Pain relief provided by an outgroup member enhances analgesia

Grit Hein[1], Jan B. Engelmann[2], and Philippe N. Tobler[3]

[1] Department of Psychiatry, Psychosomatic and Psychotherapy, Translational Social Neuroscience Unit, University of Wurzburg, Wurzburg, 97080, Germany

[2] Center for Research in Experimental Economics and Political Decision Making (CREED), Amsterdam School of Economics, and Amsterdam Brain and Cognition (ABC), University of Amsterdam, Amsterdam, 1001, The Netherlands

[3] Department of Economics, Laboratory for Social and Neural Systems Research, University of Zurich, Zurich, 8006, Switzerland

Supplementary material

1. Supplementary methods

Participants. Forty healthy men (mean age = 22.7, SE = 0.41) participated in the study, and were randomly assigned to an ingroup or an outgroup treatment group. There were no age differences between the groups, $t(38) = -0.34$, $P = 0.73$. Four data sets had to be discarded, two because of motion artefacts and two due to technical problems with the response box, resulting in groups of 18 (ingroup treatment) and 18 (outgroup treatment). Four students were trained to act as ingroup or outgroup treatment provider; the assignment of confederates to ingroup vs. outgroup was counterbalanced across participants. We chose an all-male instead of a gender-mixed group of participants to both limit total the number of confederates and avoid the complications of gender-mixed pairing of participants and confederates. To make efficient use of the complex set-up and scanning time, participants performed in a second independent task that is published elsewhere (1). None of the data in the present paper have been published previously.

Social context manipulation. We used a group manipulation with high ecological validity in our country (Switzerland). Ingroup treatment providers were individuals who ostensibly shared the participant's nationality (Swiss), while outgroup treatment providers were ostensibly of Balkan descent, that is, representatives of one of the largest minority groups in

Switzerland whose presence is often portrayed as problematic. To indicate his group membership, the ingroup treatment provider introduced himself with a typical Swiss name and the outgroup treatment provider with a typical Balkan name. Apart from that, the ingroup and outgroup treatment conditions were identical. Before scanning, the participants were asked to collect and write down stereotypical attributes of a Balkan male, which is a well-established procedure for activating stereotypes (2). To check the success of the group manipulation, and to assess how participants evaluated the ingroup and outgroup context prior to the treatment, participants rated their impression of the ingroup and outgroup treatment provider on a well-established impression scale (1, 3). The scale ranges from 1 to 9 and includes questions regarding perceived group membership (e.g., "To what extent do you see yourself and this other person as part of the same group"), similarity (e.g., "How much do you think you and this person have in common?"), likability ("How likable do you find this person?") etc. Participants of both treatment groups received identical instructions regarding the role of the treatment provider, that is, they were informed that the treatment provider would make decisions that could affect their pain stimulation.

Pre- and post-treatment sessions (pain processing). In the pre- and post-treatment sessions, participants received pain stimulation (10 trials each). This relatively small number was chosen for ethical reasons, in order to minimize overall pain exposure and distress during the procedure. After a pain anticipation phase consisting of a green arrow cue (500 ms) and a fixation cross (1500 ms), a green lightning bolt symbol was presented. After 1000 ms, it turned yellow for 1000 ms and simultaneously a painful electrical shock was delivered. After a fixation period of 3000 ms, a rating scale was presented on which participants rated how they felt receiving the painful shock on a scale from -4 (*very bad*) to +4 (*very good*).
Treatment session (learning). The treatment session consisted of 20 trials in which the participants expected to receive painful shocks. Each trial started with a pain anticipation phase, in which a green arrow cue indicated the delivery of painful stimulation (500 ms), followed by a fixation cross (1000 ms). Next, the ingroup treatment provider (ingroup treatment group) or the outgroup treatment provider (outgroup treatment group) ostensibly modified the pain stimulation of the participant by pressing one of two keys (jittered from 2500-3500 ms). In 75% of the trials (15 out of 20), the treatment provider ostensibly decided to prevent painful stimulation for the participant. In this case, a crossed-out lightning bolt (pain relief symbol) was shown at the end of the trial (1000 ms). In the rest of the trials, the treatment provider ostensibly decided to apply painful stimulation; these trials were indicated

by an intact lightning bolt (pain symbol) accompanied by a painful shock. Please note that the symbols were the same in both treatment groups. To maximize power, the main analysis focused on dynamic pain relief anticipation in all trials rather than analyzing the rarer surprising pain outcomes separately. At the end of each treatment trial, the participant rated his emotions regarding the ingroup or outgroup treatment provider on a scale from -4 (very negative) to +4 (very positive). These emotion ratings captured changes in the evaluation of the social context (i.e., the ingroup or outgroup treatment provider) as a result of experienced pain relief decisions. We therefore used them as a behavioral indicator of learning pain relief anticipation (see below). Pain relief and pain application decisions occurred in random order, with the restriction that the number of consecutive pain application decisions was limited to two.

