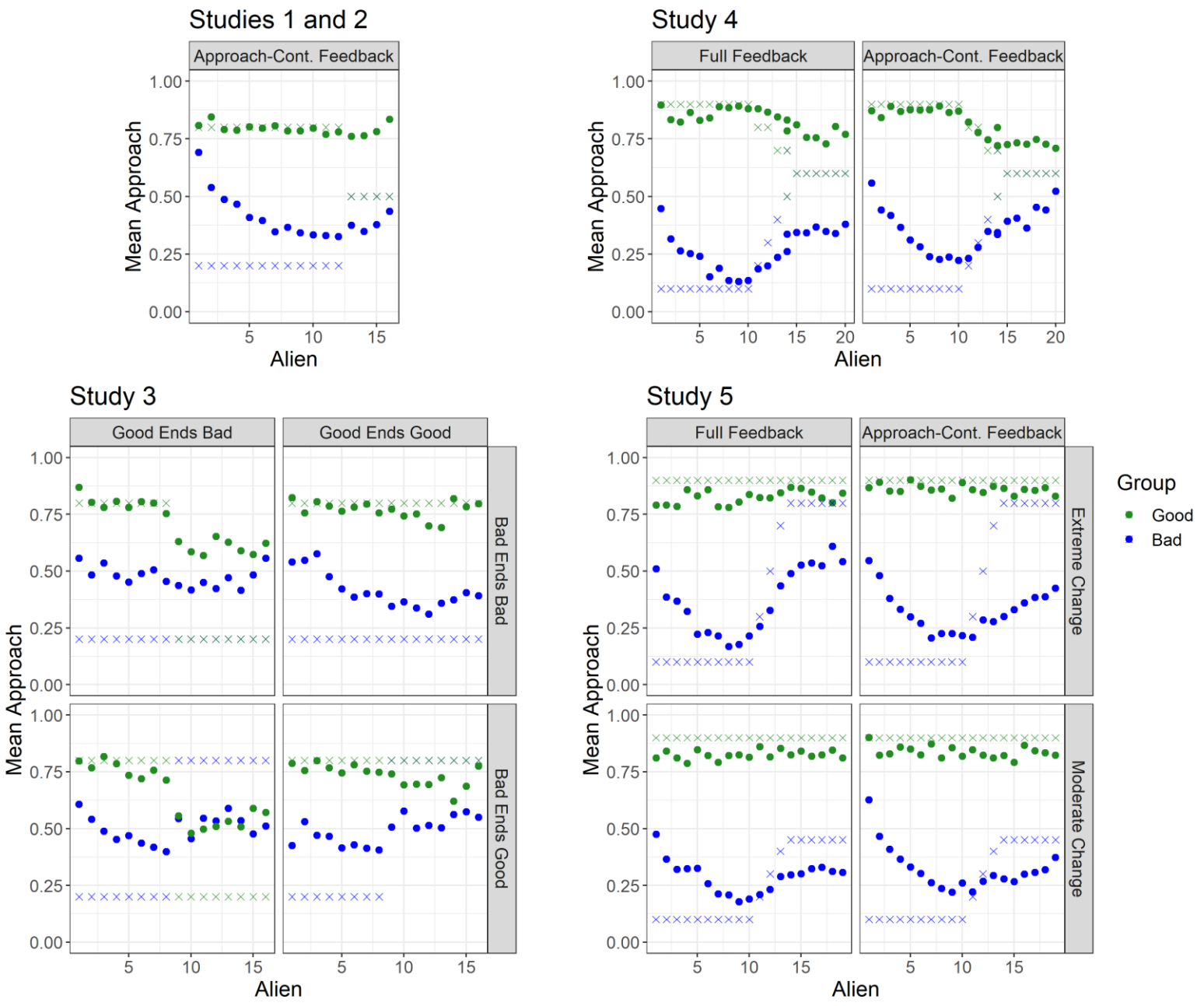
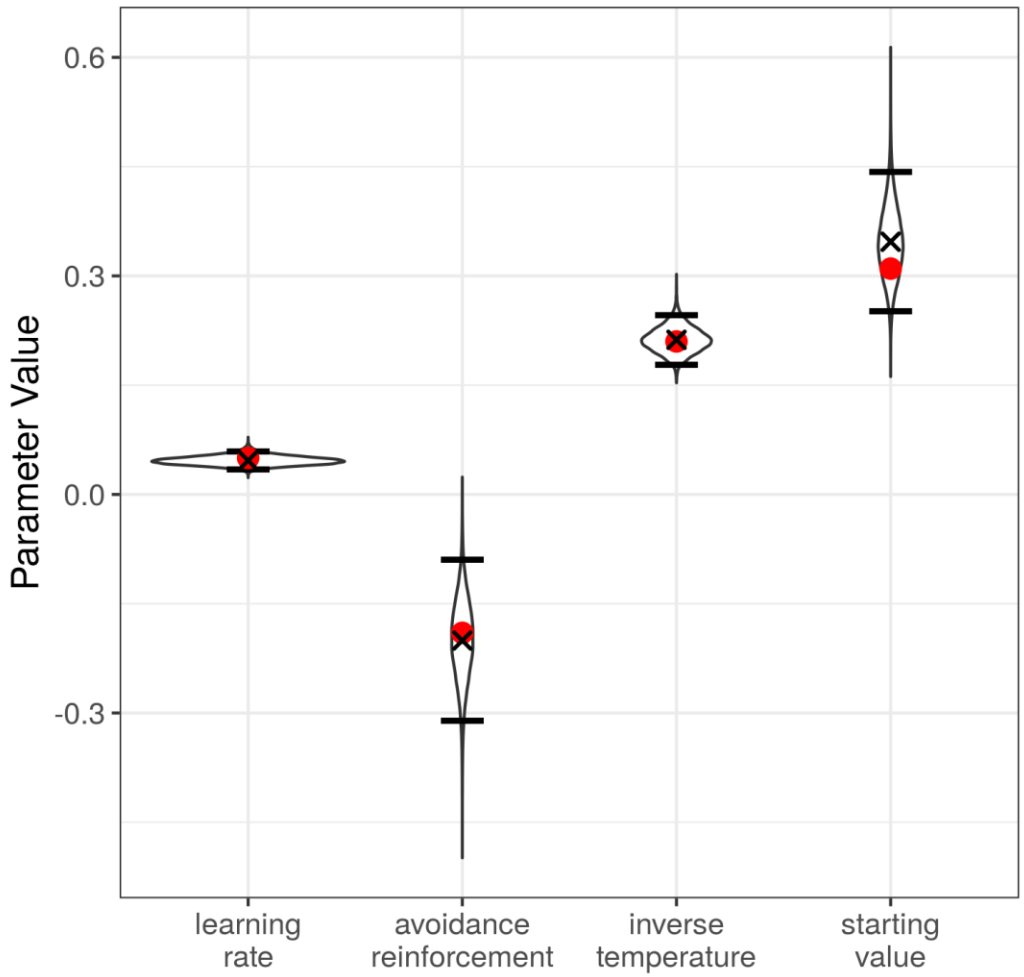
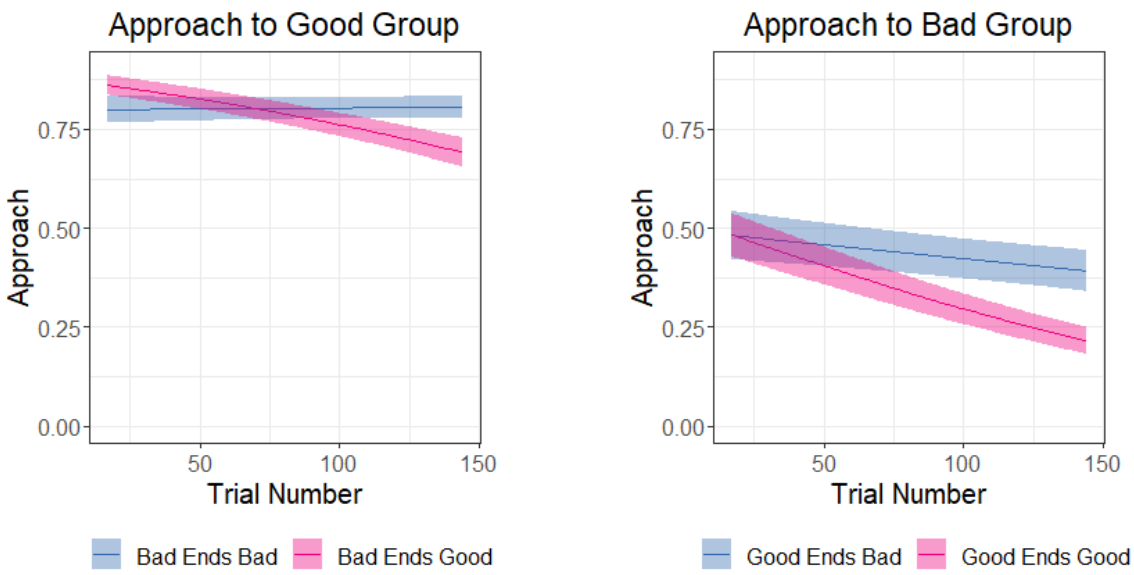


