Empathy induces sustained social closeness

Anne Saulin, Chih-Chung Ting, Jan B. Engelmann, & Grit Hein

Supplementary materials

# Supplementary Figures



**Figure S2.1.1**. Visualization of task particpants performed between blocks of the closeness learning task and observable results. (**A**) In the allocation task, participants repeatedly chose between a prosocial (here 600 points for participant and partner) and an egoistic option (here 880 points for the participant and 320 points for the partner) to divide points between themselves (green bar and values) and the respective partner (here red bar and values). After participants indicated which option they want to choose, a green rectangle highlighted the chosen option. For details on the task, please refer to (Saulin, Horn, Lotze, Kaiser, & Hein, 2022). (**B**) In each condition, participants performed three blocks of the allocation task. First at the very beginning (block 1), inbetween motive task blocks (block 2), and after the second block of the motive task (block 3). Here the mean relative frequency of prosocial decisions and SEMs are shown for each time point and condition for the fMRI study (left panel) and behavioral replication study (middle panel), and the control study (right panel)

Graphical user interface, application, website

Description automatically generated

**Figure S2.1.2**. Trial structure of the motive task in the fMRI and the behavioral replication study (**A**), as well as the control study (**B**). At the beginning of each trial, participants indicated in all studies how close they felt to the other person. Next, participants were given feedback about whether the other person received painful stimulation /had decided to help them (high pain trial/help trial = reinforced trial) or non-painful stimulation/not to help (no pain trial/no help trial = non-reinforced trial). Then participants rated how they felt after observing this feedback. After an inter-trial-interval of 4000-6000 ms, the next trial started.

**Calendar

Description automatically generated with low confidence**

**Figure S2.1.3.** Visualization of absolute model fit (mean values and SEMs) for all studies and the three models from model space I. (**A**) fMRI study. The *basic model* (model 1, dark blue) starkly underestimates participants’ closeness behavior in the second block of the treatment condition, while the *differential model* (model 2, cyan) overestimates closeness behavior in the first block of the treatment condition. The winning model (*individual calibration model*, model 3, magenta) captures participants’ closeness ratings quite well, across both blocks of the treatment condition and the control condition. (**B**) This pattern is replicated in the independent laboratory replication study. (**C**) All three models comparably well capture participants’ behavior in the different blocks and conditions in the laboratory control study.

A picture containing text

Description automatically generated

**Figure S2.1.4.** Visualization of the neural activation in the regions for which neural activation was larger in response to observed pain as compared to observed non-pain. **A** Bilateral temporo-parietal junction. **B** Right inferior frontal gyrus/anterior insula. Effects are whole-brain FWE cluster-corrected and visualized at p < .001 uncorrected and k >100 voxels.

# Supplementary Tables

**Table S2.1.1.** Demographics and characteristics of participants in the three studies.

|  |  |  |  |
| --- | --- | --- | --- |
|  | fMRI study | replication study | control study |
| N | 46 | 27 | 27 |
| Age [years] | 24.06 (4.52) | 22.89 (3.36) | 23.07 (3.35) |
| trait empathic concern | 19.07 (3.15) | 20 (2.59) | 19.45 (3.63) |
| trait perspective-taking | 17.82 (3.14) | 18.26 (3.39) | 18.07 (3.17) |

**Table S2.1.2**. Results of the Bayesian model comparison of the three models in model space I for the fMRI study (empathy fMRI, N=46), laboratory replication study (empathy lab, N=27), and the laboratory control study (reciprocity, N=27). Exceedance probabilities (EP) indicate the likelihood for a given model to have generated the data given the model space. Estimated model frequencies (EF) indicate the likelihood for a model to have generated the data of any randomly selected subject. The absolute value of the laplace approximation to the model evidence (LAME) indicates how well a given model fits the empirical data taking model complexity into account. Lower values indicate better model fit. Where applicable mean values ± SEs are reported.