Image acquisition. The experiment was conducted on a 3-Tesla Philips whole-body MR scanner (Philips Medical Systems, Best, The Netherlands) equipped with an 8-channel Philips SENSitivity Encoded (SENSE) head coil. Structural image acquisition consisted of 180 T1-weighted transversal images (voxel size of 1 mm). For functional imaging, we used T2*-weighted echo-planar imaging (35 slices, slice thickness of 3 mm, interslice gap of 0.5 mm, ascending acquisition, TR = 2100 ms; TE = 30 ms; flip angle = 80°; field of view = 240 mm; matrix 80 x 80). In the pre- and post-treatment sessions, a total of 495 images were acquired. In the treatment session, a total of 110 images were acquired.

Imaging data analyses

Preprocessing. All functional volumes were realigned to the first volume using b-spline interpolation and subsequently unwarped using fieldmaps estimated by SPM to remove residual-movement-related variance due to susceptibility-by-movement interactions. To improve co-registration, bias correction and co-registration of anatomical and mean EPI images were performed with the New Segment toolbox in SPM. The forward deformation fields created via nonlinear normalization of individual gray matter tissue probability maps were then employed to normalize the functional images to the Montreal Neurological Institute (MNI) T1 template. Finally, functional data underwent spatial smoothing using an isotropic 6-mm full-width-at-half-maximum Gaussian kernel.

First-level analyses. Two separate first-level models were estimated for each participant, one that captured the neural correlates of learning during the treatment session (learning model) and one that captured pain-related activations in the pre-treatment and post-treatment sessions (pain model). In both models, time and dispersion derivatives of the canonical hemodynamic response functions were added to account for subject-to-subject and voxel-to-voxel variation in response onset and width. To account for physiological noise and motion, we added a global signal regressor to both models (4), which was created by extracting the normalized mean signal from a whole brain mask using the Marsbar toolbox and customized Matlab scripts.

In the learning model, the regressor of primary interest tracked learning of pain relief anticipation in the presence of a pain-predicting cue (arrow), corresponding to a duration of 1500 ms. Specifically, the trial-by-trial learning-induced pain relief anticipation as described in the *reinforcement learning model* section (Supplementary Methods) was used as parametric modulator. The decision phase of the ingroup/ outgroup treatment provider and the outcome phase (pain relief symbol in pain relief trials; pain symbol and pain stimulation in pain trials) were modelled as regressors of no interest.

For the pain model, regressors of interest modeled the delivery of pain, indicated by the yellow lightning bolt, which lasted 1000 ms. All other events were modelled as regressors of no interest (arrow cue, fixation cross, green lightning bolt, rating scale).

Second-level analyses. The second-level analyses were based on contrast images that resulted from linearly contrasting parameter estimates for the regressors of interest in the learning and pain models. To assess the impact of outgroup treatment on neural pain processing, we used the first-level images that captured the neural response to pain stimulation in the pre-treatment session and the post-treatment session (pain model), contrasted the pre-treatment versus post-treatment images, and tested the contrast images against zero using a t-test. To test whether the pre- vs. post-treatment difference in pain-related activation in the outgroup treatment group was driven by learning, we used a second-level regression to assess the relationship between participants' pre- vs. post-treatment difference and the magnitude of the individual neural learning signal extracted from right AI. Please note that the presence of ingroup versus outgroup members is constant and controlled for in these analyses. To test the hypothesis that AI cortex activation reflects pain-related learning and the resulting pre-vs-post treatment differences in pain processing, we analyzed our data in bilateral anatomical masks of the insular cortex (5), using small-volume family-wise error (SV FWE) correction ($P < 0.05$).

Moreover, we conducted exploratory whole brain analyses (uncorrected, $P < 0.001$, $k = 5$; Tables S1, S3 and S4).

Reinforcement learning model.

To test for neural learning signals reflecting trial-by-trial changes in pain relief anticipation during the treatment session, we used a standard reinforcement learning model (6, 7):

$$\text{EQ1. } V_{t+1} = V_t + \alpha_t\delta_t \qquad \text{with } \delta_t = \alpha\,(\lambda_t - V_t)$$

V_t corresponds to the value of anticipated pain relief in the current trial, which is based on previous learning up to this point, but before the learning experience on the current trial (this term was used as parametric modulator at the neuroimaging level to capture learned pain relief anticipation), V_{t+1} corresponds to the pain relief anticipation in the next trial and is a combination of V_t and learning rate (α)-weighted prediction error (δ_t). Specifically, δ_t corresponds to the error in the prediction of pain relief, i.e., the difference between experienced (λ_t) and anticipated pain relief (V_t) in the current trial. λ_t was set to 1 for experiences of pain relief and -1 for experiences of no pain relief (i.e., of pain). Thus, high values of V reflect strong anticipation of pain relief. We used a fixed learning rate (α) of 0.3, which is commonly reported in reinforcement learning paradigms (8) and also captured the changes in trial-by-trial emotion ratings in our paradigm (Supplementary Results section).