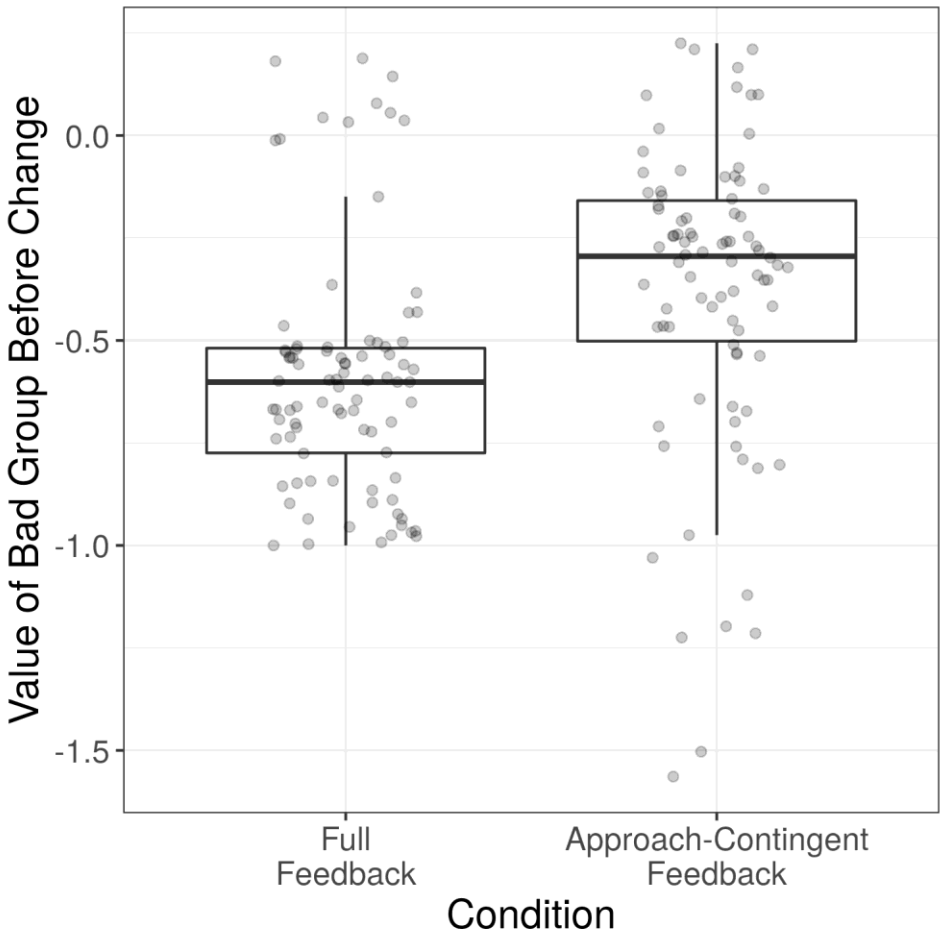
Supplementary Figure 1. Approach behaviour towards each alien in all studies. Between 16 and 20 aliens per group are encountered in each study. Aliens are numbered in the order they appear in the task, with each alien encountered multiple times throughout the experiment. Dots represent the average participant response to each alien, coloured by group, and crosses represent the alien's true probability of cooperating with the participant.