|  |  |  |  |
| --- | --- | --- | --- |
|  | simple model | differential model | individual recalibration model |
| EP (empathy fMRI) | 0 | 0 | 1 |
| EF (empathy fMRI) | .0075±.00015 | .0243±.00049 | .9682±.00064 |
| LAME (empathy fMRI) | 9.1009±.7085 | 10.6256±.5710 | 6.7835±.4922 |
| EP (empathy lab) | 0 | 0 | 1 |
| EF (empathy lab) | .0127±.00043 | .0995±.0031 | .8879±.0034 |
| LAME (empathy lab) | 9.6402±1.0080 | 10.4064±.5754 | 7.6749±.8901 |
| EP (reciprocity) | .5381 | .0783 | .3836 |
| EF (reciprocity) | .3989±.0082 | .2451±.0064 | .3611±.0080 |
| LAME (reciprocity) | 9.4270±.9857 | 9.7715±.7127 | 9.5725±1.0070 |

**Table S2.1.3.** Results of the second-level analysis with the contrast parametric modulator trialtype > control during the emotion rating phase. This analysis shows which neural regions are more active during reinforced compared to non-reinforced trials in all blocks and conditions. P<.001 uncorrected, k = 20.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| region | hemisphere | T | P(cluster-level) | k | coordinates |
| TPJ | left | 6.21 | <.001 | 898 | -52 -52 20 |
|  | right | 4.74 | <.001 | 532 | 62 -48 22 |
| occipital pole | left | 3.71 | .903 | 26 | -18 -94 8 |
|  | right | 5.11 | .005 | 214 | 16 -92 8 |
| anterior insula | right | 5.06 | .406 | 60 | 28 16 -18 |
| middle temporal gyrus | right | 4.94 | .470 | 55 | 52 -28 -8 |
| inferior frontal gyrus | right | 4.64 | .033 | 143 | 38 28 -4 |
| calcarine cortex | left | 4.11 | .762 | 36 | -8 -88 2 |
| posterior orbital gyrus | left | 4.01 | .730 | 38 | -32 24 -16 |
| precuneus | right | 3.69 | .808 | 33 | 4 -46 48 |

# Supplementary analysis and results

## Computational results

Based on the results obtained from the first model comparison and in order to more closely investigate the computational basis of empathy-related social closeness sustainability, we developed a second model space. Considering the emotion reversal effect we observed on a behavioral level for empathy-based social closeness (see **Figure S2.1.4**), we tested, whether the individual recalibration of the outcome value in these two groups depended on either the condition (treatment vs. control), block (block 1 vs. block 2) or both. The second model space hence comprised four different models either assuming only one general recalibration parameter ω, assuming condition specific recalibration parameters ωtreatment and ωcontrol, assuming block-specific recalibration parameters ωblock 1 and ωblock 2, or assuming condition- and block-specific recalibration parameters ωtreat1, ωtreat2, ωcontrol1 and ωcontrol2.

Diagram, schematic

Description automatically generated

**Figure S2.1.4**. Model space II. In order to test whether recalibration of the feedback can explain the emotion reversal effect in the two empathy groups, we conducted Bayesian model comparison using a second model space. (**A**) The first model corresponds to the third model in model space I and allows for one general recalibration parameter. (**B**) The second model assumes different recalibration values for the treatment and the control condition. (**C**) The third model assumes different recalibration values for block 1 and block 2 in both conditions, and (**D**) the fourth model assumes different recalibration values for block 1 and 2 in the treatment condition and block 1 and 2 in the control condition.

We hypothesized that the sustained closeness in the extinction block may be mechanistically subserved by a reversal of the respective feedback value for reinforced and non-reinforced trials (i.e., non-reinforced trials actually become the reinforced trials in block 2 of the treatment condition). Such a reversal should entail a large recalibration of the feedback value in block 2 of the treatment condition, since the larger the recalibration value ω, the more the feedback value of reinforced trials moves closer to 0 (i.e., the original value of non-reinforced trials), and the more the feedback value of non-reinforced trials moves closer to 1 (i.e., the original value of reinforced trials, see **Equation 4** in the methods section). Hence, it is plausible to assume, that the extent of recalibration may be large in block 2 of the treatment condition, but small in block 1 of the treatment condition. No such differentiation should be observed for the control condition.

