Supplementary Information

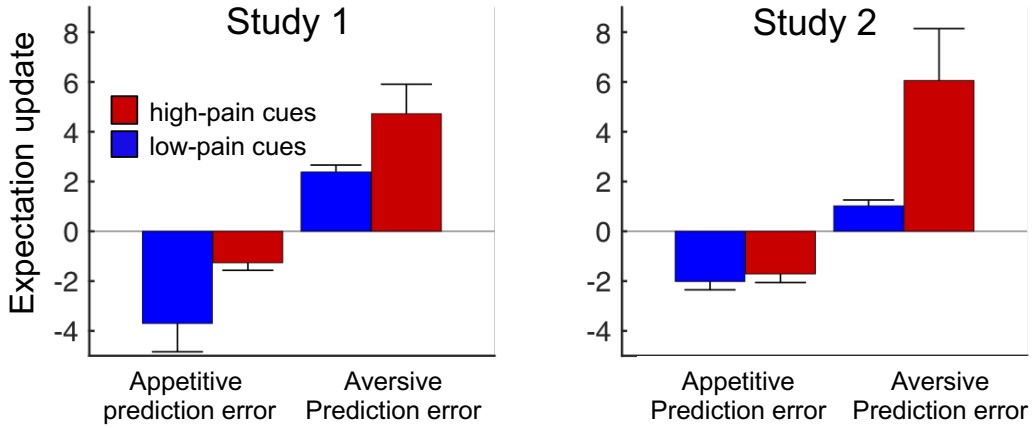
Behavioural and neural evidence for self-reinforcing expectancy effects on pain

Marieke Jepma, Leonie Koban, Johnny van Doorn, Matt Jones, and Tor D. Wager

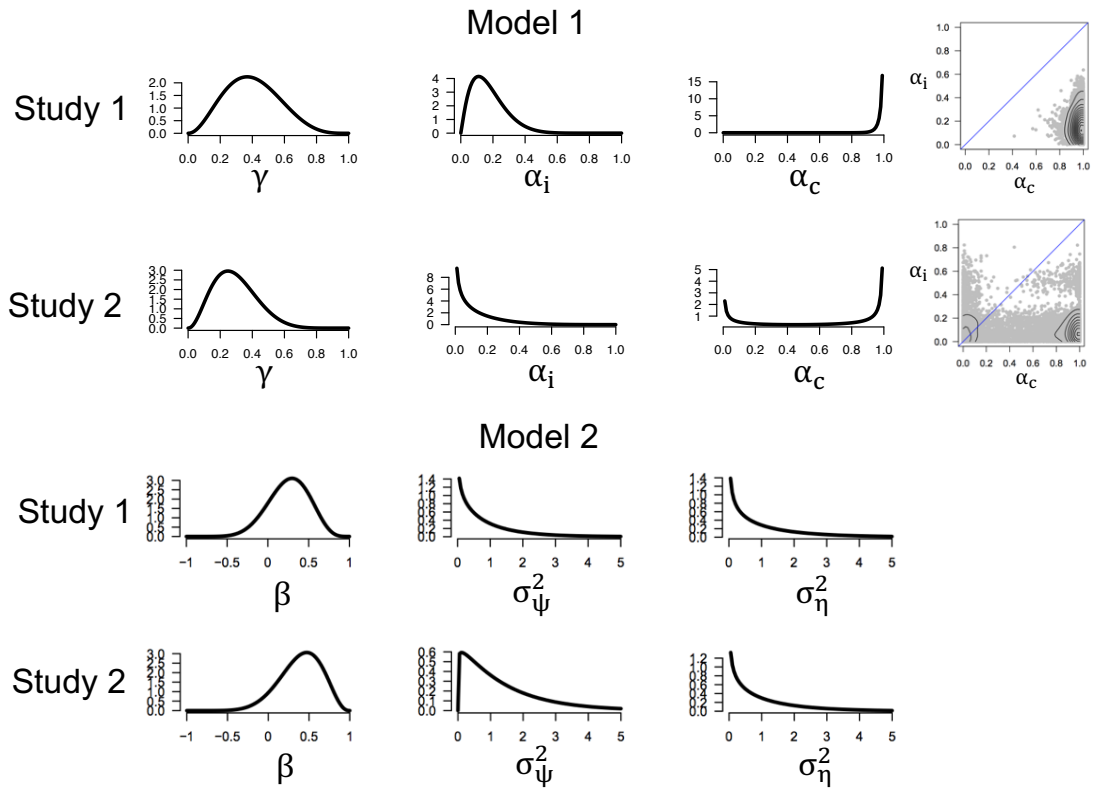
Supplementary Results: Confirmation bias on raw expectation updates