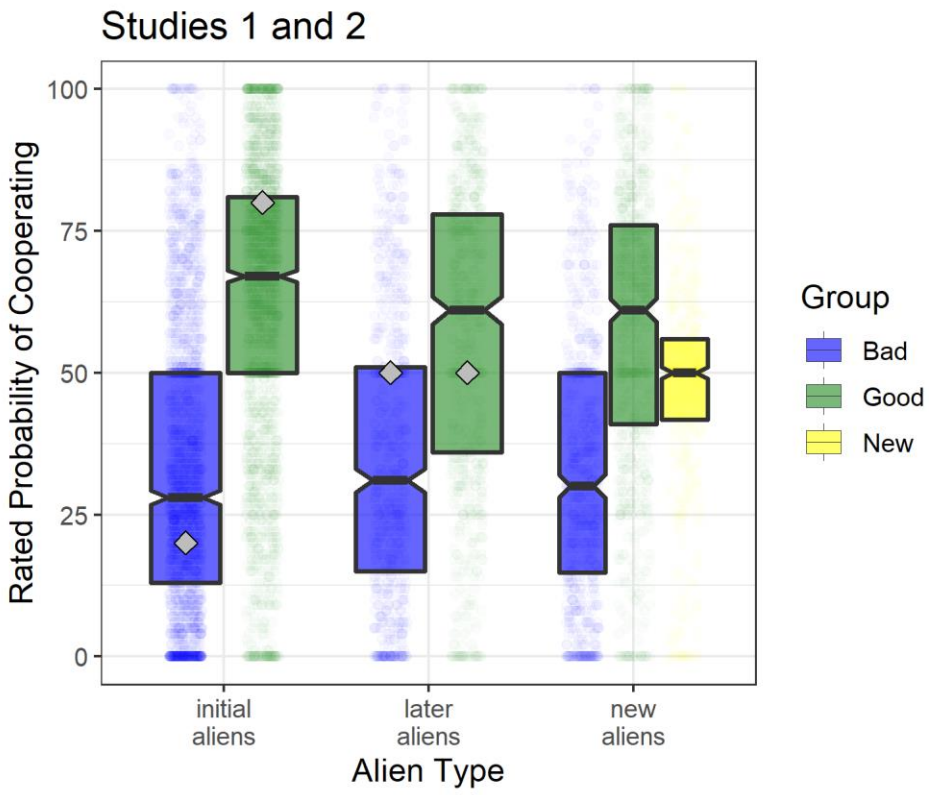
Supplementary Figure 2. Results of parameter recovery simulations (study 1). The best-fit group parameters and model from study 1 were used to simulate data for each subject, and the model was then fit to the simulated data to see how accurately the group parameters could be recovered. Red dots represent true parameters, black crosses represent modal recovered parameters, and black horizontal lines represent 95% HDIs around the recovered parameters. To aid in model convergence by keeping all parameters on a similar scale, the value of inverse temperature parameter shown here was multiplied by 20 in order to get the true value of the parameter that was used in the model.