Regression analyses. To identify the impact of neural learning signals on pre- vs. post-treatment changes in pain ratings, we conducted an ANOVA based on the following ordinary least squares (OLS) regression model:

$$y_i = \beta_0 + \beta_1 AIS_i + \beta_2 G_i + \beta_3 AIS_i G_i + \beta_4 X_i + \epsilon_i,$$

where the dependent variable y_i represents the change in average pain ratings (pre-treatment – post-treatment) for individual i. AIS_i is the beta value extracted from right anterior insula reflecting the model-based learning value for individual i. G_i is a dummy variable reflecting the type of treatment that individual i received (1 for outgroup treatment, 0 for ingroup treatment). X_i is a set of control variables that include age and average impression ratings. The regression model was estimated using OLS implemented in R (lm) and summarized using

the Anova command from the R package "car" to extract Type III sums of squares. Note that the regression approach is equivalent to a standard ANCOVA approach, and was used here for visualization of the relationship between continuous dependent (pre-post treatment differences in pain ratings) and predictor variables (learning signal from AI).

2. Supplementary results

Confirmation of assumed learning rate

To confirm that the assumption of $\alpha = 0.3$ holds for our sample, we additionally estimated the average behavioral learning rate by using trial-by-trial emotion ratings. To do so, we adopted equation 1 and used the interaction partner's decisions to provide pain relief and pain as rewarding and punishing outcomes (λ_t). Specifically, we modeled how these outcomes affected changes in emotion ratings (V) as a function of learning rate (α)-weighted prediction errors (δ) defined as the difference between the observed reinforcement and the current emotion rating (6, 9). Missing emotion ratings occurred on three trials (0.42 % of all trials) and were replaced via mean substitution. We used non-linear least squares (lsqcurvefit implemented in Matlab) to estimate the learning parameter α by fitting the modeled emotion rating (V) to the observed ratings (scaled to outcomes so as to fall between the values -1 and 1). Moreover, to test the robustness of the estimations, we iterated the starting parameters for estimating the learning rate (α) between 0.1 and 0.9 and found that final estimates of the learning rate did not differ significantly from our assumed learning rate of 0.3.

Neural correlates of learning: additional analysis

In an additional analysis we used the classical prediction error (δ_t) on the previous trial as parametric modulator of the neural responses in the pain anticipation window of each trial. The results revealed AI cortex activation (Fig. S1, yellow) that overlapped with the activation in right AI cortex found for trial-by-trial changes in pain relief values (Fig. S1, orange). Finally, we used a learning model that contained participants´ average pain ratings from the pre-intervention phase as prior, instead of zero. Inclusive masking showed that right AI cortex was also sensitive to subjective prior expectations, in addition to tracking pain relief values and the prediction error signal.

Main results cannot be explained by potential imaging artefact

It could be argued that the correlation between the AI signals in different parts of the study might be driven by an imaging artefact such as a poor EPI signal in this region. In this

case, the pre- vs. post-treatment differences in pain-related brain responses should also correlate with the signal from an insula region that did not significantly correlate with learning, e.g., the left AI cortex. To test this possibility, we regressed the pre-vs-post differences in pain-related activation against the left AI signal elicited during learning in the intervention phase. We found no significant results (even at P uncorrected < 0.05), which renders the assumption unlikely that the observed significant correlation (Fig. 4) is driven by imaging artefacts.

References

1. Hein G, Engelmann JB, Vollberg MC, Tobler PN. How learning shapes the empathic brain. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(1):80-5.
2. Dijksterhuis A, van Knippenberg A. The relation between perception and behavior, or how to win a game of trivial pursuit. Journal of personality and social psychology. 1998;74(4):865-77.
3. Hein G, Silani G, Preuschoff K, Batson CD, Singer T. Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. Neuron. 2010;68(1):149-60.
4. Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. NeuroImage. 2014;84:320-41.
5. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage. 2002;15(1):273-89.
6. Dayan P, Abbott L. Theoretical neuroscience: computational and mathematical modeling of neural systems. Journal of Cognitive Neuroscience. 2003;15(1):154-5.
7. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory. 1972;2:64-99.
8. Gershman SJ. Do learning rates adapt to the distribution of rewards? Psychonomic bulletin & review. 2015;22(5):1320-7.
9. Behrens TE, Hunt LT, Woolrich MW, Rushworth MF. Associative learning of social value. Nature. 2008;456(7219):245.

