

<https://doi.org/10.1038/s42003-024-06574-y>

# An externally validated resting-state brain connectivity signature of pain-related learning

Check for updates

Balint Kincses<sup>1,2</sup>✉, Katarina Forkmann<sup>1</sup>, Frederik Schlitt<sup>1</sup>, Robert Jan Pawlik<sup>3</sup>, Katharina Schmidt<sup>1</sup>, Dagmar Timmann<sup>1</sup>, Sigrid Elsenbruch<sup>1,3</sup>, Katja Wiech<sup>4</sup>, Ulrike Bingel<sup>1,5</sup> & Tamas Spisak<sup>1,2,5</sup>

Pain can be conceptualized as a precision signal for reinforcement learning in the brain and alterations in these processes are a hallmark of chronic pain conditions. Investigating individual differences in pain-related learning therefore holds important clinical and translational relevance. Here, we developed and externally validated a novel resting-state brain connectivity-based predictive model of pain-related learning. The pre-registered external validation indicates that the proposed model explains 8–12% of the inter-individual variance in pain-related learning. Model predictions are driven by connections of the amygdala, posterior insula, sensorimotor, frontoparietal, and cerebellar regions, outlining a network commonly described in aversive learning and pain. We propose the resulting model as a robust and highly accessible biomarker candidate for clinical and translational pain research, with promising implications for personalized treatment approaches and with a high potential to advance our understanding of the neural mechanisms of pain-related learning.

Chronic pain exhibits considerable inter-individual variability which has been linked to a combination of physical, neural, psychological, and social factors<sup>1–3</sup>. Recent multivariate studies using brain-based predictive models<sup>4–8</sup> and “neurotraits”<sup>9</sup> hold promise to provide a composite window into pain states and, ultimately, escape the constraints of the case-control diagnostic paradigm by offering a set of objective measures of individual-level pain characteristics<sup>10</sup>.

An underexplored avenue for expanding the pool of potential brain-based biomarkers of pain is to target individual differences in pain-related learning processes. Pain and learning are intimately linked, and can be conceptualized within a common framework, with pain being a precision signal for reinforcement learning in the brain<sup>11</sup>. Pavlovian conditioning plays a fundamental role in many aspects of pain behavior<sup>12</sup>, including fear and threat detection<sup>13</sup>, escape and relief learning<sup>14</sup>, and even in placebo analgesia<sup>15</sup>. In acute pain, we learn to associate painful experiences with certain stimuli or situations, which helps us to adapt by learning to avoid or minimize future harm. In persistent pain states, this learning can become maladaptive and a key driver of pain chronification<sup>16,17</sup>. Understanding individual differences in pain-related learning and its underlying neural mechanisms is therefore key not only to a comprehensive understanding of

(chronic) pain but also to the development of effective prevention and treatment options.

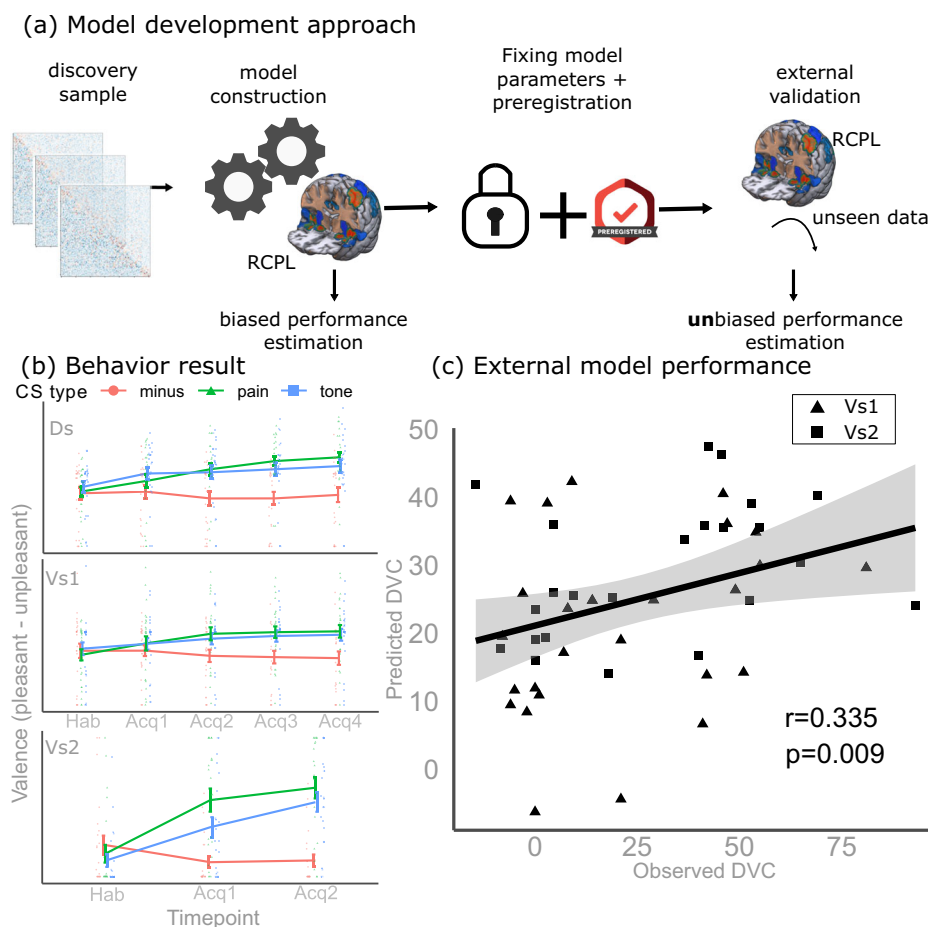
Previous structural and functional imaging studies suggest that the underlying neural processes of aversive learning are rooted in a network of brain regions such as the amygdala<sup>18–22</sup>, hippocampus<sup>19,23,24</sup>, dorsal anterior cingulate cortex<sup>24,25</sup>, posterior insula<sup>24,26</sup>, cerebellum<sup>24,27</sup>, and prefrontal regions<sup>19,24</sup>. Although pain-related aversive learning and pain conditioning clearly use the neural mechanisms of general aversive learning and threat conditioning, recent research also suggests a clear distinction in mechanisms and the presence of networks that are specific to pain-related learning<sup>28,29</sup>. Pain-related learning performance is relatively consistent within an individual over time<sup>30</sup> and the corresponding neural substrates also seem to be reflected in resting brain activity, i.e., in the absence of an explicit task<sup>31–36</sup>. However, it has yet to be determined whether resting-state brain connectivity can characterize and provide a better understanding of individual differences in pain-related learning.

Here we aimed to develop and evaluate a resting-state functional brain connectivity-based multivariate predictive marker of individual differences in pain-related learning, as measured by an established classical conditioning paradigm.

<sup>1</sup>Department of Neurology, Center for Translational Neuro- and Behavioral Sciences, University Medicine Essen, Essen, Germany. <sup>2</sup>Institute for Diagnostic and Interventional Radiology and Neuroradiology, University Medicine Essen, Essen, Germany. <sup>3</sup>Department of Medical Psychology and Medical Sociology, Faculty of Medicine, Ruhr University Bochum, Bochum, Germany. <sup>4</sup>Wellcome Centre for Integrative Neuroimaging, FMRI, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. <sup>5</sup>These authors contributed equally: Ulrike Bingel, Tamas Spisak. ✉e-mail: [balint.kincses@uk-essen.de](mailto:balint.kincses@uk-essen.de)

### Fig. 1 | Functional connectivity model predicts pain-related learning performance.

**a** Generalizable predictive performance estimates were obtained via a “registered model” design. Our resting-state connectivity-based marker of pain-related learning (RCPL) model was developed in a discovery sample. Prior to estimating the model’s true, unbiased performance in an independent external sample, we fixed all model parameters and pre-registered our approach. **b** Behavioral results: in all three datasets (Ds – discovery sample ( $n = 25$ ), Vs1 – validation sample 1 ( $n = 26$ ), Vs2 – validation sample 2 ( $n = 23$ )), we observed an increase in valence ratings of the pain- and tone-related conditioned stimuli ( $CS^+_{\text{pain}}$  and  $CS^+_{\text{tone}}$ , respectively), as compared to the safety signal ( $CS^-$ ), indicating successful conditioning. Mean and standard error of the ratings in each experimental phase are depicted by line plots, separately for all three datasets. Individual rating values are visualized (red dot –  $CS^-$ ; green triangle –  $CS^+_{\text{pain}}$ ; blue rectangle –  $CS^+_{\text{tone}}$ ). **c** Model performance: our model successfully predicted individual differential valence ratings (DVC, see Eq. 1) in the external validation sample ( $r = 0.335$ ,  $p = 0.009$ ). Validation sample 1 ( $n = 26$ ) and 2 ( $n = 23$ ) are denoted by marker shape (validation sample 1 – triangle, validation sample 2 – rectangle).



To ensure transparency and reliability, we have followed a so-called “registered model” design<sup>37</sup> that has addressed concerns about overly optimistic performance estimates during model discovery<sup>38–44</sup>. This included (i) limiting analytical flexibility by using a standard preprocessing pipeline and methodological recommendations from the literature, (ii) fixing and publicly registering all model parameters (including data preprocessing) after the model discovery phase, (iii) externally validating the proposed model on prospective and retrospective datasets, to provide a robust and unbiased assessment of predictive performance. Finally, we aimed to evaluate the model in terms of specificity for pain-related learning, bias toward potential confounders, and neuroscientific validity/plausibility of the underlying predictive brain network.