A potential problem with testing effects on learning rate—defined as expectation update divided by prediction error— is that learning rates could be very high on trials with very small prediction errors, disproportionately driving results. Therefore, we repeated the confirmation-bias analysis reported in the main text using raw expectancy update (not scaled by prediction error) as the dependent variable. As expected, aversive and appetitive prediction errors triggered upward and downward expectation updating, respectively, reflected in a main effect of prediction-error sign on signed expectancy updating ($t(25) = 5.6$, bootstrap $p < .001$, $d = 1.9$, CI = 1.9 to 2.9 and $t(20) = 4.5$, bootstrap $p < .001$, $d = 1.2$, CI = 1.6 to 2.9 in Study 1 and 2, respectively). In addition, upward expectation updating was stronger for high- than low-pain cues, reflected in a main effect of cue type on signed expectancy updating ($t(25) = 2.1$, bootstrap $p < .001$, $d = .77$, CI = .28 to .82 and $t(20) = 2.0$, bootstrap $p = .003$, $d = .45$, CI = .19 to 1.4 in Study 1 and 2, respectively). Furthermore, consistent with a confirmation bias, absolute update size was largest following aversive prediction errors on high-cue trials and appetitive prediction errors on low-cue trials, reflected in an interaction between cue type and prediction-error sign on absolute update size ($t(25) = 2.1$, bootstrap $p < .001$, $d = .77$, CI = .28 to .82 and $t(20) = 2.0$, bootstrap $p = .003$, $d = .45$, CI = .19 to 1.4 in Study 1 and 2, respectively; see Supplementary Figure 1).

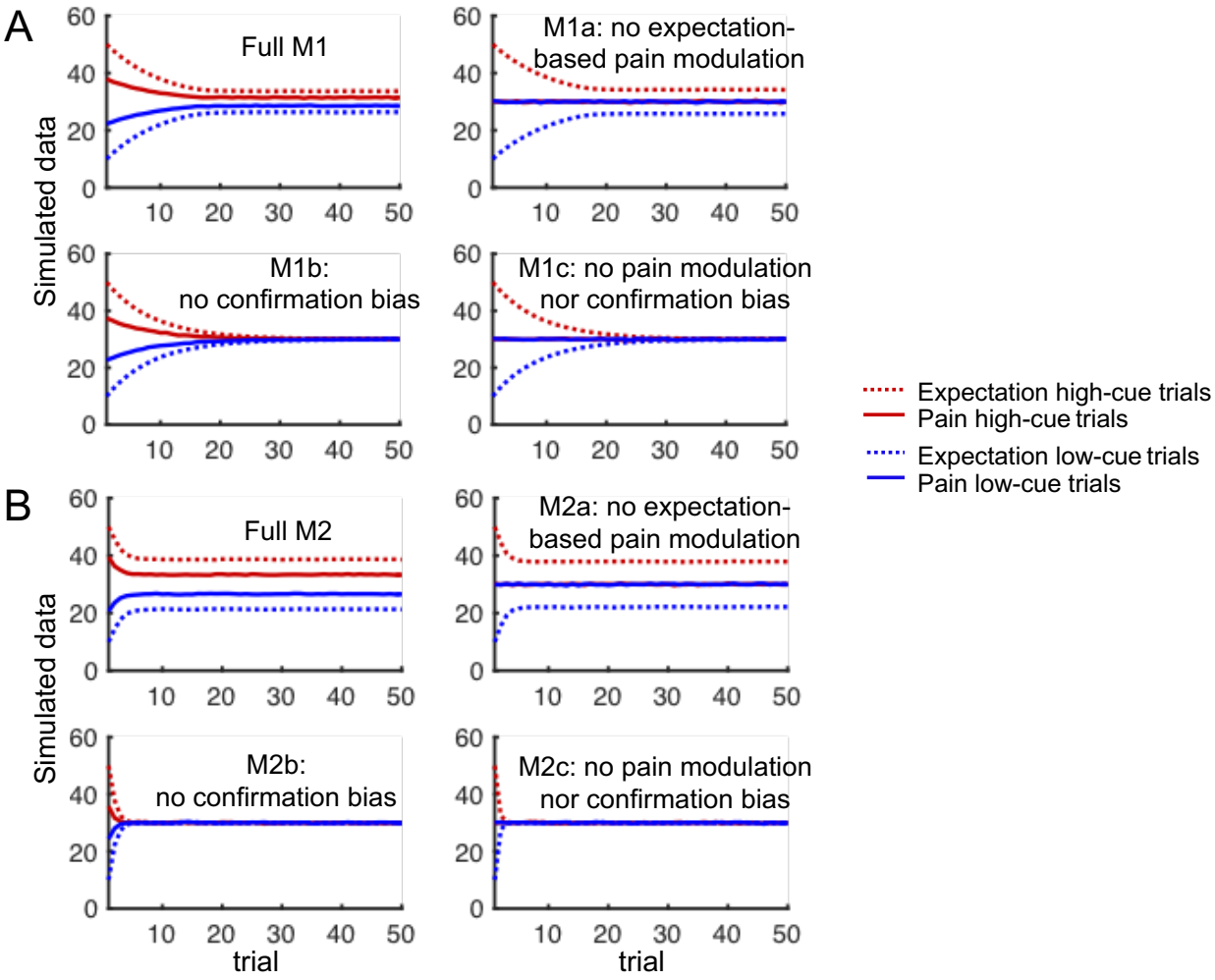
Supplementary Figure 1. Mean signed expectation update as a function of prediction-error sign and cue type in each study. Negative and positive expectation updates indicate decreases and increases in expected pain, respectively. For Study 1, the first three conditions include data from 28 participants, and the last condition includes data from 26 participants. For Study 2, the first three plots include data from 33 participants, and the last plot includes data from 21 participants. Fewer participants contributed to the last condition because some participants never experienced aversive prediction errors on high-cue trials. Error bars are between-subject standard errors.



