**Comparisons between the four experimental conditions**

We performed TOST Paired Samples T-Test to examine potential differences and equivalences between conditions (tab. S1) by reversing the roles of H0 and H1, taking into account that the data are paired, i.e. the same focal rats were tested repeatedly (1). This test is based on a parametric t-test, whereas our data were not normally distributed. However, based on our sample size this assumption violation does not impact the detection rate of this test (2). Further, we validated this approach by comparing our results to a permutated Paired Samples T-Test (R package: RVAideMemoire, code: perm.t.test with 5000 permutations). The results did not deviate from the results obtained with the TOST Paired Samples T-Tests, which seems therefore a valid test for our data (3).

This TOST Paired Samples T-Test requires defining a threshold above which two conditions are considered as equal. A medium sized effect size is d= 0.5, i.e. two groups differ by half the standard deviation, which is commonly used for equivalence tests (4). Hence, we set the upper and lower equivalence boundaries symmetrically to ± 0.5. If the true effect (± 90% confidence interval) was within this boundary, it was considered to be equal. To test whether the effect is within the boundaries, two one-sided tests are conducted. If both one-sided tests are significant, H0 can be rejected, i.e. the conditions can be considered to be equal.

1. Lakens D. Equivalence tests: a practical primer for t-tests, correlations, and meta-analyses. Soc Psychol Personal Sci. 2017;8(4):355–62.

2. Mara CA, Cribbie RA. Paired-samples tests of equivalence. Commun Stat Simul Comput. 2012;41(10):1928–43.

3. Konietschke F, Pauly M. Bootstrapping and permuting paired t-test type statistics. Stat Comput (2014) 24:283–96.

4. Lakens D, Scheel AM, Isager PM. Equivalence testing for psychological research: a tutorial. Adv Methods Pract Psychol Sci. 2018;1(2):259–69.

**Tab. S1 Comparisons between the four experimental conditions**

|  |  |  |  |
| --- | --- | --- | --- |
| Condition | Dissimilarity | Similarity | |
| TOST upper boundary | TOST lower boundary |
| C vs. D | t20= 2.59, ***p*= 0.02** | t20= -0.30, *p*= 0.61 | t20= 4.88, *p*< 0.001 |
| DDDC vs. CCCD | t20= 2.29, ***p*= 0.03** | t20= -0.01, *p*= 0.50 | t20= 4.58, *p*< 0.001 |
| D vs CCCD | t20= 1.56, *p*= 0.14 | t20= -0.74, *p*= 0.24 | t20= 3.85, *p*< 0.001 |
| C vs. DDDC | t20= 1.85, *p*= 0.08 | t20= -0.44, *p*= 0.33 | t20= 4.14, *p*< 0.001 |

Focal rats experienced four partners that were experienced once as cooperators (C) and defectors (D), as well as four times as overall cooperators that were defecting in the last round (CCCD) and overall defectors that cooperated in the last round (DDDC). Comparisons were made using the TOST Paired Samples T-Test, testing for differences and similarities between conditions.