## Results

### Pain and aversive tone stimuli elicited associative learning behavior

We assessed individual learning performance related to pain and aversive tone using a differential conditioning paradigm (Supplementary Fig. 1) in a total of  $n = 99$  ( $n = 74$ , after exclusion) young healthy volunteers (age:  $25.3 \pm 3.9$  years (mean ( $\pm$ sd)) in three separate studies (discovery sample + validation samples 1 and 2) see Supplementary Table 1 for study-wise demographic summaries). In brief, the unconditioned stimuli (US pain and tone) were preceded with geometric figures as conditioned stimulus (CS) and participants rated the unpleasantness of these CS (via valence ratings). Another geometric figure that was never followed by a US served as control ( $CS^-$ ). Over the course of the experiment, participants learned to associate the CSs with the USs which was reflected in the change of valence ratings (see section “Methods” and Fig. 1b for more details).

In all three studies, the behavioral results showed evidence for threat learning with both CS (see Fig. 1b – Behavior results,  $p < 0.05$  for all, Table 1, Supplementary Tables 2–5).

Individual pain-related learning performance was characterized by differential valence ratings (see Eq. 1), i.e., the degree to which the valence difference between  $CS^+_{\text{pain}}$  and  $CS^-$  (see Eq. 1) changed over the course of the experiment for a given individual. Differential valence ratings served as the target measure for the discovery and validation of our brain-based predictive model of pain-related learning.

### Resting connectivity-based signature predicted individual pain-related learning

Resting-state functional MRI data were collected from all 99 participants ( $n = 74$ , after exclusion) in the three studies for a duration of 8–10 min/participant, prior to the behavioral experiments (see Supplementary Table 1). The model development was performed exclusively in the discovery study ( $n = 25$ , after exclusions), prior to acquiring the validation data. We trained a machine learning model, based on whole-brain resting-state functional brain connectivity, assessed between  $M = 122$  functionally defined regions (BASC atlas)<sup>45</sup> to predict pain-related learning performance as quantified by differential valence ratings (Eq. 1, Supplementary Fig. 2). Nested leave-one-participant-out (LOPO) cross-validation indicated a significant predictive performance in the discovery sample (Supplementary Fig. 3). The final regularized linear model (Ridge) was fitted on the discovery sample and contained ten predictive connections with a regularization parameter of  $\alpha = 0.001$ . However, recent studies<sup>39,44</sup> have raised concerns about the reliability, replicability, and generalizability of effect sizes measured in small-sample studies, particularly when using LOPO cross-validation<sup>38</sup>. Therefore, we evaluated the true, unbiased predictive effect size of our model by means of an independent, pre-registered external validation, based on two, independent, heterogeneous samples that were not available during model discovery.

Prior to obtaining the external validation data, we froze all model weights, fixed all preprocessing parameters, and pre-registered our

**Table 1 | Behavioral results**

	Pain		Aversive tone	
	Valence $\Delta(\Delta CS^+_{\text{pain}}, \Delta CS^-) (95\% \text{ CI})$	Contingency <sup>a</sup> $CS^+_{\text{pain}} (95\% \text{ CI})$	Valence $\Delta(\Delta CS^+_{\text{tone}}, \Delta CS^-) (95\% \text{ CI})$	Contingency <sup>a</sup> $CS^+_{\text{tone}} (95\% \text{ CI})$
Discovery study ( <i>n</i> = 25)	25 (16–34)	56 (40–70)	15 (7–25)	22 (1–42)
Validation study ( <i>n</i> = 49)				
Sub-sample 1 ( <i>n</i> = 26)	21 (12–31)	48 (29–65)	15 (4–25)	19 (1–37)
Sub-sample 2 ( <i>n</i> = 23)	29 (18–41)	66 (50–80)	27 (16–38)	70 (55–84)

Results show evidence for threat learning in both the pain and the unpleasant tone conditions, in all three studies. Values represent the mean and bootstrapped confidence intervals value in parenthesis.  $\Delta CS^+_{\text{pain}}$  – valence change between the end of acquisition and habituation for the conditioned stimulus associated with pain,  $\Delta CS^+_{\text{tone}}$  – valence change between the end of acquisition and habituation for the conditioned stimulus associated with aversive tone,  $\Delta CS^-$  – valence change between the end of acquisition and habituation for the conditioned stimulus associated with no aversive stimulus. See Eq. 1 for more details.  $CS^+_{\text{pain}}$  – contingency ratings of the conditioned stimulus associated with pain,  $CS^+_{\text{tone}}$  – contingency ratings of the conditioned stimulus associated with aversive tone. A similar table with all the participants (without exclusion) can be found in the Supplementary Material (Supplementary Table 4).

<sup>a</sup>Contingency ratings were measured with different scales in the discovery and validation study 1 as compared to validation study 2 (see the section “Methods”, subsection “Behavioral paradigm”).

approach (<https://osf.io/b8znd>, Fig. 1a – Model development approach). We refer to this model as the Resting-state functional Connectivity signature of Pain-related Learning (RCPL).

The external validation involved no model fitting. Instead, we simply obtained the RCPL predictions independently for each participant and contrasted them to the behavioral results, pooled across all participants in the external validation sample.

In this independent data, we observed a highly significant predictive performance ( $r = 0.34$ ,  $CI_{90\%}: [0.14 \text{ } 0.53]$ ,  $p = 0.009$ , Fig. 1c – External model performance). The observed effect size estimates were consequent in both sub-samples, ( $r = 0.28$ ,  $CI_{90\%}: [-0.01 \text{ } 0.54]$  and  $r = 0.35$ ,  $CI_{90\%}: [0.04 \text{ } 0.68]$  in validation sample 1 and 2, respectively) showing that the RCPL model can be expected to explain 8–12% of variance in independent datasets (based in validation study 1 and 2, respectively; see Supplementary Fig. 3 – External samples). The root mean squared error values were 26.8, 25.03, and 26.9 for the merged sample, validation sample 1, and validation sample 2, respectively. The range of the observed values was  $[-14.5, 98.0]$  within the validation dataset. Interestingly, completely dropping our pre-registered motion-exclusion (~20% more individuals), the performance of the model remained unchanged ( $r = 0.35$ ,  $p = 0.0028$ , see also Supplementary Notes: Model performance on the extended sample, Supplementary Fig. 4).

### Specificity to pain-related learning and confounding bias

To be considered as a valuable tool for basic and translational clinical research, our findings on the externally validated, out-of-sample predictive performance of our proposed model must be complemented by an initial investigation of the model’s convergent and divergent validity<sup>41</sup>. Here we aimed to characterize the model’s (i) generalizability to other measures of pain-related learning (convergent validity to contingency ratings and extinction learning), (ii) its specificity to pain (divergent validity to tone-related learning), and (iii) confounding bias (divergent validity to various confounding variables, such as fear of pain or in-scanner motion artifacts).

Our criteria for generalizability, specificity, and confounding go beyond the common practice of drawing conclusions simply from bivariate associations between the model predictions and the (convergent and divergent) validator variables. Instead of this arguably misleading approach, we focus on the conditional independence structure across all three variables involved (i.e., target, prediction, validator) with a dedicated statistical test<sup>46</sup>. In our definition, predictions are specific to the target variable, if they are conditionally independent of the validator variable, i.e., their (marginal) association is not stronger than would be expected from the association between the target and the validator variable. Conversely, if the predictions are conditionally dependent on the validator variable (i.e., the validator explains variance in the predictions over and beyond the expected marginal association), we say that the predictions generalize to the validator variable. In our view, confounding bias is equivalent to an (unwanted) generalization to a divergent validator (for more details on our approach see ref. 46). The concepts of convergent and divergent validity are not yet fully developed in the field of machine learning. Importantly, the term generalization to other

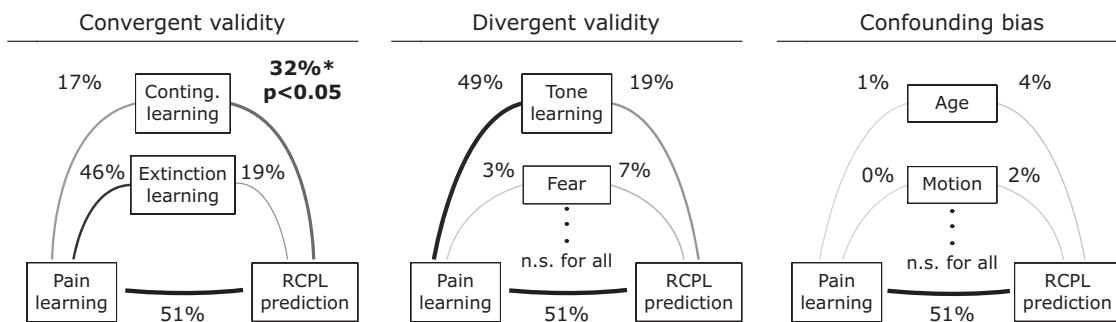
measures of pain (or convergent validity) should not be confused with the model generalization to new, unseen data (i.e., external validity see above). Additionally, our specificity analysis (divergent validity) tests whether the model captures aversive learning (tone-related) in general and differs from the similar term used in classification problems.

Differences between specificity assessment with bivariate measures and conditional independence can be illustrated by testing the convergent validity of our model with another measure of pain-related learning such as contingency rating. In contrast to the affective aspect of learning (measured by the change in valence ratings, Eq. 1), contingency captures the cognitive component of pain-related learning. While the relationship between contingency learning and the prediction of our model was significant ( $R^2 = 32\%$ ,  $p = 0.003$ ), we also have to take into account that contingency learning and valence learning were also highly correlated ( $R^2 = 17\%$ ,  $p = 0.038$ ). Therefore, to demonstrate that the model generalizes to contingency learning, we need to reject the null hypothesis of conditional independence between the prediction and contingency learning (given the association of both with differential valence rating). The specific statistical test was able to reject this null hypothesis with  $p = 0.05$ , meaning that the association between the prediction and contingency learning was not only significantly different from zero, but it was significantly higher than expected from the association between valence learning and contingency learning (Fig. 2 – Convergent validity, Table 2, see also Supplementary Notes: Model validators in the validation sample, Supplementary Table 6).