Supplementary figures and tables

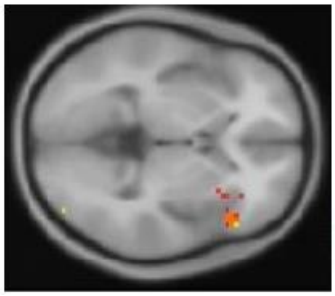

Figure S1. Impact of social context and learning on pre- vs. post-treatment changes in pain ratings. The right anterior insular cortex shows overlapping activation (orange) reflecting trial-to-trial changes in pain relief value (red) and prediction errors (yellow; visualized at $P < 0.001$, uncorrected).

Table S1. Whole brain analyses across treatment groups (uncorrected, $P < 0.001$, $k = 5$) revealing neural regions that are correlated with change in pain relief anticipation.

	Coordinates			***T*-value**	***Z*-value**	**Voxels in Cluster**
Brain region	X	Y	Z			
right anterior insula/ inferior frontal gyrus*	48 39	26 26	2 5	4.56 4.22	4.01 3.77	61
left temporal pole	-33	17	-22	4.57	4.02	16
left SMA	-15	5	62	4.40	3.90	13
right SMA	6	11	59	4.39	3.89	14
right middle frontal gyrus	42	-4	53	4.21	3.76	17
right middle occipital gyrus*	36	-88	8	5.57	4.68	58
right fusiform gyrus	30	-67	-13	4.46	3.94	29
left superior occipital gyrus	-15	-97	17	4.3	3.83	9
left middle temporal gyrus	-51	-55	5	4.34	3.86	6
left lingual gyrus	-18	-85	-10	4.2	3.75	6
right lingual gyrus	21	-85	-13	3.99	3.60	8
left precentral gyrus	-45	2	29	4.01	3.61	6
right precentral gyrus	51	5	32	3.59	3.27	11
cerebellar vermis	3	-40	-13	3.88	3.51	9

SMA, supplementary motor area; *FWE corrected (whole brain cluster level) < 0.05

Table S2. Results of ANOVA (left panel) and sequential linear regression analyses testing for the effects of learning (anterior insula learning signal, right panel a), the effect of social context (treatment type, right panel b), and the interaction between social context and learning (right panel c), on the individual pre- vs. post-treatment differences in pain ratings (= delta pain rating). Age and impression ratings were included as control variables. DV = dependent variable, delta pain rating = pre- vs. post-treatment differences in pain ratings.

DV = delta pain rating	**ANOVA**	**Sequential regression analyses**		
		a. B (SE)	**b.** B (SE)	**c.** B (SE)
anterior insula learning signal	F(1,30) = 0.387	0.162* (0.072)		-0.059 (0.096)
treatment type (ingroup/ outgroup)	F(1,30) = 0.037		0.543* (0.234)	0.049 (0.257)
anterior insula learning signal x treatment type	F(1,30) = 7.421 *			0.354** (0.129)
impression ratings	F(1,30) = 0.809	0.001 (0.018)	-0.001 (0.017)	0.017 (0.016)
Age	F(1,30) = 0.557	0.012 (0.039)	0.009 (0.039)	0.02 (0.034)
Intercept	F(1,30) = 0.762	-0.219 (1.042)	-0.132 (1.026)	-0.833 (0.927)

*P = 0.05; ** P < 0.05; B = beta coefficient; SE = standard error

Table S3. Whole brain analyses (uncorrected, $P < 0.001$, $k \geq 5$) revealing neural regions that are correlated with the pre-to-post treatment difference in the outgroup treatment group.

	Coordinates			***T*-value**	***Z*-value**	**Voxels in Cluster**
Brain region	X	Y	Z			
left anterior insula	-42	2	5	4.65	3.68	8
left postcentral gyrus	-63	-22	20	4.88	3.81	13
left inferior parietal lobe	-51	-25	41	4.63	3.67	6

Table S4. Whole brain analyses (uncorrected, $P < 0.001$, $k \geq 5$) revealing pre- vs. post-changes in pain-related activation after outgroup treatment predicted by the individual learning signal extracted from right anterior insula.

	Coordinates			***T*-value**	***Z*-value**	**Voxels in Cluster**
Brain region	X	Y	Z			
right anterior insula*	39	17	-7	6.25	4.38	36
left anterior insula/ left inferior frontal gyrus	-39	20	-4	4.89	3.77	14
left inferior frontal gyrus*	-57	14	5	5.47	4.05	41
left SMA	-18	-10	59	4.51	3.57	8
right SMA	15	11	62	4.08	3.33	6
left middle temporal gyrus	-54	-16	-16	4.62	3.63	5

SMA, supplementary motor area; *FWE corrected (whole brain cluster level) < 0.05.