In order to test this hypothesis, we tested which out of four new models best described participants’ behavior (see **Figure S2.1.4** for visualization of model space II). Model 1 corresponds to the winning model from model space I (*individual calibration model*), model 2 is an extension of this model in that it assumes different values of recalibration for the treatment and the control conditionbut across both blocks (*condition-specific recalibration model*), model 3 assumes different values of recalibration for the block 1 and block 2 across both conditions (*block-specific recalibration model*), and model 4 assumes different values of recalibration for block 1 of the treatment condition, block 2 of the treatment condition, block 1 of the control condition, and block 2 of the control condition(*condition and block-specific recalibration model*). If the emotion reversal effect can indeed be understood in terms of outcome value reversal from the first to the second block of the treatment condition, the most complex model (*condition and block specific recalibration model*) should be most likely to have generated the data, revealing moderate recalibration in the two blocks of the control condition, low recalibration in the first block of the treatment condition, and high recalibration in the second block of the treatment condition.

In line with this hypothesis, model comparison results revealed that for the fMRI study as well as the behavioral replication study, the most complex model wins relative to the other models (see **Figure S2.1.5** for visualization of the Bayesian model comparison results, **Table S2.1.3** for comparison metrics, and **Figure S2.1.7** for visualization of absolute model fit).

Post-hoc contrasts revealed that the recalibration values did not differ between block 1 and block 2 for the control condition(fMRI study: t(45) = -.579, P = .568, CI = [-.139, .077]replication study: t(26) = -1.027, P = .314, CI = [-.176, .059]), but were significantly larger in block 2 compared to block 1 for the treatment condition in the fMRI study (t(45) = 2.753, P = .009, CI = [.345, .054]) and also marginally in the replication study (t(26) = 2.0, P = .056, CI = [-.005, .384], for visualization of the median and spread of the extracted parameters, see **Figure S2.1.6**).

Chart, waterfall chart

Description automatically generated

**Figure S2.1.5**. Results of the Bayesian model comparison of model space II. (**A**) In the fMRI study, the model assuming condition and block specific recalibration is more likely than the other models in the model space in terms of exceedance probability as well as estimated model frequencies. (**B**) In the laboratory replication study, this pattern is replicated regarding both metrics. Model 1 = *individual calibration model*, model 2 = *condition-specific recalibration model*, model 3 = *block-specific recalibration model*, model 4 = *condition and block specific recalibration model.*

Chart, box and whisker chart

Description automatically generated

**Figure S2.1.6**. Boxplots visualizing median and spread of the extracted ω parameters by block number and condition extracted from the winning model of model space II (*condition and block wise recalibration model*) (**A**) ω parameters resulting from modelling the behavior of the fMRI. (**B**) ω parameters resulting from modelling the behavior of replication study.

**Chart

Description automatically generated with medium confidence**

**Figure S2.1.7.** Visualization of absolute model fit (mean values and SEMs) for the fMRI study and the laboratory replication study comparing the fit of the *individual calibration model* (model 1, dark blue) and the *individual block- and condition-specific recalibration model* (model 4, magenta) from model space II. (**A**) In the fMRI study, across both blocks of the treatment condition, model 4 better describes participants behavior as compared to model 1, especially in the treatment condition. (**B**) This pattern is replicated for participants’ behavior in the laboratory study.