Similarly, the predictions of the RCPL signature were significantly associated with differential valence ratings (internally validated estimate:  $R^2 = 51\%$ ,  $p < 0.001$ , nested LOPO cross-validation) not only during the acquisition phase but also during extinction ( $R^2 = 18.6\%$ ,  $p = 0.03$ ) (i.e., the recovery of valence ratings after cessation of conditioning ( $CS^+$ -US pairing)). However, when testing for conditional independence between extinction and the model predictions, we found no evidence that this association was more than a consequence of the correlation between acquisition and extinction learning alone ( $R^2 = 46\%$ ,  $p = 0.86$ ; Fig. 2 – Convergent validity, Table 2, see also Supplementary Notes: Model validators in the validation sample, Supplementary Table 6).

The model predictions were also significantly correlated with valence learning induced by the aversive tone ( $R^2 = 18.9\%$ ,  $p = 0.03$ ). However, the specific statistical test suggested that this association could be explained as a secondary consequence of the association between pain-related learning and learning induced by the aversive tone ( $R^2 = 49\%$ ). Thus, our analysis could not provide evidence against the specificity of our model for pain-related learning. Note that if our model would primarily capture general aversive learning (i.e., neural mechanisms shared between pain and tone learning), its predictions would also be in conditional dependence with tone learning (Fig. 2 – Divergent validity, Table 2, see also Supplementary Notes: Model validators in the validation sample, Supplementary Table 6).

In all three datasets, we have obtained a comprehensive set of additional demographic and psychometric measures as well as measures of in-scanner motion (see all the variables in Supplementary Table 3). The partial



**Fig. 2 | Analysis of convergent validity, divergent validity, and confounder bias.**

Convergent validity: testing the generalizability of the model to the cognitive aspect of learning suggests that the RCPL signature captures pain-specific learning, as reflected in the rejection of conditional independence of contingency ratings ( $p = 0.05$ ). However, the model prediction has not generalized to extinction learning ( $p > 0.05$ ). Divergent validity: model prediction was tested against tone-related aversive learning. The model was not driven by tone learning, suggesting that the model primarily captures pain-related learning and not tone. Other divergent validators were also tested, including fear of pain/tone, pain catastrophizing, and pain anxiety, but none of them significantly biased the model. Confounding bias: none of the investigated confounder variables including age, in-scanner motion parameters (mean, median and maximum framewise displacement, percent of scrubbed volumes), and depression score biased our model (see Supplementary Table 3).

The thickness and the opacity of the lines are proportional to the  $R^2$  values. Percentages refer to the coefficient of determination  $R^2$  between the two variables.  $p < 0.05$  indicates that the model prediction is significantly associated with the variable over and above the baseline association between the observed pain-related differential valence ratings (DVC, Eq. 1) and the variable itself (e.g., contingency learning). The observed and predicted valence learnings were significantly correlated  $R^2 = 51\%$ ,  $p < 0.001$ . Conting. Learning –  $CS^+$ <sub>pain</sub> contingency after acquisition training, Extinction learning – differential valence rating in extinction learning (see the study preregistration for more detail), Tone learning – differential valence rating for tone (see Eq. 1), Fear – fear of pain (see “Questionnaires”), Motion – head motion during the resting-state fMRI (mean framewise displacement (FD)).

**Table 2 | Result of partial generalization test**

		$R^2$ with pain learning (%)	$R^2$ with model prediction (%)	Significance of partial generalization test ( $p$ )	
Convergent validity	Contingency $CS^+$ <sub>pain,acq</sub>	17.4	31.9	<b>0.05*</b>	
	Contingency $CS^+$ <sub>pain,ext</sub>	1.4	25.3	<b>0.03*</b>	
	Extinction learning <sub>pain</sub>	46.1	18.6	0.86	
Divergent validity	Tone learning	49.1	18.9	0.80	
	Extinction learning <sub>tone</sub>	33.5	7.5	0.92	
	Fear of pain	2.9	6.7	0.30	
	Fear of tone	0.5	0.0	0.96	
	Pain catastrophizing	0.0	0.0	0.91	
	Pain anxiety	PASS D1	0.3	0.0	0.99
		PASS D2	6.5	0.8	0.78
PASS D3		2.0	1.1	0.64	
PASS D4		0.4	1.1	0.62	
Confounder bias	Age	1.3	3.5	0.41	
	Motion	Mean FD	0	2.0	0.49
		Median FD	0.2	1.8	0.52
		Max FD	2.5	7.5	0.24
		Percent scrubbed	0.2	1.0	0.63
	Depression	0.5	0.7	0.69	

We tested convergent and divergent validators, and confounder variables for model prediction. See the main text for a detailed description of convergent, divergent validators, and confounder bias. Significant results in the partial generalization test are denoted with an asterisk and bold character.

confounder test<sup>46</sup> did not provide evidence for confounding bias with any of these variables (Fig. 2 – Confounding bias).

While our power analysis (Supplementary Fig. 5) suggests that larger samples are needed to reliably disentangle general and pain-specific mechanisms in aversive learning, our results provide a promising basis for further large-scale efforts to clarify the divergent validity of the proposed marker.

**Pain-related learning was predicted by a neuroscientifically plausible neural system**

Our machine learning procedure involved a data-driven feature selection step and identified 10 predictive connections (out of 7503), 5 of them being positively and 5 negatively associated with pain-related learning (see Table 3 and Fig. 3a). As previous studies report that, even for robust predictive models, model coefficients can be instable and unreliable<sup>47</sup>, we did not consider these as the primary measure for quantifying predictive importance. Instead, we have constructed a more reliable, externally validated importance measure of single connections and regions by repeating the whole model fitting procedure in the discovery dataset with a reduced feature space that excludes (i) a single connection or (ii) all connections of a single brain region. Predictive importance is then quantified as the loss of explained variance when testing these “truncated” models in the external validation samples (hereinafter referred to as: “exclusion loss”). As compared to the full model, larger exclusion loss indicates higher predictive importance of the left-out connection/region.

With this approach, the following connections were found to be most predictive to pain-related learning: amygdala – parieto-occipital associative area (exclusion loss: 10.68% of variance explained), posterior insula – left sensorimotor cortex (exclusion loss: 10.33%), angular gyrus – right anterior PFC (exclusion loss: 9.3%) and cerebellum (I–V) – posterior cingulum (exclusion loss: 5.81%) (Fig. 3c). In terms of regional importance (i.e., when excluding all connections of a brain region), the posterior cingulum (exclusion loss: 11.07%), the amygdala (exclusion loss: 9.98%), the posterior insula (exclusion loss: 9.44%), the dorsal PCC (exclusion loss: 7.65%), and the left sensorimotor region (exclusion loss: 7.84%) were found to be the most important predictors (Fig. 3b).

**Discussion**

Here we present the RCPL signature, a novel resting-state functional brain connectivity-based predictive model of individual pain-related learning

**Table 3 | Predictive connections of RCPL signature**

$\beta$ coef value	Region 1						Region 2						LOFO (%)
	Region name	BASC idx	LORO (%)	x	y	z	Region name	BASC idx	LORO (%)	x	y	z	
-79.3	ParOcc assoc	100	5.9	17.13	-76.26	25.61	Amy	71	9.98	3.06	-7.47	-19.9	<b>10.68*</b>
84.4	plns	21	9.44	-2.01	-7.26	-3.49	SensMot	8	7.84	-40.99	-22.62	61.9	<b>10.33*</b>
-92.7	Angular	50	5.56	-29.83	-72.86	31.94	r aPFC	92	6.87	42.57	42.79	10.85	<b>9.3*</b>
-80.8	Cer V	48	2.99	-0.4	-46.11	-23.17	pCing	19	11.07	-1.47	-63.33	8.28	<b>5.81*</b>
43.3	Cer VII	18	-0.09	1.9	-79.54	-48.01	Coll s	79	0.15	-4.4	-41.99	-12.66	1.57
48.4	iParietal	114	-0.45	-48.23	-65.34	33.46	Marg s	118	-1.71	-0.23	-30.3	47.49	0.77
-30.5	alns	85	2.53	10.94	22.89	4.96	Vis assoc	43	2.35	12.89	-83.54	18.77	-0.12
6.8	aPFC	87	2.21	1.64	62.77	13.2	STG	88	6.44	0.66	-5.24	-0.93	-0.63
-49.2	Put	25	6.03	-2.8	-2.14	4.11	Vis assoc	16	-2.66	2.23	-95.23	14.76	-2.21
38.6	dPCC	72	7.65	-1.71	-49.05	7.1	sParietal	76	-5.45	0.57	-25.7	73.98	-2.29

The table lists all predictive connections in the RCPL signature, including the  $\beta$  coefficient of the connections in the model (the positive (negative) sign of the  $\beta$  coefficient means that a higher connection between the two regions leads to higher (lower) differential valence rating), the involved regions, their coordinates, their index in the applied functional brain atlas (generated with the bootstrap analysis of stable clusters method (BASC)), and the exclusion loss in the leave-one-out-region (LORO) and leave-one-out-feature (LOFO) analysis. The table is sorted based on feature importance as measured by exclusion loss (the loss of explained variance in the external sample when excluding the feature during model construction). Significant drop in explained variance in the LOFO analysis is denoted with asterisk and bold character.