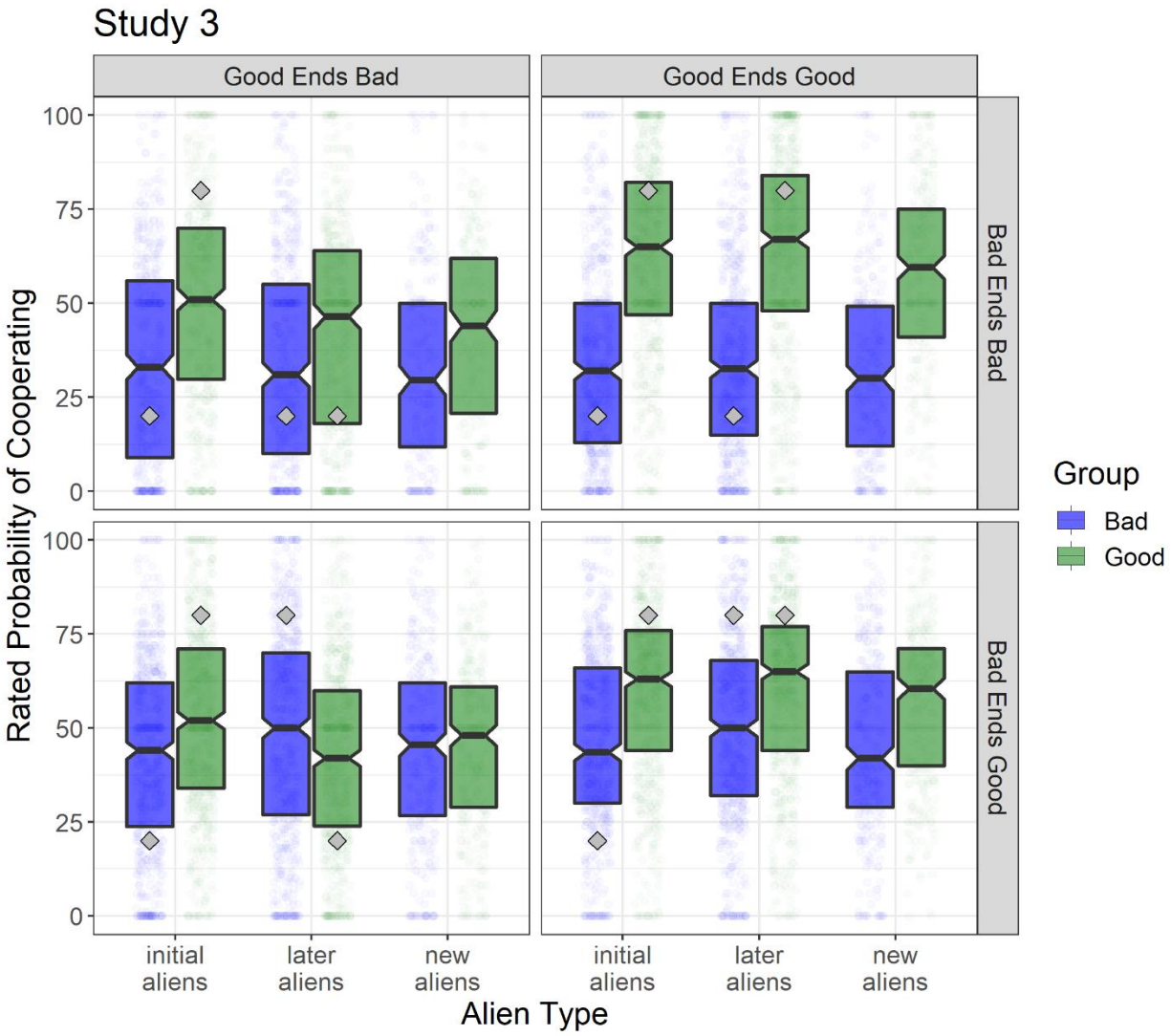
Supplementary Figure 3. Effect of changes in one group on approach to the other (study 3). The left panel shows approach to members of the initially cooperative ("good") group over time when there has been no change in this group, as a function of whether or not the initially uncooperative ("bad") group has changed. The right panel shows approach to members of the bad group over time when the bad group has not changed, as a function of whether the good group has changed. Both panels show that even in the absence of any changes in the group, participants adjust their behaviour towards one group when the behaviour of the other group changes.