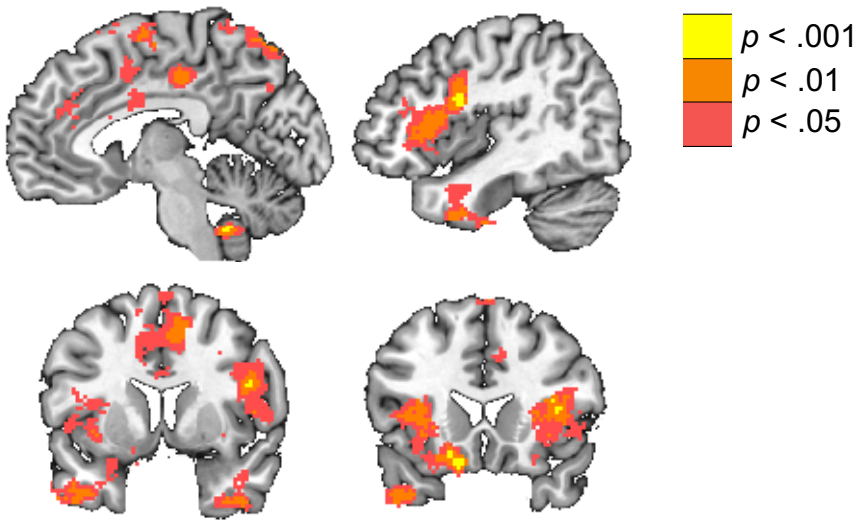
Supplementary Figure 2. Estimated population distributions of each model's parameters. Note that α_c could not be reliably estimated for a few participants in Study 2, who did not experience enough trials in which the sign of the prediction error was consistent with the initial low or high pain value of the preceding cue. The individual posterior distributions of α_c for those participants were very broad (spanning the entire range from 0 to 1), which explains the bimodal population distribution for α_c in Study 2.



Supplementary Figure 3. Model simulations. On each of 100 trials, the models were presented with a low- or a high-pain cue (on half of the trials each) followed by a noxious stimulus with an intensity of 25 or 35 (on half of the low- and high-cue each, in random order). The initial pain expectations associated with the high- and low-pain cues were 50 and 10, respectively. We ran each model 1000 times, and plotted the average simulated expectation and pain ratings. Parameter values were set to the median of their group-level posterior distributions fitted to the data from Study 1. **A**. Results for the full and reduced versions of Model 1. **B**. Results for the full and reduced versions of Model 2. Only models including both expectation-based pain modulation and a confirmation model (upper left plots) predict persistent cue effects on both expectations and pain.



Supplementary Figure 4. Parametric modulation of anticipatory brain activity by expected-pain rating. All colored regions contain at least 5 voxels that are significant at $p < .001$, uncorrected. There was no significant FDR-corrected activity. Plots are based on data from 34 participants



Supplementary Table 1. Log_e Bayes factors for comparisons of the different versions of Model 1, including the extended versions of Model 1 ($M1^+$ and $M1b^+$) that allowed for a persistent effect of initial beliefs on pain perception.