$\beta$  coef coefficient of  $\beta$  value in the model, *Region name* custom-defined name of the regions, *BASC idx* the index of the region from the applied BASC atlas, *LORO* exclusion loss in the leave-one-region-out approach (see section "Methods"), high values mean large drop in explained variance compared to the original model - x, y, z: "Center of Mass. We report the center of mass of the parcel in MNI coordinates. Note that the center of mass may lie outside the parcel for bilateral parcels<sup>45</sup>"; Supplementary Table 7 contains MNI coordinates of left and right parcels separately for bilateral parcels, *LOFO* exclusion loss in the leave-one-feature-approach (see section "Methods"), high values mean larger drop in explained variance, *ParOcc* parieto occipital association area, *Amy* amygdala, *plns* posterior insula, *SensMot* sensorimotor area, *Angular* angular region, *r aPFC* right anterior prefrontal cortex, *Cer V* cerebellar lobules I-V, *pCing* posterior cingulum, *Cer VII* cerebellar lobule VII, *Coll s* collateral sulcus, *iParietal* inferior parietal lobule, *Marg s* marginal sulcus, *alns* anterior insula, *Vis assoc* visual associative area, *aPFC* anterior prefrontal cortex, *STG* superior temporal gyrus, *Put* putamen, *dPCC* dorsal posterior cingulate cortex, *sParietal* superior parietal lobule.

\*Significant drop in explained variance.

performance. We discuss the relevance of the presented brain biomarker of pain-related learning for future research efforts and clinical applications, as well as the proposed model's externally validated predictive performance, specificity, and its neurobiological plausibility.

The RCPL signature offers a wealth of opportunities to advance clinical and translational pain research, both as a stand-alone tool to quantify pain-related learning, and as a potential component of future composite pain biomarkers<sup>10</sup>.

First, RCPL scores can be interpreted as the signature of an underlying neural mechanism that may be differentially expressed in different pain conditions. Using this signature as an easily quantifiable indicator could help to identify those at risk of chronification due to (maladaptive) pain-related learning.

Second, our brain signature may allow for a more in-depth mechanisms-based characterization of patients presenting with similar pain phenotypes. As with structural connections, the expression of functional links between brain regions varies widely between individuals, determining the extent to which they learn. Insights into the strength and integrity of these connections could therefore provide valuable information about an individual's capability for pain-related learning and serve as both a prognostic and diagnostic tool.

Third, the RCPL signature could be used to assess or predict treatment efficacy, as well as a method for patient stratification. Combined with improvements in subjective outcome measures, changes in the RCPL signature could indicate whether an intervention has effectively altered pain-related learning (as opposed to other, less relevant processes). As the RCPL signature is derived solely from resting-state functional connectivity, and therefore requires only a 10-min non-invasive scan without the use of noxious stimuli or time-consuming learning paradigms, it is a highly accessible tool for these clinical and translational purposes.

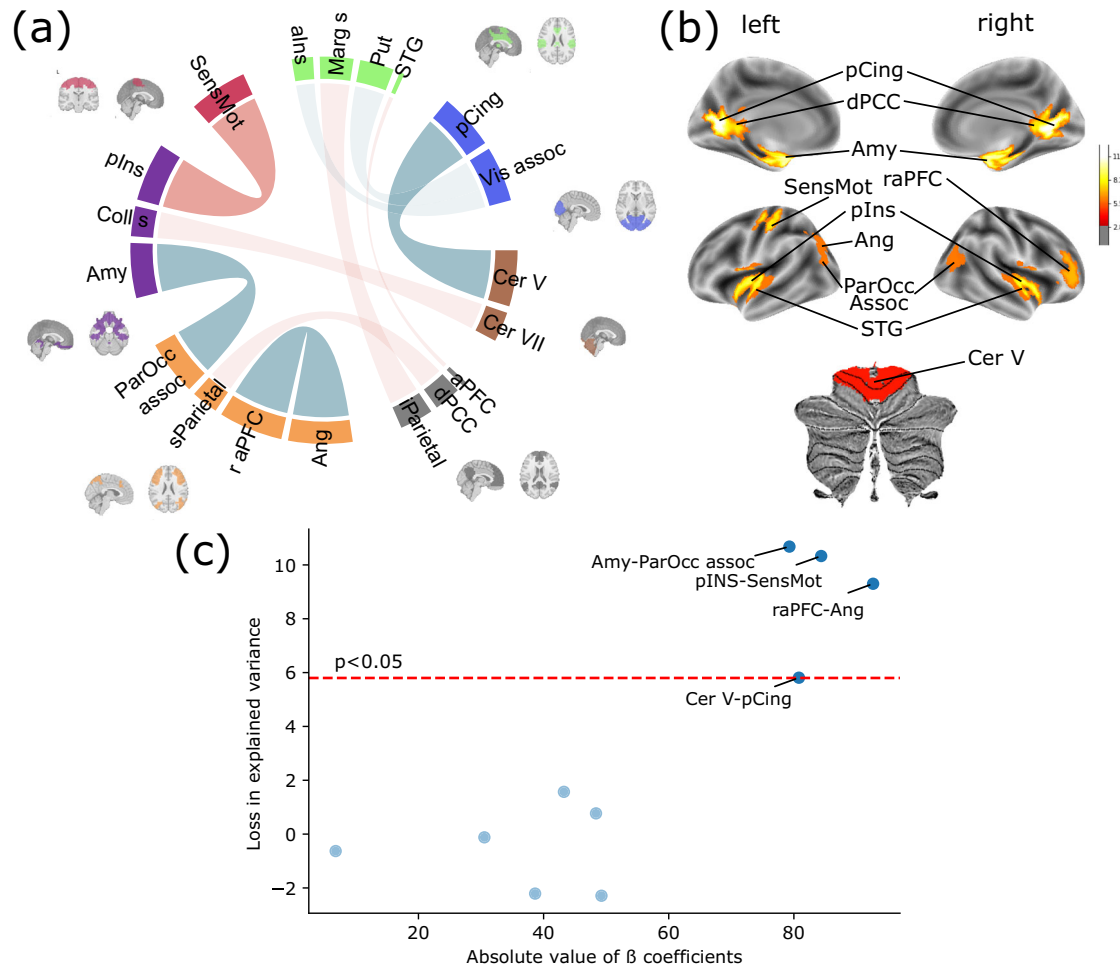
While the development of pain-related brain biomarkers has been widely hailed as an important step<sup>10</sup>, a thorough evaluation of their predictive performance, specificity, and confounders is warranted. In our pre-registered external validation, the RCPL signature was found to explain 8–12% of the variance in pain-related learning. This is a substantial predictive effect size, both according to Cohen's rule of thumb<sup>48</sup> and as

compared to typical effect sizes in multivariate brain-wide association studies. For example, 90% of the psychometric traits in the Adolescent Brain Cognitive Development dataset show weaker predictability and the proposed model is on par with the top 10% phenotypes in the same dataset<sup>43</sup>.

While large effect sizes are desirable for a biomarker for individual predictions in clinical settings, effect sizes that are comparable to, or considerably lower than the hereby reported effect size, proved to be useful in the case of polygenic scores<sup>49–51</sup>. For example, in studies assessing inter-individual variance in cognitive performance, genetic data typically explain around 7–10% of the variance, even with data from more than a million individuals<sup>52</sup>. With effect sizes, similar to that of the proposed model, polygenic risk scores are commonly deployed as research tools to explore the genetic architecture of complex traits and diseases, to uncover new genetic associations, to identify novel pathways, and to gain a deeper understanding of the underlying biology<sup>53</sup>. Similarly, investigating the "expression" of the identified brain network signature of pain-related learning in various experimental and clinical pain conditions provides a promising avenue for further research, with the potential to inform the development of novel, personalized treatment approaches.

Of note, we observed a remarkable difference between internally and externally validated estimates of prediction performance. The reasons may be twofold. First, in small-sample model discovery, cross-validation is known to provide unbiased, but highly variable estimates<sup>38,39</sup>. Second, in the case of complex preprocessing and feature engineering workflows, it is often not feasible to include all free parameters of the model in the nested cross-validation procedure<sup>37,38</sup>. Seemingly "innocent" adjustments of such workflows during model discovery can lead to biased internal performance estimates. Our study shows that both pitfalls can be overcome with pre-registered external validation; a technique that should become standard practice, in order to unlock small-sample model discovery and accelerate the discovery of robust and replicable multivariate brain-behavior associations<sup>43,44</sup>.

The model was trained to predict pain-related valence learning. However, that does not automatically imply that the model is specific to pain-related valence learning as the predictions may be driven by more



**Fig. 3 | Resting-state functional connectivity predictors of pain-related learning.** **a** Chord plot of all 10 predictive connections between 20 regions. The sign of the  $\beta$ -coefficient is color coded, red (blue) indicates a positive (negative) sign, i.e., higher connections between the two regions lead to higher (lower) differential valence rating. The thickness of the chords is proportional to the absolute magnitude of the  $\beta$ -coefficient. Regions are grouped according to their parent network based on the seven parcellations of the BASC atlas. Red – somatomotor network; green – ventral attention + salience network + basal ganglia + thalamus; blue – visual network; brown – cerebellar network; gray – default mode network; orange – fronto-parietal network + visual downstream; purple – mesolimbic network. **b** Exclusion loss after leave-one-region-out (LORO). Leaving out some regions, such as the amygdala or posterior insula leads to a large decrease in the explained variance of the model when

tested in the external validation datasets. Only cortical regions with significant ( $p < 0.05$ ) exclusion loss are visualized. **c** Contrasting the association between the predictive connections'  $\beta$ -coefficients (the standard technique for assessing predictive performance) and exclusion loss (the proposed more robust measure of importance) in the leave-one-feature-out analysis, shows that six out of ten predictive connectivities have little or no effect on the external model performance. However, the four most important connections, on the other hand, clearly stand out from the rest; excluding them during model training results in models that explain 6–10% less variance in the external samples. The red horizontal dashed line represents the significance threshold for the exclusion loss (5.8%). Abbreviations of the regions can be found in the footnote of Table 3.

general aspects of aversive learning, with shared underlying mechanisms<sup>28,54,55</sup>, that generalize across modalities (e.g., aversive tone-related learning) and across emotional and cognitive aspects (e.g., contingency ratings). Therefore, the specificity of the model was additionally tested against these other aspects of aversive learning. We found that the RCPL does significantly generalize to (i.e., explained more variance than expected in, convergent validity) the cognitive aspect of pain-related learning (contingency ratings). On the other hand, we found no evidence for “generalization” to aversive tone-related learning (divergent validity).