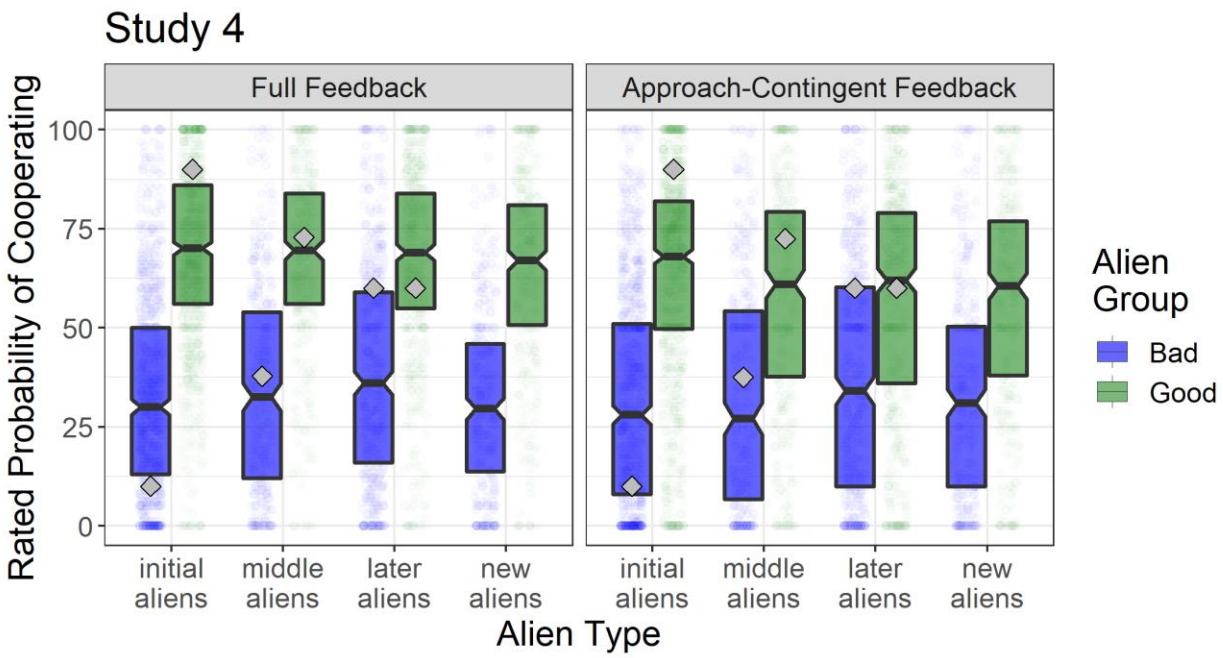
Supplementary Figure 4. Value of bad group before change (study 4). Values associated with the initially uncooperative alien group in the two feedback conditions before the group has started to change to become better.