Study 1						
	M1	M1a	M1b	M1c	$M1^+$	$M1b^+$
M1	0	391	129	529	12	90
M1a		0	-262	138	-374	-296
M1b			0	400	-110	-33
M1c				0	-510	-432
$M1^+$					0	78
$M1b^+$						0
Study 2						
	M1	M1a	M1b	M1c	$M1^+$	$M1b^+$
M1	0	263	86	334	-8	94
M1a		0	-177	71	-268	-166
M1b			0	248	-97	4
M1c				0	-345	-244
$M1^+$					0	102
$M1b^+$						0

Note. M1 is Model 1, and a, b and c refer to the reduced models that do not include (a) expectancy-based pain modulation, (b) a confirmation bias, and (c) either of those components. $M1^+$ and $M1b^+$ are extended with a direct effect of initial beliefs on pain. Thus $M1b^+$ is $M1^+$ without a confirmation bias in learning rate.

Supplementary Table 2 (related to Figure 7B)

Note. Brain regions in which updating of anticipatory activity following aversive prediction errors on high- vs. low-cue trials correlated with individual differences in confirmation bias on learning rate (N=21). Results are whole-brain FDR-corrected ($q < .05$) and clusters are defined based on contiguity with voxels at uncorrected $p < .001$ (at least 5 voxels) and $p < .01$ (at least 5 voxels). Vox: number of voxels.

Region	**x**	**y**	**z**	**Vox**	**Max *t* value**
Positive correlation with confirmation bias on learning rate					
Brainstem	4	-32	-46	297	13.60
R temporal fusiform cortex	26	-8	-42	101	12.69
L cerebellum	-20	-78	-42	28	11.05
R temporal pole	28	20	-34	14	10.54
Brainstem	10	-40	-36	28	11.11
R cerebellum	14	-50	-20	151	9.39
L frontal orbital cortex	-16	16	-22	73	16.98
Brainstem/midbrain, encompassing PAG	2	-34	-18	57	13.04
R thalamus	8	-22	-2	80	11.31
R putamen	20	16	-2	121	9.44
L putamen	-24	18	0	27	10.67
R superior temporal gyrus/middle temporal gyrus	48	-36	0	28	14.37
L frontal operculum cortex	-36	12	10	20	10.49
L superior temporal gyrus	-70	-40	10	8	9.14
R frontal operculum cortex	32	20	12	38	10.53
Anterior cingulate cortex	-6	12	28	21	10.69
L supplementary motor cortex	-12	-12	48	13	12.11
Precentral gyrus	4	-26	70	87	13.18
Supplementary motor cortex	6	-2	70	194	16.60
Negative correlation with confirmation bias on learning rate					
L cerebellum	-26	-40	-40	74	11.97
L cerebellum	-40	-40	-38	20	16.50
L parahippocampal gyrys	-14	0	-36	32	14.83
R cerebellum	46	-80	-32	7	11.43
L temporal fusiform cortex	-38	-44	-20	67	11.51
L frontal orbital cortex	-34	16	-20	109	12.79
L frontal medial cortex/subcallosal cortex	-6	28	-14	120	10.97
Frontal pole/frontal medial cortex	-8	54	-22	30	12.23
Posterior cingulate cortex	-4	-44	-2	148	14.40
Frontal pole	-8	70	4	73	11.34
R hippocampus	24	-36	-2	37	13.96
Occipital pole	28	-96	6	35	12.27
R central opercular cortex	38	-6	20	5	12.21

Supplementary Table 3 (related to Figure 7C)

Note. Brain regions in which updating of anticipatory activity following appetitive prediction errors on low- vs. high-cue trials correlated with individual differences in confirmation bias on learning rate (N=21). Results are whole-brain FDR-corrected ($q < .05$) and clusters are defined based on contiguity with voxels at uncorrected $p < .001$ (at least 5 voxels) and $p < .01$ (at least 5 voxels). Vox: number of voxels.

Region	**x**	**y**	**z**	**Vox**	**Max *t* value**
Positive correlation with confirmation bias on learning rate					
R cerebellum	38	-74	-46	158	10.70
R temporal pole	22	14	-38	87	19.77
L temporal pole	-36	10	-44	203	14.57
Brainstem	0	-40	-34	64	11.02
R parahippocampal gyrus	22	0	-28	19	10.98
L frontal orbital cortex	-14	28	-22	11	12.42
L middle temporal gyrus	-56	-22	-12	29	11.48
Anterior cingulate cortex	0	34	-4	23	9.37
Posterior cingulate cortex	2	-44	24	49	10.01
Negative correlation with confirmation bias on learning rate					
R inferior temporal gyrus	46	-8	-40	71	13.29
L temporal pole/ frontal orbital cortex	-24	12	-30	18	12.27
L occipital pole	-22	-92	-20	78	11.83
L frontal pole	-34	46	-14	25	14.90
R frontal pole	24	54	-14	143	13.38
R frontal pole	50	46	-8	30	17.60
R occipital pole	22	-96	2	107	12.67
L inferior frontal gyrus	-56	28	14	227	12.09
L middle frontal gyrus	-46	18	36	166	12.53