The absence of bias in the predictions toward potential confounding variables is also a critical factor in determining the clinical relevance and translational potential of brain-based predictive models<sup>46</sup>. For instance, psychological factors such as anxiety and depression are known to be associated with aversive learning<sup>56</sup>. In MRI-based predictive models, motion artifacts are also known to be problematic, especially when the variable of interest is also correlated with motion in the scanner<sup>57,58</sup>. Our confounder analysis found no evidence that the predictions of the proposed model are

biased by variables of no interest, such as in-scanner motion, demographics, or depression, making the RCPL signature as a promising candidate for further independent validation.

Using a machine learning approach with inherent feature selection functionality, we identified a set of ten resting-state connections that together predicted pain-related learning (Table 3). Five of these connections were positively associated with learning, while the remaining five were negatively associated. As recent work has highlighted potential reliability issues with feature weights from predictive models<sup>47</sup>, we retrained models with feature exclusion to obtain more reliable, externally validated importance estimates. We found that 4 of the 10 connections and 9 out of the 20 involved brain areas had statistically significant external validity. Most of the identified predictors are known to play a critical role in aversive learning, pain-related neural processes, or both.

First, mesolimbic regions, and the amygdala in particular, make the largest contribution to the predictive capabilities of the RCPL signature. Connectivity between the amygdala and the frontal part of the central executive network

(CEN) (which is also known as frontoparietal network<sup>59</sup>) is one of the most prevalent findings in aversive learning and is commonly associated with behavioral learning parameters (US expectancy, fear ratings)<sup>31,33</sup>. Although, unlike our study, these studies examined resting-state connectivity changes induced by learning (as opposed to how baseline connectivity shapes subsequent learning performance), they confirm the important role of these regions in the underlying neural processes. Our results substantially contribute to these findings by highlighting the critical role of the connection between the amygdala and the parietal part of the CEN (occipitoparietal association area), a network widely recognized for its involvement in pain and perception more generally<sup>59</sup>. The amygdala has been reported to directly mediate aspects of fear learning and to facilitate fear memory operations in other regions, including the hippocampus and prefrontal cortex<sup>60</sup>. Amygdala–CEN connectivity has been found to be increased in patients with chronic back pain<sup>61</sup>, in whom the evaluative aspect of threat learning has been reported to be impaired<sup>62</sup>.

Second, the strong contribution of posterior insula–sensorimotor connection to the model may be a key component of our model's apparent specificity for pain-related learning. Modality-specific threat cues have been reported to have shared and distinct biological components<sup>28,29</sup> and there is recent evidence for the specificity of the posterior insula for pain-related learning as compared to tone<sup>63</sup>. The insula has long been proposed as a multisensory integration site in pain<sup>64,65</sup>. Our study extends these findings and suggests that, in addition to the well-documented role of the insula in processing acute pain, pain-related learning features are also encoded in the interplay between sensory and posterior insular regions during spontaneous resting-state activity.

Third, we found that connectivity between the anterior prefrontal cortex and the angular region was negatively associated with pain-related learning. The anterior prefrontal cortex is part of the CEN and has been implicated in various aspects of pain processing<sup>59</sup>. Among other processes, it has been reported to predict cognitive modulation of pain<sup>66</sup>, and anterior prefrontal and angular regions (very similar to our findings) were found to be deactivated during fear conditioning<sup>24</sup> and activated during placebo analgesia in a large individual participant meta-analysis<sup>67</sup>.

Finally, the cerebellum is known to be involved in fear conditioning behavior<sup>68–70</sup> and there is now also accumulating evidence for a direct involvement of the cerebellum in pain modulation<sup>67,71</sup>. However, the cerebellum is often overlooked in related research. Here we found that omitting cerebellar connectivity significantly reduced predictive performance (~6% less variance explained). These results provide additional evidence that the commonly reported pain- and learning-associated cerebellar activity is not a secondary consequence of the experimental paradigms in these studies (e.g., suppression of motor responses) but it has a unique predictive contribution to the underlying mechanisms that is not redundant with activity and connectivity in other parts of the brain.

While the involvement of these connections in pain-related learning is plausible, further studies are needed to elucidate their specific functional significance. For example, given that both the posterior insula and the sensorimotor area have primarily been implicated in the processing of sensory rather than cognitive-affective information<sup>72,73</sup>, their contribution to the RCPL signature suggests a prominent role for sensory processing in pain-related learning. Whether this is indeed the case, needs to be investigated in further studies.

Moreover, the hippocampal formation, that consistently shows increased signal levels during pain-related learning, is represented only by the collateral sulcus (a region tightly linked to the hippocampus<sup>74</sup>), and analysis of its importance in the external validation sample was inconclusive. Therefore, further investigation of its contribution as a unitary region versus as part of an extended network of functionally connected brain areas is needed for a more detailed understanding of its role.

The low reliability of self-reported outcome measures can be an important limitation in terms of the power and reliability of brain-based predictive models<sup>75,76</sup>. This may also apply to the model presented here, which was trained to predict pain-related learning indexed as differential valence ratings, i.e., based on subjective self-reports. Thus, the reported

predictive performance is likely to be lower than what could optimally be achieved with more reliable target measures.

Our study used a small discovery sample combined with pre-registered external validation. While this approach allows for an efficient and reliable construction of brain-based predictive models<sup>37</sup>, our power calculation (Supplementary Fig. 5) indicated that moderate to large samples<sup>44</sup> will be needed to thoroughly characterize the model in terms of its specificity to pain-related mechanisms, its relationship to other aspects of pain-related learning (e.g., resistance to extinction), and its validity in clinical populations.

Disentangling the common and distinct mechanisms of pain- and tone-related learning is further limited in the present study by their potential interactions in our paradigm, through transfer learning and cross-modality generalization.

Like any fMRI-based study, our study is also limited by low SNR in regions prone to artifacts (e.g., susceptibility artifacts), including regions that are known to be involved in pain-related learning, such as the orbitofrontal cortex. Results relating to these regions should therefore be interpreted with caution.

Beyond the presented neural correlates of individual differences in pain-related learning, the RCPL signature holds potential as a reliable research tool for objectively characterizing such differences in future studies. By employing the RCPL signature in large and diverse populations, including individuals with different chronic pain conditions, we can deepen our understanding of the role of learning processes across different contexts. To facilitate the broader utilization of the RCPL signature in translational and clinical investigations, we offer a containerized implementation of our model. This implementation allows for predictions of individual pain-related learning performance from any resting-state fMRI dataset in a single step ([https://github.com/kincsesbalint/paintone\\_rsn](https://github.com/kincsesbalint/paintone_rsn))<sup>77</sup>.

In this study, we developed and externally validated a novel brain connectivity-based biomarker candidate for individual differences in pain-related learning. Using an unbiased performance evaluation approach, we demonstrate substantial predictive performance of the proposed RCPL signature with potential clinical relevance. Initial assessments of convergent and divergent validity indicate that the model may generalize to different measures of pain-related learning, but not to aversive tone, fear of pain, or other potential confounders. The model's predictions are based on predictors with high neurobiological plausibility in light of previous studies. Together, this makes the presented model a robust and highly accessible biomarker candidate for clinical and translational pain research, with promising implications for personalized treatment approaches.

## Methods

### Study design

We aimed to train a predictive model based on individual resting-state connectivity which predicts individual differences in a subsequent differential conditioning paradigm. To address previously reported<sup>38</sup> challenges in terms of reproducibility, credibility, and robustness of predictive models, the project was preregistered (<https://osf.io/b8znd/>) after the discovery study was conducted but prior to the collection of the validation studies (see Participants and samples). The preregistration did not only fix the details of the MRI preprocessing pipeline, the limits of in-scanner motion parameters for exclusion, the differential conditioning paradigm, and the preprocessing of the behavior data (see below) for validation sample 1 but also the results of the model discovery phase (hyperparameter values, model coefficients, model pipeline). For the complete document and full details, visit the study OSF page (<https://osf.io/b8znd/>).

### Participants and samples

All three studies involved young healthy individuals with no known neurological or psychiatric conditions. Exclusion was based on pre-registered criteria, such as missing behavior or MRI data, incidental findings in MRI, falling asleep during the resting-state measurement (self-report), and in-scanner motion limits (see "MRI preprocessing"). For additional details on

exclusion see the preregistration (<https://osf.io/b8znd/>) and the previously published manuscripts on the behavior and task fMRI results<sup>29,78,79</sup>. In all studies, the resting-state paradigm preceded the behavior paradigm.

**Discovery sample.** Thirty-eight participants were measured with a resting-state paradigm before they subsequently performed the behavior paradigm (see “Behavior paradigm”) inside the scanner. After exclusion, we used data from 25 participants for model training.

**Validation sample 1.** Thirty-three participants were first measured with a resting-state paradigm and the participants performed the behavior paradigm on a separate day out-of-scanner (median 3 days later, range: 1–5 days) (see Supplementary Table 1). After exclusion, 26 participants’ data were used.