**Table S2.1.3**. Results of the Bayesian model comparison of the four models in model space II for the fMRI study (empathy fMRI, N=46), laboratory replication study (empathy lab, N=27), and the laboratory control study (reciprocity, N=27). Exceedance probabilities (EP) indicate the likelihood for a given model to have generated the data given the model space. Estimated model frequencies (EF) indicate the likelihood for a model to have generated the data of any randomly selected subject. The absolute value of the laplace approximation to the model evidence (LAME) indicates how well a given model fits the empirical data taking model complexity into account. Lower values indicate better model fit. Where applicable mean values ± SEs are reported.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | recalibration with 1 ω | condition specific recalibration (ωcontrol, ωtreatment) | block specific recalibration (ωblock1, ωblock2) | block and condition specific recalibration (ωtreatment1, ωtreatment2, ωcontol1, ωcontrol2) |
| EP (empathy fMRI) | 0 | 0 | 0 | 1 |
| EF (empathy fMRI) | .126±.002 | .072±.001 | .053±.001 | .739±.004 |
| LAME (empathy fMRI) | -74.168±.10.534 | -74.623±10.469 | -81.854±9.637 | -85.776±8.882 |
| EP (empathy lab) | 0 | 0 | 0 | 1 |
| EF (empathy lab) | .010±.0003 | .010±.0003 | .064±.002 | .917±.003 |
| LAME (empathy lab) | -59.409±13.654 | -60.615±.13.289 | -73.391±10.793 | -78.145±10.009 |
| EP (reciprocity) | .014 | .058 | .005 | .923 |
| EF (reciprocity) | .169±.005 | .234±.006 | .128±.004 | .470±.009 |
| LAME (reciprocity) | -36.974±13.849 | -39.234±.13.344 | -39.974±13.671 | -42.900±13.282 |

## Neural results

### Feedback phase

We first computed the contrast of neural activation during the feedback phase in the treatment condition>feedback phase in the control condition. We observed that neural activation was significantly larger in right precentral gyrus (peak coordinates: x = 52, y = -4, z = 40, T(44) = 4.85, P = .011, k = 175) and marginally larger in left putamen (peak coordinates: x = -26, y = -10, z = 10, T(44)=4.18, P=.063, k = 116) in the treatment condition compared to the control condition. Please note that both conditions yielded in sum the same number of reinforced and non-reinforced trials across the two respective blocks, allowing the conclusion, that any differences between conditions are driven by the difference in reinforcer schedules.

When adding individual values of ω (extracted from the winning model 3) as covariate, neural activation in dorso-lateral prefrontal cortex (dlPFC, peak coordintaes: x = -22, y = 34, z = 32, T(44) = 4.84, P = .042, k = 129) and ventro-lateral prefrontal cortex (vlPFC, peak coordinates: x = -28, y = 60, z = 6, T(44)=4.48, P=.061, k = 129) was modulated by the recalibration parameter more strongly in the treatment as compared to the control condition. Hence, the larger the individual recalibration value, the larger the difference in neural activation between the treatment and the control conditionduring the feedback phase in left dlPFC and vlPFC.

Further, when contrasting the neural tracking of trial type (reinforced vs. non-reinforced trial) during the feedback phase in the treatment condition and the control condition(i.e., the contrast parametric modulator of trial type during feedback in the treatment condition > parametric modulator of trial type during the control condition) and adding the individual values of ω as covariate revealed significant neural activation in left supramarginal gyrus (peak coordinates: x = -56, y = -28, z = 32, T(44) = 4.63, P = .003, k = 236). Thus, the larger an individual’s recalibration parameter, the more strongly neural activation was associated with trial type on a trial-by-trial basis in the treatment condition as compared to the control condition.

Across both conditions, neural activation was not larger for reinforced vs. non-reinforced trials (**Table S2.1.3**) or vice versa, nor did the extent of individual recalibration significantly modulate neural tracking of trial type across both conditions on a whole-brain level (Ps>.246).

**Table S2.1.3.** Results of the second-level analysis with the contrast parametric modulator trialtype > implicit control during the feedback phase. This analysis shows which neural regions are more active during reinforced compared to non-reinforced trials in all blocks and conditions. P<.001 uncorrected, k = 20.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| region | hemisphere | T | P(cluster-level) | k | coordinates |
| insula | left | 4.83 | .246 | 77 | -42 -6 12 |
| frontal operculum | left | 4.37 | .953 | 21 | -30 14 14 |
| supra marginal gyrus | left | 3.53 | .953 | 21 | -54 32 54 |

