Empathy incites a sustainable prosocial decisions bias

Anne Saulin & Grit Hein

Supplementary material

# Supplementary Figures

**Diagram

Description automatically generated**

**Figure S2.2.1.** Trial structure and behavioral results of the empathy activation blocks in the fMRI study and the laboratory replication study. **A** Each of the 25 trials per block started with a fixation cross, followed by the rating of how close participant(green mannequin) felt to the respective interaction partner (here blue mannequin). After another fixation period, participants saw a fully filled flash which symbolized that the partner now received a painful stimulation (high pain trial) or they saw a partly filled flash symbolizing non-painful stimulation of the partner (no pain trial). Following a fixation, participants then rated how they felt in response to the other’s pain or no pain on a continuous ratings scale ranging from *very bad* to *very good*. **B** In the fMRI study, participants ratings of social closeness increased over the course of the first block in the control (light blue) as well as the treatment condition (dark blue), and plateaued in block, again in both conditions. **C** This pattern was replicated in the behavioral replication study conducted in the laboratory.

**Diagram

Description automatically generatedFigure S2.2.2.** Trial structure and behavioral results of the reciprocity activation blocks. **A** Each of the 25 trials per block started with a fixation cross, followed by the rating of how close participant (green mannequin) felt to the respective interaction partner (here red mannequin). After another fixation period, participants saw that the partner decided to forgo a monetary reward in order to reduce the amount of painful stimulation for the participant (help trial) or they saw that the partner had decided to take the money and thus did not reduce the participants painful stimulations (no help trial). Following a fixation, participants then rated how they felt in response to the other’s help or no help on a continuous ratings scale ranging from *very bad* to *very good*. **B** In the treatment condition, social closeness ratings increased over the course of block 1 and strongly decreased in block 2. In the control condition, social closeness stayed at a relatively constant level throughout both blocks.

Chart, scatter chart

Description automatically generated

**Figure S2.2.3.** Estimated psychometric functions for empathy-based decisions in the treatment and the control condition. Vertical dashed lines denote the points of subjective equality (PSEs) for the control condition (light blue, mean = 123 points, Inf = 110 points, Sup = 138 points) and the treatment condition (dark blue, mean = 118 points, Inf = 106 points, Sup = 130 points). Thus, PSEs were comparable across the two conditions.

Chart, scatter chart

Description automatically generated

**Figure S2.2.4.** Estimated psychometric functions for reciprocity-based decisions in the treatment and the control condition. Vertical dashed lines denote the points of subjective equality (PSEs) for the control condition (light red, mean = 221 points, Inf = 207 points, Sup = 236 points) and the treatment condition (dark red, mean = 208 points, Inf = 195 points, Sup = 222 points). Thus, PSEs were comparable across the two conditions.

# Supplementary Tables

**Table S2.2.1**. DIC (deviance information criterion) values for all models tested for the empathy motive (data pooled across the fMRI study and the laboratory replication study) and the reciprocity motive (laboratory control study). The winning model is highlighted in bold font

|  |  |  |  |
| --- | --- | --- | --- |
| motive | model | model specification | DIC |
| empathy | M0 | - | 31672.26 |
|  | V1 | *v ~ other possible gain* | 29450.93 |
|  | V2 | *v ~ other possible gain + condition* | 29346.74 |
|  | V3 | *v ~ other possible gain + condition +block number* | 29301.53 |
|  | VZ1 | *v ~ other possible gain*  *z ~ condition* | 29184.64 |
|  | VZ2 | *v ~ other possible gain*  *z ~ condition + block number* | 29010.1 |
|  | VZ3 | *v ~ other possible gain*  *z ~ condition\*block number* | **28963.7** |
| reciprocity | M0 | - | 14655.38 |
|  | V1 | *v ~ other possible gain* | 13496.07 |
|  | V2 | *v ~ other possible gain + condition* | 13370.4 |
|  | V3 | *v ~ other possible gain + condition +block number* | 13157.64 |
|  | VZ1 | *v ~ other possible gain*  *z ~ condition* | 13275.5 |
|  | VZ2 | *v ~ other possible gain*  *z ~ condition + block number* | 12958.59 |
|  | VZ3 | *v ~ other possible gain*  *z ~ condition\*block number* | **12782.74** |