**Validation sample 2.** An independent study group shared 28 participants’ data. This study (study 3) was acquired by different personnel and the participants performed a different paradigm in-scanner right after the resting-state paradigm (see “Behavioral paradigm”). After exclusion, 23 participants’ data were used.

The behavior and task-related fMRI data from the discovery sample and validation sample 2 have been previously published<sup>29,78,79</sup>. Validation sample 1 has not been published before.

The study was conducted in accordance with the Declaration of Helsinki and had been approved by the local Ethics Committee (University of Duisburg-Essen, Germany). All participants provided written informed consent. All ethical regulations relevant to human research participants were followed.

### MRI acquisition

All three studies were conducted using a 3T Siemens MAGNETOM Skyra MRI scanner. A high-resolution anatomical image (MPRAGE sequence, parameters in the discovery study and validation sample 1: repetition time (TR): 2300 ms, echo time (TE): 2.07 ms, inversion time (IR): 900 ms, flip angle (FA): 9°, resolution 1 × 1 × 1 mm; validation sample 2: TR: 1900 ms, TE: 2.13 ms, IR: 900 ms, FA: 9°, resolution: 1 × 1 × 1 mm) and an open-eye resting-state functional image (rs-fMRI) were collected with an EPI sequence (discovery study and validation sample 1: 38 slices, resolution 2.5 × 2.5 × 3 mm, TE: 28.0 ms, TR: 2300 ms, GRAPPA: 2, distance factor 15%, 260 volumes; validation sample 2: 46 slices, 3 × 3 × 3 mm, TE: 30 ms, TR: 2500 ms, GRAPPA: 3, 192 volumes) (Supplementary Table 1).

### Behavioral paradigm

All participants performed a differential conditioning paradigm including habituation, an acquisition training phase, and an extinction training phase.

**Discovery sample.** The details of the task are described in the preregistration (<https://osf.io/b8znd/>) and in ref. 78. Most importantly, geometric figures were used as CS to predict the delivery (CS<sup>+</sup><sub>pain</sub> or CS<sup>+</sup><sub>tone</sub>) or the absence (CS<sup>-</sup>) of two different US (US<sub>pain</sub> and US<sub>tone</sub>). In the habituation phase, the CS and US were presented/delivered once with a duration of 9 and 2.5 s, respectively. During the acquisition phase, 12 out of 16 CS<sup>+</sup> were followed by a US (75% reinforcement rate) (Supplementary Fig. 1). In the reinforced trials, 8 s after CS presentation the US was delivered (1 s overlap). In the extinction phase, 12 CS of each condition were presented but no US were applied. The intertrial interval ranged between 6 and 11 s. Heat pain stimuli and unpleasantness-matched auditory stimuli (see “Stimuli and matching procedure” below) were individually calibrated and served as US (US<sub>pain</sub> and US<sub>tone</sub>). The participants were informed about the possible association between the CS and US. We measured the following parameters on a 0–100 visual analog scale (VAS): valence ratings (verbal anchors: 0 = “very pleasant”, 50 = “neutral”, 100 = “very unpleasant”, once during habituation, four times during acquisition, and three times during extinction), contingency ratings after acquisition and extinction (verbal anchors: 0 = “100%

auditory stimulus”, 50 = “0%”, 100 = “100% painful stimulus”), US unpleasantness ratings (verbal anchors: 0 = “not unpleasant at all” and 100 = “unbearably unpleasant”, once during habituation and four times during acquisition), as well as fear of pain and tone before the experiment. The extinction phase followed the acquisition phase, in which the CSs were not associated with any US.

**Validation sample 1.** The same behavior paradigm was used as in the discovery sample.

**Validation sample 2.** Participants underwent a similar differential conditioning paradigm with two US (75% reinforcement rate), receiving visceral instead of somatic heat pain (Supplementary Table 1). The details of the paradigm can be found in refs. 29,79. During acquisition, 9 out of 12 CS were paired with US. In the reinforced trials, the 14 s long US was preceded by a CS for 6–12 s and terminated together. The intertrial interval was 8 s long. Participants provided CS valence ratings before, in the middle, and after acquisition on a 0–100 VAS (verbal anchors: “not unpleasant at all” and “very unpleasant”). As the anchors of the used scale differed from the one in the discovery study, we corrected for that (dividing the raw values by 2 and adding 50 to them), therefore, the two scales are comparable. Contingency ratings were assessed with a 0–100 VAS (representing the probability of the CS is followed by the specific US).

We derived the pre-registered differential valence change (DVC) from the valence ratings (Eq. 1) in which each element stands for the following: CS<sup>+</sup><sub>pain,Acq</sub> – is the valence rating of the CS predicting pain given at the end of acquisition training; CS<sup>+</sup><sub>pain,Hab</sub> – is the valence rating of the CS pain given in the habituation phase (before acquisition training); CS<sup>-</sup><sub>Acq</sub> – is the valence rating of the CS predicting US omission (not associated with any US) given at the end of acquisition training; CS<sup>-</sup><sub>Hab</sub> – is the valence rating of the CS predicting US omission (not associated with any US) given in the habituation phase (before acquisition). DVC in the experiment is meant to show the evaluative aspect of learning. A higher value refers to higher learning. We used this value as our main outcome.

$$\text{differential valence change} = \left( \text{CS}_{\text{pain,Acq}}^{+} - \text{CS}_{\text{pain,Hab}}^{+} \right) - \left( \text{CS}_{\text{Acq}}^{-} - \text{CS}_{\text{Hab}}^{-} \right) \quad (1)$$

### Stimuli and matching procedure

In the discovery sample and in the validation sample 1, the task was implemented in the software *Presentation* (<https://www.neurobs.com/>). Visual geometric stimuli were used as conditioned stimuli (RGB code: 142, 180, 227) on a black background. Heat pain stimulation was delivered with a thermode (Pathway System, CHEPS thermode, MEDOC, Israel, <http://www.medoc-web.com>, baseline temperature 35° and heating, cooling rise and fall rates of 70 and 40°/s) to the left volar forearm for 2.5 s. Individual temperature levels with an unpleasantness rating of VAS70 on a 0–100 VAS were determined using a calibration procedure and used subsequently (see Supplementary Methods: Calibration of aversive stimuli and ref. 78 for details). The maximum temperature of 50° was used as a safety limit to avoid potential tissue damage. An auditory stimulus was generated using the software *Audacity* software (v: 1.3.10-beta; <http://www.audacity.sourceforge.net/>) with the following parameters: sawtooth waveform profile, duration: 2.5 s, fade-in: 180 ms, fade-out: 300 ms. The volume was adjusted to match the unpleasantness of the pain stimulus in a separate calibration phase as follows. Immediately after the application of the heat stimulus, the auditory stimulus was delivered. Participants had to indicate whether the tone was less unpleasant, more unpleasant, or as unpleasant as the heat stimulus. Based on their response, the volume of the tone was automatically adjusted, but the procedure was stopped if both types of stimuli were equally unpleasant. This was repeated five times, and the mean volume level was used. As a final check after the matching procedure, an additional test was performed, in which each pain and tone stimulus was delivered three times and their unpleasantness was rated. If there was a considerable discrepancy,

the volume and/or the temperature were adjusted to ensure that stimuli were equally unpleasant.

In validation sample 2, similar geometric figures were used as CS. However, instead of the heat pain stimulation, visceral pain was used as US (see “Behavioral paradigm” above). The pressure-controlled rectal distension was elicited with a barostat system (ISOBAR3 device, G&J Electronics, Toronto, ON, Canada). The individually calibrated pressure for visceral pain was determined with a rating between 60 and 80 on a VAS100 scale (see Supplementary Material and in refs. 29,79,80). A similar system as in the other two studies (*Audicity* software) was used to elicit the auditory signal with the same parameters. A similar matching procedure ensured that the pain and unpleasant tone stimuli were matched in unpleasantness (see details above). The intensity of the unpleasant tone was adjusted until unpleasantness ratings of the two stimuli were comparable in two consecutive trials.

### Questionnaires

To assess individual psychological state and trait characteristics, participants filled different questionnaires regarding depression (ADS-K<sup>81</sup>), pain-related anxiety (Pain Anxiety Symptoms Scale – PASS-D<sup>82</sup>), and pain catastrophizing (Pain Catastrophizing Scale<sup>83</sup>). Additionally, participants provided fear ratings of pain and tone stimuli before the experiment on 0–100 VAS (verbal anchors: 0 = “not afraid at all” and 100 = “extremely afraid”). We calculated the difference between the two ratings, to further characterize individual differences.

### MRI preprocessing

All acquired resting-state data were preprocessed with our pre-registered, in-house built pipeline: the RPN-signature pipeline (version 0.2.6)<sup>6</sup>. The docker version of the pipeline can be found here (<https://spisakt.github.io/RPN-signature/>). The preprocessing utilized various tools from the *FSL*<sup>84</sup>, *AFNI*<sup>85</sup>, and *ANTS*<sup>86</sup> software libraries and employed *nipype*<sup>87</sup> to perform the following steps: brain extraction, boundary-based co-registration from functional to anatomical space, high precision co-registration from anatomical to standard space, motion correction, nuisance regression (24 motion regressors<sup>88</sup> and temporal SNR-based *compcor*<sup>89</sup> correction with 5 components), bandpass filtering (0.08–0.008 Hz) and scrubbing of high motion time frames (framewise displacement,  $FD > 0.15$ ). We visually checked all the brain extraction and registration results. The brain masks were manually corrected in case of error. A functional brain atlas (data-driven parcellation derived with the BASC algorithm)<sup>45</sup> was used with 122 non-overlapping functional regions (the used atlas is available on the project GitHub page [https://github.com/kincsesbalint/paintone\\_rsn/tree/master/data\\_in](https://github.com/kincsesbalint/paintone_rsn/tree/master/data_in)) and additionally, the global gray matter signal was also kept. The atlas regions were masked with gray matter and transformed back to the functional space. This approach is expected to decrease the heterogeneity introduced by registration across subjects and ensures that the resulting signal derives from gray matter voxels. An example region of interest of the amygdala is visualized in Supplementary Fig. 6. The individual connectivity matrices were calculated via partial correlation. We excluded participants based on our pre-registered criteria, namely, mean  $FD > 0.15$  and percent of scrubbed volumes  $> 25\%$ . Based on these motion criteria, we excluded  $n = 11$ ,  $n = 7$ , and  $n = 5$  participants from the discovery sample, validation sample 1, and validation sample 2, respectively.

