computational Modeling in Cognition: Principles and Practice. California, CA: SAGE Publications, Inc.

Moeller S, Yacoub E, Olman CA, Auerbach E, Strupp J, Harel N, Uğurbil K (2010) Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn Reson Med 63:1144–1153. CrossRef Medline

Mutschler I, Wieckhorst B, Kowalevski S, Derix J, Wentlandt J, Schulze-Bonhage A, Ball T (2009) Functional organization of the human anterior insular cortex. Neurosci Lett 457:66–70. CrossRef Medline

Nieuwenhuys R (2012) The insular cortex: a review. Prog Brain Res 195: 123–163. CrossRef Medline

Nomi JS, Schettini E, Broce I, Dick AS, Uddin LQ (2017) Structural connections of functionally defined human insular subdivisions. Cereb Cortex 2017:1–12. CrossRef Medline

Okada K, Matchin W, Hickok G (2018) Neural evidence for predictive coding in auditory cortex during speech production. Psychon Bull Rev 25: 423–430. CrossRef Medline

Ostrowsky K, Magnin M, Ryvlin P, Isnard J, Guenot M, Mauguière F (2002) Representation of pain and somatic sensation in the human insula: a study of responses to direct electrical cortical stimulation. Cereb Cortex 12:376–385. CrossRef Medline

Ploner M, Lee MC, Wiech K, Bingel U, Tracey I (2010) Prestimulus functional connectivity determines pain perception in humans. Proc Natl Acad Sci U S A 107:355–360. CrossRef Medline

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79– 87. CrossRef Medline

Segerdahl AR, Mezue M, Okell TW, Farrar JT, Tracey I (2015) The dorsal posterior insula subserves a fundamental role in human pain. Nat Neurosci 18:499–500. CrossRef Medline

Setsompop K, Gagoski BA, Polimeni JR, Witzel T, Wedeen VJ, Wald LL (2012) Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. Magn Reson Med 67:1210–1224. CrossRef Medline

Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS (2004) Temporal difference models describe higher-order learning in humans. Nature 429:664– 667. CrossRef Medline

Sterzer P, Frith C, Petrovic P (2008) Believing is seeing: expectations alter visual awareness. Curr Biol 18:R697–R698. CrossRef Medline

Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T (2008) Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci 11:1004–1006. CrossRef Medline

Taylor KS, Seminowicz DA, Davis KD (2009) Two systems of resting state connectivity between the insula and cingulate cortex. Hum Brain Mapp 30:2731–2745. CrossRef Medline

Todorovic A, de Lange FP (2012) Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. J Neurosci 32:13389–13395. CrossRef Medline

Todorovic A, van Ede F, Maris E, de Lange FP (2011) Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. J Neurosci 31:9118–9123. CrossRef Medline

Tracey I (2010) Getting the pain you expect: mechanisms of placebo, nocebo and reappraisal effects in humans. Nat Med 16:1277–1283. CrossRef Medline

Uddin LQ, Kinnison J, Pessoa L, Anderson ML (2014) Beyond the tripartite cognition-emotion-interoception model of the human insular cortex. J Cogn Neurosci 26:16–27. CrossRef Medline

Wager TD, Atlas LY, Leotti LA, Rilling JK (2011) Predicting individual differences in placebo analgesia: contributions of brain activity during anticipation and pain experience. J Neurosci 31:439–452. CrossRef Medline

Wager TD, Atlas LY, Lindquist MA, Roy M, Woo CW, Kross E (2013) An fMRI-based neurologic signature of physical pain. N Engl J Med 368: 1388–1397. CrossRef Medline

Wiech K (2016) Deconstructing the sensation of pain: the influence of cognitive processes on pain perception. Science 354:584–587. CrossRef Medline

Wiech K, Ploner M, Tracey I (2008) Neurocognitive aspects of pain perception. Trends Cogn Sci 12:306–313. CrossRef Medline

Wiech K, Lin CS, Brodersen KH, Bingel U, Ploner M, Tracey I (2010) Anterior insula integrates information about salience into perceptual decisions about pain. J Neurosci 30:16324–16331. CrossRef Medline

Wiech K, Jbabdi S, Lin CS, Andersson J, Tracey I (2014) Differential structural and resting state connectivity between insular subdivisions and other pain-related brain regions. Pain 155:2047–2055. CrossRef Medline

Xu J, Moeller S, Auerbach EJ, Strupp J, Smith SM, Feinberg DA, Yacoub E, Uğurbil K (2013) Evaluation of slice accelerations using multiband echo planar imaging at 3 T. Neuroimage 83:991–1001. CrossRef Medline

Ylinen S, Huuskonen M, Mikkola K, Saure E, Sinkkonen T, Paavilainen P (2016) Predictive coding of phonological rules in auditory cortex: a mismatch negativity study. Brain Lang 162:72– 80. CrossRef Medline