### Closeness rating phase

Adding the individual values of ω (extracted from the winning model for model space I) as covariate to this second level regression revealed neural activation in bilateral precuneus/PCC (left hemisphere peak coordinates: x = -4, y = -56, z = 56, T(44) = 4.62,; right hemisphere peak coordinates: x = 12, y = -52, z = 34, T(44) = 4.97, P < .001, k = 598), bilateral TPJ (left hemisphere peak coordinates: x = -52, y =-52, z = 44, T(44) = 4.96, P = .044, k = 140; right hemisphere peak coordinates: x = 38, y = -68, z = 36, T(44) = 5.07, P <.001, k = 495), left middle temporal gyrus (peak coordinate: x = -52, y = -50, z = -2, T(44) = 5.16, P =.006, k = 219), left inferior frontal gyrus (peak coordinates: x = -38, y = 40, z = 0, T(44) = 4.82, P = .048, k = 137), and right middle frontal gyrus (peak coordinates: x = 42, y = 16, z = 40, T(44) = 4.15, P = .015, k = 181). Again, the larger an individual’s recalibration value, the stronger the association between trial type and neural activation during the closeness ratings phase in the treatment condition as compared to the control condition in these regions.

Across both conditions (treatment as well as control), neural activation was stronger in reinforced compared to non-reinforced trials in right dorso-medial prefrontal cortex (peak coordinates: x = 18, y = 54, z = 32, T(44) = 4.64, P < .001, k = 488), right putamen (peak coordinates: x = 30, y = 2, z = -8, T(44) = 5.03, P = .047, k = 131), bilateral inferior frontal gyrus/ anterior insula (left hemisphere peak coordinates: x = -36, y = 28 z = -4, T(44) = 4.95, P = .098, k = 106; right hemisphere peak coordinated: x = 42, y = -28, z = -12, T(44) = 4.49, P < .001, k = 405), amongst other regions (see **Table S2.1.4** for full results with cluster-size k >100 and p < .001 uncorrected).

**Table S2.1.4.** Results of the second-level analysis with the contrast parametric modulator trialtype > control during the closeness rating phase. This analysis shows which neural regions are more active during reinforced compared to non-reinforced trials in all blocks and conditions. P<.001 uncorrected, k = 20.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| region | hemisphere | T | P(cluster-level) | k | coordinates |
| cerebellum and fusiform gyrus | left | 6.35 | <.001 | 1373 | -26 -80 -34 |
|  | right | 5.72 | <.001 | 348 | 24 -78 -34 |
| temporal pole |  |  |  |  |  |
|  | right | 5.81 | .009 | 191 | 52 6 -36 |
| occipital pole | right | 5.26 | .005 | 213 | 20 -98 16 |
| putamen | right | 5.03 | .047 | 131 | 30 2 -8 |
| inferior frontal gyrus /anterior insula | left | 4.95 | .098 | 106 | -32 28 -4 |
|  | right | 4.49 | <.001 | 405 | 42 28 -12 |
| dorso medial prefrontal cortex | right | 4.64 | <.001 | 488 | 18 54 22 |
| middle temporal gyrus | left | 4.36 | .075 | 115 | -56 -34 -6 |
| caudate | right | 4.13 | .010 | 185 | 8 12 4 |

Interestingly, across both conditions (treatment as well as control), the extent of individual recalibration was associated with increased neural activation in non-reinforced as opposed to reinforced trials in bilateral dorso-medial/ventro-medial prefrontal cortex (dmPFC/vmPFC, left hemisphere peak coordinates: x = 6, y = 48, z = 16, right hemisphere peak coordinates: x = -4, y = 56, z = 8, T(44) = 4.62, P < .001, k = 395) and right middle frontal gyrus (peak coordinates: x = 42, y = 2, z = 58, T(44) = 4.48, P = .013, k = 176). Thus, the more participants recalibrated the feedback value, the stronger the neural activation in response to non-reinforced as compared to reinforced trials in the vmPFC and middle frontal gyrus. The reverse contrast did not yield any significant neural activation (all Ps>.993).