### Statistics and reproducibility

**Model discovery.** The timeseries of the regions were standardized and used to calculate individual partial correlation matrices resulting in 7503 functional connections. Our machine learning pipeline consisted of a *k*-best feature selection and a regularized linear model with L2 penalty (Ridge), executed in a cross-validation compliant manner similar to previous studies<sup>6,7</sup>. We used a nested leave-one-out cross-validation scheme for model training. The inner loop was used to optimize hyperparameters (*k*: number of selected features and alpha: regularization parameter) and the outer loop was to estimate model

performance in the discovery phase. We fit a final model on the whole discovery sample with the identified best hyperparameters at the end of model training, to be tested in the external validation study. Predictive performance was estimated with the explained variance and the correlation between the observed and predicted values in the outer cv-loop. For significance testing, we used a nonparametric permutation test. Exact *p* values were reported,  $p < 0.05$  was considered as significant.

**External validation.** Following recommendations in predictive model development<sup>41</sup>, we tested the model generalizability on the external validation sample 1 and validation sample 2 (external validity). The preprocessed connectivity matrices from validation sample 1 and validation sample 2 were fed into the developed model and model predictions were correlated with the observed DVC values. The association between the RCPL prediction and the observed values was estimated with correlation coefficients (one-tailed) for the two samples separately and for the pooled sample as well. Exact *p* values were reported,  $p < 0.05$  was considered as significant.

**Analysis of validators and confounder.** To test the effect of convergent, divergent validators, and confounders on model predictions, we evaluated conditional independence with the partial confounder test from the *mlconfound* software package<sup>46</sup>. Briefly, the correlation between the predicted value and the convergent, divergent validator or the confounder was tested given the correlation between the convergent, divergent validator or the confounder and the outcome variable. We tested against the null hypothesis that the model is not biased by the third variable (convergent validator, divergent validator, confounder). Convergent validators were contingency ratings after acquisition and extinction of CS<sup>+</sup><sub>pain</sub> and extinction learning. Divergent validators were tone-related learning (in acquisition and extinction), fear of pain, and pain anxiety. We investigated the following confounders: age, scores on depression questionnaires, and in-scanner motion (mean  $FD$ , max  $FD$ , median  $FD$ , percent of scrubbed volumes). Exact *p* values were reported,  $p < 0.05$  was considered as significant.

**Predictive importance.** To investigate the impact of brain connections and regions on the model performance, we used an externally validated leave-one-feature-out (LOFO) and a leave-one-region-out (LORO) approach. In detail, we excluded a single connection or all connections of a region (for LOFO and LORO, respectively) and retrained a model on the remaining subset of connections. Afterward, we compared the retrained model’s explained variance to the original model’s explained variance in the external validation samples. High values mean a drop in model performance estimation and indicate that the excluded connection/region has a high impact on the predictions.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Preprocessed fMRI timeseries data and behavior data are available on the project’s GitHub page ([https://github.com/kincsesbalint/paintone\\_rsn](https://github.com/kincsesbalint/paintone_rsn))<sup>77</sup>. The raw MRI data are deposited on the project OSF page (<https://osf.io/b8znd/files/osfstorage>).

### Code availability

All codes are available on the project’s GitHub page ([https://github.com/kincsesbalint/paintone\\_rsn](https://github.com/kincsesbalint/paintone_rsn))<sup>77</sup>.

Received: 9 October 2023; Accepted: 10 July 2024;

Published online: 17 July 2024

## References

- Kuner, R. & Flor, H. Structural plasticity and reorganisation in chronic pain. *Nat. Rev. Neurosci.* **18**, 20–30 (2016).
- Meints, S. M. & Edwards, R. R. Evaluating psychosocial contributions to chronic pain outcomes. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **87**, 168–182 (2018).
- Naylor, B., Boag, S. & Gustin, S. M. New evidence for a pain personality? A critical review of the last 120 years of pain and personality. *Scand. J. Pain* **17**, 58–67 (2017).
- Wager, T. D. et al. An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* **368**, 1388–1397 (2013).
- Woo, C. W. et al. Quantifying cerebral contributions to pain beyond nociception. *Nat. Commun.* **8**, 14211 (2017).
- Spisak, T. et al. Pain-free resting-state functional brain connectivity predicts individual pain sensitivity. *Nat. Commun.* **11**, 1–12 (2020).
- Kotikalapudi, R. et al. Brain morphology predicts individual sensitivity to pain: a multicenter machine learning approach. <https://doi.org/10.1097/j.pain.0000000000002958> (2023).
- Lee, J. J. et al. A neuroimaging biomarker for sustained experimental and clinical pain. *Nat. Med.* **27**, 174–182 (2021).
- Vachon-Presseau, E. et al. Identification of traits and functional connectivity-based neurotraits of chronic pain. *PLoS Biol.* **17**, e3000349 (2019).
- Tracey, I., Woolf, C. J. & Andrews, N. A. Composite pain biomarker signatures for objective assessment and effective treatment. *Neuron* **101**, 783–800 (2019).
- Seymour, B. Pain: a precision signal for reinforcement learning and control. *Neuron* **101**, 1029–1041 (2019).
- Zhang, S., Mano, H., Ganesh, G., Robbins, T. & Seymour, B. Dissociable learning processes underlie human pain conditioning. *Curr. Biol.* **26**, 52–58 (2016).
- LeDoux, J. E. Coming to terms with fear. *Proc. Natl Acad. Sci. USA* **111**, 2871–2878 (2014).
- Gerber, B. et al. Pain-relief learning in flies, rats, and man: Basic research and applied perspectives. *Learn. Mem.* **21**, 232–252 (2014).
- Wager, T. D. & Atlas, L. Y. The neuroscience of placebo effects: connecting context, learning and health. *Nat. Rev. Neurosci.* **16**, 403–418 (2015).
- Mansour, A. R., Farmer, M. A., Baliki, M. N. & Apkarian, A. V. Chronic pain: the role of learning and brain plasticity. *Restor. Neurol. Neurosci.* **32**, 129–139 (2014).
- Löffler, M. et al. Corticostriatal circuits in the transition to chronic back pain: the predictive role of reward learning. *Cell Rep. Med.* **3**, 100677 (2022).
- Winkelmann, T. et al. Brain morphology correlates of interindividual differences in conditioned fear acquisition and extinction learning. *Brain Struct. Funct.* **221**, 1927–1937 (2016).
- Carter, R. M. K., O'Doherty, J. P., Seymour, B., Koch, C. & Dolan, R. J. Contingency awareness in human aversive conditioning involves the middle frontal gyrus. *Neuroimage* **29**, 1007–1012 (2006).
- MacNamara, A. et al. Neural correlates of individual differences in fear learning. *Behav. Brain Res.* **287**, 34–41 (2015).
- Büchel, C., Dolan, R. J., Armony, J. L. & Friston, K. J. Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. *J. Neurosci.* **19**, 10869–10876 (1999).
- Van Well, S., Visser, R. M., Scholte, H. S. & Kindt, M. Neural substrates of individual differences in human fear learning: evidence from concurrent fMRI, fear-potentiated startle, and US-expectancy data. *Cogn. Affect. Behav. Neurosci.* **12**, 499–512 (2012).
- Pohlack, S. T. et al. Hippocampal but not amygdalar volume affects contextual fear conditioning in humans. *Hum. Brain Mapp.* **33**, 478–488 (2012).
- Fullana, M. A. et al. Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Mol. Psychiatry* **21**, 500–508 (2016).
- Milad, M. R. et al. A role for the human dorsal anterior cingulate cortex in fear expression. *Biol. Psychiatry* **62**, 1191–1194 (2007).
- Hartley, C. A., Fischl, B. & Phelps, E. A. Brain structure correlates of individual differences in the acquisition and inhibition of conditioned fear. *Cereb. Cortex* **21**, 1954–1962 (2011).
- Utz, A. et al. Cerebellar vermis contributes to the extinction of conditioned fear. *Neurosci. Lett.* **604**, 173–177 (2015).
- Biggs, E. E. et al. The neural correlates of pain-related fear: a meta-analysis comparing fear conditioning studies using painful and non-painful stimuli. *Neurosci. Biobehav. Rev.* **119**, 52–65 (2020).
- Koenen, L. R. et al. Associative learning and extinction of conditioned threat predictors across sensory modalities. *Commun. Biol.* **4**, 553 (2021).
- Cooper, S. E., Dunsmoor, J. E., Koval, K. A., Pino, E. R. & Steinman, S. A. Test-retest reliability of human threat conditioning and generalization across a 1-to-2-week interval. *Psychophysiology* <https://doi.org/10.1111/psyp.14242> (2022).
- Schultz, D. H., Balderston, N. L. & Helmstetter, F. J. Resting-state connectivity of the amygdala is altered following pavlovian fear conditioning. *Front. Hum. Neurosci.* **6**, 242 (2012).
- Linnman, C., Zeidan, M. A., Pitman, R. K. & Milad, M. R. Resting cerebral metabolism correlates with skin conductance and functional brain activation during fear conditioning. *Biol. Psychol.* **89**, 450–459 (2012).
- Feng, P., Feng, T., Chen, Z. & Lei, X. Memory consolidation of fear conditioning: bi-stable amygdala connectivity with dorsal anterior cingulate and medial prefrontal cortex. *Soc. Cogn. Affect. Neurosci.* **9**, 1730–1737 (2014).
- Feng, P., Zheng, Y. & Feng, T. Spontaneous brain activity following fear reminder of fear conditioning by using resting-state functional MRI. *Sci. Rep.* **5**, 16701 (2015).
- Fraenz, C. et al. Fear learning sculpts functional brain connectivity at rest beyond the traditional fear network in humans. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.05.26.115840> (2020).
- Martynova, O. et al. Longitudinal changes of resting-state functional connectivity of amygdala following fear learning and extinction. *Int. J. Psychophysiol.* **149**, 15–24 (2020).
- Gallitto, G. et al. External validation of machine learning models with adaptive sample splitting. Preprint at *BioRxiv* <https://doi.org/10.1101/2023.12.01.569626> (2023).
- Varoquaux, G. Cross-validation failure: small sample sizes lead to large error bars. *Neuroimage* **180**, 68–77 (2018).
- Varoquaux, G. & Cheplygina, V. Machine learning for medical imaging: methodological failures and recommendations for the future. *NPJ Digit. Med.* **5**, 48 (2022).
- Chyzyk, D., Varoquaux, G., Milham, M. & Thirion, B. How to remove or control confounds in predictive models, with applications to brain biomarkers. *Gigascience* **11**, giac014 (2022).
- Woo, C. W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging. *Nat. Neurosci.* **20**, 365–377 (2017).
- Marek, S. et al. Reproducible brain-wide association studies require thousands of individuals. *Nature* **603**, 654–660 (2022).
- Tervo-Clemmens, B. et al. Reply to: multivariate BWAS can be replicable with moderate sample sizes. *Nature* **615**, E8–E12 (2023).
- Spisak, T., Bingel, U. & Wager, T. D. Multivariate BWAS can be replicable with moderate sample sizes. *Nature* **615**, E4–E7 (2023).
- Urchs, S. et al. MIST: a multi-resolution parcellation of functional brain networks. *MNI Open Res.* **1**, 3 (2019).
- Spisak, T. Statistical quantification of confounding bias in machine learning models. *Gigascience* **11**, giac082 (2022).
- Tian, Y. & Zalesky, A. Machine learning prediction of cognition from functional connectivity: are feature weights reliable? *Neuroimage* **245**, 118648 (2021).
- Muller, K. & Cohen, J. Statistical power analysis for the behavioral sciences. *Technometrics* **31**, 499–500 (1989).