**Table S2.2.2.** Results of the second-level analysis showing the main effect of block number during the decision process. P<.001 uncorrected, k = 100.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| region | hemisphere | T | P  (peak-level) | P  (cluster-level) | k | coordinates |
| striatum | left | 5.88 | .001 | <.001 | 814 | -20 10 -10 |
| right | 5.08 | .030 | <.001 | 487 | 18 8 6 |
| inferior frontal gyrus | left | 5.49 | .006 | .022 | 174 | -46 8 22 |
|  | right | 3.80 | .934 | .043 | 147 | 46 6 34 |
| temporo-parietal junction | left | 4.67 | .138 | <.001 | 742 | -38 -36 18 |
| postcentral gyrus | right | 4.24 | .495 | .014 | 191 | 28 -34 48 |
| supplementary motor area | right | 4.08 | .687 | <.001 | 347 | 4 -4 60 |
| superior temporal gyrus | right | 4.26 | .462 | .142 | 103 | 56 -4 3 |
| anterior insula | right | 4.05 | .711 | .045 | 146 | 38 -2 12 |
| precentral gyrus | left | 4.57 | .192 | .048 | 143 | -34 -16 52 |
| left | 4.07 | .688 | .114 | 111 | -58 2 36 |
| right | 5.33 | .011 | <.001 | 1778 | 18 -96 20 |
| right | 4.88 | .065 | <.001 | 874 | 38 -10 60 |
| fusiform gyrus | left | 4.23 | .500 | .154 | 100 | -38 -84 -12 |
| anterior cingulate gyrus | left | 4.09 | .668 | .142 | 103 | -10 36 12 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| model | χ² | P-value | β | SE |
| (Intercept) | 0.4223 | 0.5162 | -0.073 | 0.112 |
| strength | 3.5660 | 0.0599. | -0.213 | 0.113 |
| effect | 3.0975 | 0.0793. | 0.278 | 0.158 |
| regions | 0.3538 | 0.5524 | 0.094 | 0.158 |
| indicators | 0.0000 | 1.0000 | 0.000 | 0.158 |
| strength:effect | 3.1580 | 0.0765. | 0.284 | 0.160 |
| strength:regions | 0.0407 | 0.8403 | 0.032 | 0.160 |
| effect:regions | 0.3864 | 0.5346 | -0.139 | 0.223 |
| strength:indicators | 4.9798 | 0.0263\* | 0.346 | 0.155 |
| effect:indicators | 0.0000 | 1.0000 | 0.000 | 0.223 |
| regions:indicators | 0.0000 | 1.0000 | 0.000 | 0.223 |
| strength:effect:regions | 0.0227 | 0.8804 | -0.034 | 0.226 |
| strength:effect:indicators | 2.9689 | 0.0858. | -0.378 | 0.219 |
| strength:regions:indicators | 0.0053 | 0.9420 | 0.016 | 0.219 |
| effect:regions:indicators | 0.0000 | 1.0000 | 0.000 | 0.316 |
| strength:effect:regions:indicators | 0.0002 | 0.9902 | -0.004 | 0.310 |

**Table 2.2.3** Results of the linear model with the predictors strength (exent of individual decision bias), effect (initial vs. sustained), regions (TPJ vs. mPFC), indicators (initial increase in *z*-parameter vs. sustained increase in *z*-parameter), and their interaction and the dependent variable increases in neural activation from block 1 to block2 (initial) and from block 2 to block 3 (sustained). TPJ = temporo-parietal junction, mPFC = medial prefrontal cortex. P-values are derived from the type 3 Wald χ²-test.