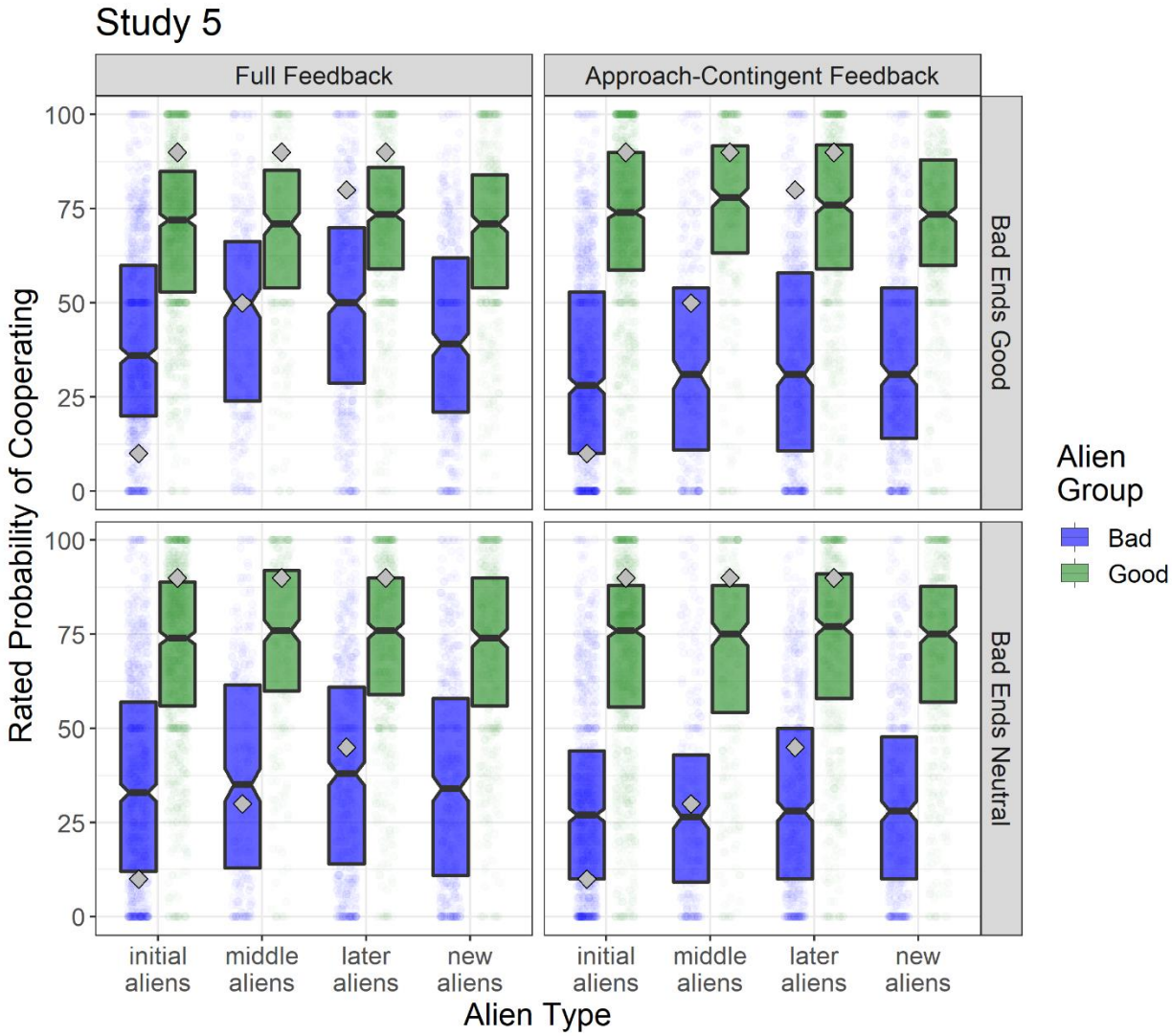
Supplementary Figure 5. Post-game ratings (studies 1 and 2). Particpants' ratings of how likely each alien was to cooperate with them after finishing the task. Aliens are separated according to whether they appeared earlier in the task and cooperated differently across groups, appeared later in the task and cooperated equally across groups, or did not appear in the task at all. Diamonds represent the aliens' true probability of cooperating, which differs across groups for the initial aliens but not for the aliens introduced towards the end of the game.



Supplementary Figure 6. Post-game ratings (study 3). Particpants' ratings of how likely each alien was to cooperate with them after finishing the task. Aliens are separated according to whether they appeared earlier in the task and had their group's initial rate of cooperation (high for the good group and low for the bad group), appeared later in the task and had their group's new rate of cooperation (which differed depending on condition), or did not appear in the task at all. Diamonds represent the aliens' true probability of cooperating in the different conditions.



Supplementary Figure 7. Post-game ratings (study 4). Particpants' ratings of how likely each alien was to cooperate with them after finishing the task. Diamonds represent the aliens' true probability of cooperating in the different conditions. Aliens are divided according to whether they were present at the start of the task and had the initial probability of cooperating for their group, introduced somewhere in the middle of the task and had a probability of cooperating in between the start and end probabilities, present at the end of the task and had the new neutral probability of cooperating, or not seen in the task at all.



Supplementary Figure 8. Post-game ratings (study 5). Particpants' ratings of how likely each alien was to cooperate with them after finishing the task. Diamonds represent the aliens' average true probability of cooperating in the different conditions. Aliens are divided according to whether they were present at the start of the task and had the initial probability of cooperating for the bad group, introduced somewhere in the middle of the task and had a probability of cooperating in between the start and end probabilities, present at the end of the task and had the new probability of cooperating, or not seen in the task at all.