49. Ikeda, M., Saito, T., Kanazawa, T. & Iwata, N. Polygenic risk score as clinical utility in psychiatry: a clinical viewpoint. *J. Hum. Genet.* **66**, 53–60 (2021).
50. Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
51. Park, J. H. et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* **42**, 570–575 (2010).
52. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
53. Lewis, C. M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* **12**, 44 (2020).
54. Constantinou, E. et al. Measuring fear: association among different measures of fear learning. *J. Behav. Ther. Exp. Psychiatry* **70**, 101618 (2021).
55. Tabbert, K. et al. Influence of contingency awareness on neural, electrodermal and evaluative responses during fear conditioning. *Soc. Cogn. Affect. Neurosci.* **6**, 495–506 (2011).
56. Lonsdorf, T. B. & Merz, C. J. More than just noise: inter-individual differences in fear acquisition, extinction and return of fear in humans – biological, experiential, temperamental factors, and methodological pitfalls. *Neurosci. Biobehav. Rev.* **80**, 703–728 (2017).
57. Eloyan, A. et al. Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging. *Front. Syst. Neurosci.* <https://doi.org/10.3389/fnsys.2012.00061> (2012).
58. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142–2154 (2012).
59. Smith, S. M. et al. Correspondence of the brain’s functional architecture during activation and rest. *Proc. Natl Acad. Sci. USA* **106**, 13040–13045 (2009).
60. LaBar, K. S. & Cabeza, R. Cognitive neuroscience of emotional memory. *Nat. Rev. Neurosci.* **7**, 54–64 (2006).
61. Jiang, Y. et al. Perturbed connectivity of the amygdala and its subregions with the central executive and default mode networks in chronic pain. *Pain* **157**, 1970–1978 (2016).
62. Schlitt, F. et al. Impaired pain-related threat and safety learning in patients with chronic back pain. <https://doi.org/10.1097/j.pain.0000000000002544> (2021).
63. Horing, B. & Büchel, C. The human insula processes both modality-independent and pain-selective learning signals. *PLoS Biol.* **20**, e3001540 (2022).
64. Brooks, J. C. W. & Tracey, I. The insula: a multidimensional integration site for pain. *Pain* **128**, 1–2 (2007).
65. Lu, C. et al. Insular cortex is critical for the perception, modulation, and chronification of pain. *Neurosci. Bull.* **32**, 191–201 (2016).
66. Kong, J. et al. Functional connectivity of the frontoparietal network predicts cognitive modulation of pain. *Pain* **154**, 459–467 (2013).
67. Zunhammer, M. et al. Meta-analysis of neural systems underlying placebo analgesia from individual participant fMRI data. *Nat. Commun.* **12**, 1391 (2021).
68. Hwang, K. D., Kim, S. J. & Lee, Y. S. Cerebellar circuits for classical fear conditioning. *Front. Cell. Neurosci.* **16**, 836948 (2022).
69. Ernst, T. M. et al. The cerebellum is involved in processing of predictions and prediction errors in a fear conditioning paradigm. *Elife* **8**, e46831 (2019).
70. Batsikadze, G. et al. The cerebellum contributes to context-effects during fear extinction learning: a 7T fMRI study. *Neuroimage* **253**, 119080 (2022).
71. Coombes, S. A. & Misra, G. Pain and motor processing in the human cerebellum. *Pain* **157**, 117–127 (2016).
72. Duerden, E. G. & Albanese, M. C. Localization of pain-related brain activation: a meta-analysis of neuroimaging data. *Hum. Brain Mapp.* **34**, 109–149 (2013).
73. Jensen, K. B. et al. Brain activations during pain: a neuroimaging meta-analysis of patients with pain and healthy controls. *Pain* **157**, 1279–1286 (2016).
74. Langner, R., Eickhoff, S. B. & Bilalić, M. A network view on brain regions involved in experts’ object and pattern recognition: implications for the neural mechanisms of skilled visual perception. *Brain Cogn.* **131**, 74–86 (2019).
75. Tiego, J. et al. Precision behavioral phenotyping as a strategy for uncovering the biological correlates of psychopathology. *Nat. Ment. Health* **1**, 304–315 (2023).
76. Gell, M. et al. The burden of reliability: how measurement noise limits brain-behaviour predictions. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.02.09.527898> (2023).
77. Kincses, B. kincsesbalint/paintone\_rsn: CommunicationsBiology (commbiol). Zenodo. <https://doi.org/10.5281/zenodo.12699040> (2024).
78. Forkmann, K., Wiech, K., Schmidt, K., Schmid-Köhler, J. & Bingel, U. Neural underpinnings of preferential pain learning and the modulatory role of fear. *Cereb. Cortex* <https://doi.org/10.1093/cercor/bhad236> (2023).
79. Pawlik, R. J. et al. Inflammation shapes neural processing of interoceptive fear predictors during extinction learning in healthy humans. *Brain Behav. Immun.* **108**, 328–339 (2023).
80. Koenen, L. R. et al. From anticipation to the experience of pain: the importance of visceral versus somatic pain modality in neural and behavioral responses to pain-predictive cues. *Psychosom. Med.* **80**, 826–835 (2018).
81. Radloff, L. S. The CES-D scale: a self-report depression scale for research in the general population. *Appl. Psychol. Meas.* **1**, 385–401 (1977).
82. McCracken, L. M., Zayfert, C. & Gross, R. T. The pain anxiety symptoms scale: development and validation of a scale to measure fear of pain. *Pain* **50**, 67–73 (1992).
83. Sullivan, M. J. L., Bishop, S. R. & Pivik, J. The pain catastrophizing scale: development and validation. *Psychol. Assess.* **7**, 524–532 (1995).
84. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage* **62**, 782–790 (2012).
85. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
86. Avants, B. B. et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011).
87. Gorgolewski, K. et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. *Front. Neuroinform.* **5**, 13 (2011).
88. Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J. & Turner, R. Movement-related effects in fMRI time-series. *Magn. Reson. Med.* **35**, 346–355 (1996).
89. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101 (2007).

## Acknowledgements

The work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 422744262 – TRR 289 (“Treatment Expectation”) (Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 422744262 – TRR 289) and Project-ID 316803389 – TRR 1280 (“Extinction Learning”).

## Author contributions

B.K., U.B., and T.S. conceived the study. Discovery study was planned and performed by K.F., K.S., and U.B.; validation study 1 was planned and performed by B.K., K.F., K.S., F.S., and U.B.; validation study 2 was planned and performed by R.J.P. and S.E. B.K. and T.S. conducted the analyses; B.K., K.W., U.B., and T.S. wrote the manuscript. K.S., D.T., R.J.P., F.S., S.E., K.W., B.K., U.B., and T.S. contributed to the interpretation, as well as manuscript revision. The whole study was supervised by B.K., U.B., and T.S.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-024-06574-y>.

**Correspondence** and requests for materials should be addressed to Balint Kincses.

**Peer review information** *Communications Biology* thanks Li Hu, Marta Ceko, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Joao Valente. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024