Type	Study	Model	Estimate (same model)	AIC	BIC
robust regression	1	no avoidance parameter	32.35	1494.55	1510.40
		one avoidance parameter	-15.84	1511.01	1526.86
		two avoidance parameters	21.93	1508.80	1524.65
	2	no avoidance parameter	49.46	1690.63	1707.12
		one avoidance parameter	3.05	1713.38	1729.88
		two avoidance parameters	-6.24	1716.57	1733.06
	3	no avoidance parameter	29.57	5115.48	5137.45
		one avoidance parameter	-1.15	5234.28	5256.25
		two avoidance parameters	3.13	5219.78	5241.75
	4	no avoidance parameter	49.77	2972.09	2991.32
		one avoidance parameter	-2.08	3019.88	3039.11
		two avoidance parameters	3.7	3008.40	3027.63
	5	no avoidance parameter	24.74	7014.69	7038.17
		one avoidance parameter	-2.79	7055.25	7078.73
		two avoidance parameters	9.55	7057.49	7080.98
removing outliers	1	no avoidance parameter	24.96	1427.41	1443.03
		one avoidance parameter	-10.88	1433.05	1448.67
		two avoidance parameters	19.65	1435.52	1451.14
	2	no avoidance parameter	47.84	1674.74	1691.18
		one avoidance parameter	8.07	1693.11	1709.55
		two avoidance parameters	-12.97	1697.09	1713.54
	3	no avoidance parameter	26.4	4914.40	4936.18
		one avoidance parameter	-14.66	4951.01	4972.79
		two avoidance parameters	22.31	4944.08	4965.86
	4	no avoidance parameter	40.56	2913.90	2933.03
		one avoidance parameter	4.69	2943.27	2962.40
		two avoidance parameters	6.43	2926.54	2945.67
	5	no avoidance parameter	18.93	6859.88	6883.26
		one avoidance parameter	1.24	6888.05	6911.43
		two avoidance parameters	11.89	6869.02	6892.40

Supplementary Table 1. Additional comparisons of the ability of the three models to predict participants' ratings.

Two additional sets of models were run to ensure that the negative skewness of the values from the models with avoidance parameters did not account for the results. These included 1) using robust regression models that minimize the influence of outliers, and 2) removing values that were below -2 to

ensure that negative outliers were not driving the effect. For each method, we compare the three models by examining which produces the highest estimate when entered into the same regression model, as well as which produced the lowest AIC and BIC values when entered into separate regression models. Across most of these transformations, the model without an avoidance parameter generally provides the best prediction of participants' ratings, lending further support to the results of the main text and suggesting that the results are not simply driven by negative outliers.

Supplementary Results

In addition to analyzing participants' approach behaviour in each study, we also examined their explicit ratings of how likely each alien was to cooperate with them. Participants rated each alien that they saw in the task as well as new aliens from each group who were not in the task. Participants in study 2 also rated members of an entirely new group of aliens.

For each study, we examine whether the biases that emerge in approach behaviour are also reflected in participants' explicit beliefs about how cooperative each alien was. Ratings of aliens who appeared in the task were predicted from the alien's group membership while controlling for the alien's actual probability of cooperating with the participant. Controlling for the real probability allows us to examine whether participants are responding to aliens differently based on their group membership even after any real differences between the groups are taken into account. If participants are simply responding appropriately to the real differences between individuals in the two groups, the alien's group membership should not affect ratings when controlling for the alien's real probability of cooperating. If, on the other hand, group membership shapes participants' behaviour even after taking into account any real differences between the two groups, the alien's group should affect responses even after controlling for the real probabilities. For studies that have between-subjects condition, we also interact the alien's group membership with condition.

Study 1. Even when controlling for the real cooperation rates, participants rated aliens in the initially cooperative group as more likely to cooperate than those in the initially uncooperative group, $b = -10.14$, $\chi^2(1) = 115.71$, $p < .001$. These biases also extended to new aliens who did not appear in the game and who differed randomly along various facial features from the aliens in the game: participants rated new members of the initially cooperative group as more cooperative than new members of the initially uncooperative group, $b = -12.03$, $\chi^2(1) = 176.27$, $p < .001$ (see Supplementary Figure 5). These

results suggest that early beliefs about the two groups generalize both to the later individuals who cooperated equally and to new members of the groups who had never been seen before.

Study 2. Participants rated members of the initially cooperative group who appeared in the task as more cooperative than members of the initially uncooperative group even after controlling for the alien's real probability of cooperating, $b = -11.74$, $\chi^2(1) = 193.15$, $p < .001$. Furthermore, these biases extended to aliens who were not seen during the game at all, with participants rating new members of the initially cooperative group as more cooperative than new members of the initially uncooperative group, $b = -13.33$, $\chi^2(1) = 259.49$, $p < .001$. As in study 1, these results replicate the findings on approach behaviour, indicating that initial experiences with members of the two groups generalized to participants' explicit beliefs about later group members.

To examine how participants generalize from their knowledge of the two groups they encountered in the task to an entirely new group, we examined participants' ratings of the third group of aliens who had not been seen before. The average of each participant's ratings of members of the initially uncooperative group and the initially cooperative group were taken, and a paired samples t-test was conducted to compare these average scores to each participant's average ratings of the members of the new group. If ratings of the new group are approximately in between ratings of the initially cooperative and uncooperative groups, this analysis should reveal no systematic differences between the ratings of the new group and the average ratings of these two groups. Indeed, this t-test indicates there are no significant differences between these ratings, $t(99) = -1.16$, $p = .250$ (see Supplementary Figure 5). To determine whether there was evidence for the null hypothesis that ratings of the new group are equal to the average ratings of the initially cooperative and initially uncooperative groups, we computed a Bayes Factor for this difference using the BayesFactor package (Morey & Rouder, 2018), which uses a Cauchy prior with a scale of √2/2 on the standardized effect size. This analysis revealed moderate support for the null hypothesis, $BF_{10} = 0.21$. Thus, it seems that there is no evidence for a

positivity or negativity bias in participants' generalizations to an entirely new group of aliens, suggesting that neither positive nor negative behaviours are considered the "default" after learning, but contribute equally to beliefs.

Study 3. Participants rated the initially cooperative group as more cooperative than the initially uncooperative group, $b = -4.90$, $\chi^2(1) = 69.66$, $p < .001$, and this depended on whether or not each group changed over the task. Specifically, participants rate the initially cooperative group as less cooperative when its members become uncooperative over the task, $b = 2.32$, $\chi^2(1) = 16.50$, $p < .001$, and they rate the initially uncooperative group as more cooperative when its members become cooperative, $b = -2.31$, $\chi^2(1) = 16.28$, $p < .001$. Extending this analysis to new members of the two groups who were not seen during the game indicates a similar pattern, with participants rating the initially cooperative group as more cooperative than the initially uncooperative group, $b = -5.83$, $\chi^2(1) = 84.23$, $p < .001$, with interactions with whether or not the initially cooperative group changes, $b = 3.36$, $\chi^2(1) = 27.98$, $p < .001$, and whether or not the initially uncooperative group changes, $b = -2.81$, $\chi^2(1) = 19.54$, $p < .001$, suggesting that they picked up on changes in the groups (see Supplementary Figure 6). These findings parallel the results on approach behaviour, suggesting that participants are able to pick up on the drastic changes in the groups, although their final ratings of each group still do not fully reflect the later group members' actual probabilities of cooperating.

Study 4. Participants rated the initially cooperative group as more cooperative than the initially uncooperative group even after controlling for the real probabilities of cooperating, $b = -11.72$, $\chi^2(1) = 109.26$, $p < .001$. A main effect of condition was also present, with participants in the full feedback condition providing slightly higher ratings than those in the approach-contingent feedback condition, $b = 2.20$, $\chi^2(1) = 8.46$, $p = .004$, but no interaction between alien group and feedback condition, $b = -1.62$, $\chi^2(1) = 2.29$, $p = .13$. Extending this analysis to new aliens who did not appear in the game does reveal a main effect of group, $b = -13.73$, $\chi^2(1) = 148.04$, $p < .001$, and an interaction between group and

condition, with participants in the full feedback condition rating members of the initially cooperative group as slightly more cooperative than participants in the approach-contingent feedback condition, b = -2.35, $\chi^2(1) = 4.32$, $p = .038$ (see Supplementary Figure 7). Thus, it seems that participants' explicit beliefs about the groups reflect generalization based on early experiences with group members. There seems to be mixed evidence for the role of feedback type on these beliefs, with greater generalization in the full feedback condition when rating new aliens who were not seen in the task, but not when rating aliens present in the task.

Study 5. Participants rated the initially cooperative group as more cooperative than the initially uncooperative group even after controlling for any real differences in the probabilities of cooperating, b = -14.48, $\chi^2(1) = 322.70$, $p < .001$, and this difference was more pronounced for participants in the approach-contingent feedback condition, $b = 1.99$, $\chi^2(1) = 6.75$, $p = .009$. Extending this analysis to new aliens who did not appear in the game produces similar results, with participants rating new members of the initially cooperative group as more cooperative than new members of the initially uncooperative group, b = -16.82, $\chi^2(1) = 413.25$, $p < .001$, with a more pronounced difference for those in the approach-contingent feedback condition, $b = 1.70$, $\chi^2(1) = 4.20$, $p = .04$ (see Supplementary Figure 8).

References

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes Factors for common designs. R package version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor