

Coordination over a unique medium of exchange under information scarcity

Aurélien Nioche,^{1,2,3,4,5,†,*} Basile Garcia,^{6,7,4,5,†} Germain Lefebvre,⁶ Thomas Boraud,^{4,5,8} Nicolas P. Rougier,^{7,4,5,9,‡} and Sacha Bourgeois-Gironde^{10,2,3,‡}

¹Aalto University, School of Electrical Engineering, Department of Communications and Networking, 02150 Espoo, Finland; ²Institut Jean Nicod, Département d'Etudes Cognitives, ENS, EHESS, PSL Research University, 75005 Paris, France; ³Institut Jean Nicod, CNRS, UMR 8129; ⁴Institut des Maladies Neurodégénératives, Université de Bordeaux, 33000 Bordeaux, France; ⁵Institut des Maladies Neurodégénératives, CNRS, UMR 5293; ⁶Laboratoire de Neurosciences Cognitives Computationnelles, Département d'Etudes Cognitives, ENS, PSL Research University, 75005 Paris, France; ⁷Laboratoire de Neurosciences Cognitives Computationnelles, INSERM, U960; ⁸Inria Bordeaux Sud-Ouest, 33405 Talence, France; ⁹Centre Expert Parkinson, CHU Bordeaux, 33000 Bordeaux, France; ¹⁰LaBRI, Université de Bordeaux, INP, CNRS, UMR 5800, 33405 Talence, France; ¹¹Laboratoire d'Economie Mathématique et de Microéconomie Appliquée, Université Panthéon Assas, 75006 Paris, France.

[†]These authors have contributed equally to this work.

[‡]These authors have also contributed equally to this work.

*Corresponding author: nioche.aurelien@gmail.com

Supplementary material

Model

Valuation functions for $G = 3$. We provide here the equations for the value function for each of the options associated with the two types of decisions that an agent can face when $G = 3$ (the process would be similar for any value of G).

Let i be the production good, j , the consumption good and k , the medium of exchange for the agent under consideration, we have:

$$\begin{aligned} v(ij) &= 1/(1 + \beta)^{\Delta_{ij}} = 1/(1 + \beta)^{\delta_{ij}} = 1/(1 + \beta)^{1/e_{ij}} \\ v(ik) &= 1/(1 + \beta)^{\Delta_{ik}} = 1/(1 + \beta)^{\delta_{ik} + \delta_{kj}} = 1/(1 + \beta)^{(e_{ik} + e_{kj})/(e_{ik} \cdot e_{kj})} \\ v(kj) &= 1/(1 + \beta)^{\Delta_{kj}} = 1/(1 + \beta)^{\delta_{kj}} = 1/(1 + \beta)^{1/e_{kj}} \\ v(ki) &= 1/(1 + \beta)^{\Delta_{ki}} = 1/(1 + \beta)^{\delta_{ki} + \delta_{ij}} = 1/(1 + \beta)^{(e_{ki} + e_{ij})/(e_{ki} \cdot e_{ij})} \end{aligned}$$

Simulations

Parameter recovery. To assess the quality of our decision-making model, we fitted the data generated by artificial agents using subjects best-fit parameters, expecting to retrieve as new best-fit parameter values the ones used to produce the data. We computed for each parameter a Pearson correlation coefficient. The Fig. S6 show the results and the Table S6 summarizes the statistics. The results show that our parameter optimization procedure correctly retrieve with a good accuracy the values of the learning rate parameter α ($r_{pearson} = 0.88$, $p < 0.001^*$, $n = 166$) as well as the values of the exploration-exploitation parameter γ ($r_{pearson} = 0.91$,

$p < 0.001^*$, $n = 166$). However, the procedure did not allow us to accurately retrieve the values of the discount factor parameter, β ($r_{pearson} = 0.01$, $p = 0.920$, $n = 166$), indicating that this parameter has a low impact of behavior.

Sensitivity analysis. In order to assess to what extent the results obtained through simulation were dependent on the free parameters (and possibly more on the value of the free parameters than on the variables we were manipulating), we conducted a sensitivity analysis on the simulated data used for producing the Fig. ???. Considering separately the economies with 3 goods in circulation and 4 goods in circulation, we computed for each parameter value, the frequency of use of the good 1 as a medium of exchange. The statistical relevancy has been assessed using a Kruskal–Wallis one-way analysis of variance (Bonferroni corrections for multiple comparisons are applied). The results are reported in Fig. S5 and Table S5 summarizes the statistics.

The learning rate (α) has an impact on the frequency of use of a medium of exchange for economies with 3 goods in circulation ($H = 76.17$, $p < 0.001^*$, $n = 10\,800$) and economies with 4 goods in circulation ($H = 45.50$, $p = 0.001^*$, $n = 10\,800$). The exploitation-exploration rate (γ) has an impact also on the frequency of use of a medium of exchange with three goods in circulation ($H = 19.71$, $p < 0.001^*$, $n = 10\,800$), but not with four goods in circulation ($H = 0.69$, $p = 1.000$, $n = 10\,800$). The discount rate has no significant impact neither in economies with 3 goods in circulation ($H = 0.01$, $p = 1.000$, $n = 10\,800$), not in economies with 4 goods in circulation ($H = 0.02$, $p = 1.000$, $n = 10\,800$). The fact that the discount rate has no significant impact on the use of a medium of exchange corroborates with the fact that we did not accurately retrieve the values of the discount factor parameter during the parameter recovery procedure.

Experiments

Training. Each subject went through an interactive tutorial, teaching him the rules and controls for the experiment. In order to get comfortable with the interface, he was then told to play a game with random-playing artificial players for 20 time-steps.

Subject exclusion from analysis. For both experiments I and II, no subject has been excluded from the analysis.

Control for the a priori preferences. In order to control a priori preferences for a certain type of good (e.g., having wheat is seen as preferable because eatable), each subject sees himself as a producer of wood, consumer of wheat, regardless if he is a producer of the abstract good 1, 2, 3 or 4.

As a consequence, it also puts a strong limit to the possibility of ‘cheating’ between subjects, by trying either to mimicking the neighbor’s behavior or creating tacit or explicit collusion.

Parametrization. Table S1 and Table S2 summarize the parameters used for the experiment I and the experiment II respectively.

Statistical tests. Due to the boundaries of our measure (it is normalized between 0 and 1), the observations could not be normally distributed. Thus, we used a non-parametric test for independent groups, the Mann–Whitney U test. For both experiments I and II, Bonferroni corrections are used because we do multiple testing across independent groups, knowing that the probability of rejecting the null hypothesis by error inflates for multiple tests and that Bonferroni correction assumes independent observations (Armstrong, 2014).

Post hoc analysis for possible demographic effects. In order to control for possible effects of demographic variables on the decision rules used by the subjects, we conducted a post hoc analysis on experimental data.

Figure S2 shows the distribution of frequencies of use of a medium of exchange according to gender, Figure S1 shows the distribution of frequencies of use of a medium of exchange according to age, and Table S4 summarizes statistical test results.

There is no significant effect of sex (Experiment I: $U = 83.5$, $p = 0.289$, $n = 28$; Experiment II: $U = 413.0$, $p = 0.335$, $n = 60$).

There is a significant correlation between age and a specific score in Experiment I ($r_{pearson} = -0.41$, $p = 0.031$, $n = 28$): it could indicate that a young age promotes the use of a medium of exchange. However, we observe no significant correlation between age and a specific score in the Experiment II ($r_{pearson} = -0.04$, $p = 0.790$, $n = 60$). As we don't have a homogeneous distribution of age, and there is a non-congruence of the results between Experiment I and Experiment II, we can not exclude the fact that the positive result for the Experiment I is due to sampling specificities.

Individual behaviors. Figures S3 and figures S4 depict, respectively for Experiment I and II, the individual indirect exchange frequency of each agent belonging to our population of interest (i.e. the type of agents that are able to use good 1 as a medium of exchange) for simulations (A.), experiments (B.), and post hoc simulations using as parametrization of the agents the best-fit parameter values of each subject (C.).

Post-hoc simulations with average best-fit parameter values (homogeneous pop.). Assuming the cognitive model as true, this could be due to the fact that the artificial agents from the same economy were having homogeneous cognitive features, while it exists certain heterogeneity among the human subjects that could make the coordination more difficult". To test the relevancy of this interpretation, after fitting the behavioral data with the model, we simulated an homogeneous population using as cognitive parameter values the average best value for each cognitive parameter after fitting the behavioral data (instead of simulating an heterogeneous population with the parameters of a single agent being the best-value parameters of a subject fit). These values are (rounded to the second decimal): $\alpha = 0.25$, $\beta = 1.05$, $\gamma = 0.26$.

The Fig. S7A and the Fig. S8A represents the results for post-hoc simulations based on Experiment I and II respectively. The Table S7 summarizes the statistic results.

The pattern remained unchanged: the non-uniform distribution of agent types promotes the use of a medium of exchange with 3 goods: agents of type (2, 3) use more the good 1 as a medium of exchange ($U = 9.5, p < 0.001^*, n = 28$), but not with four: agents of type (2, 3) do not use more the good 1 as a medium of exchange ($U = 68.5, p = 0.172, n = 30$), and neither do the agents of type (3, 4) ($U = 59.5, p = 0.078, n = 30$).

Post-hoc simulations with extended time. It is possible that more trials would have allowed subjects to coordinate over a unique medium of exchange at 4 goods. To evaluate the likelihood of such hypothesis, after fitting the behavioral data with the model, we simulated a population of (heterogeneous) agents with the parameters of every single agent being the best-fit parameter values of a single subject for a larger number of iteration ($n_t = 500$ instead of 50). The Fig. S7B and the Fig. S8B represents the results for post-hoc simulations based on Experiment I and II respectively. The Table S7 summarizes the results.

Here, the results changed as the non-uniform distribution of agent types promotes the use of a medium of exchange in both settings with a large number of trials: in the 3 goods settings, agents of type (2, 3) use more the good 1 as a medium of exchange ($U = 50.0, p = 0.029^*, n = 28$), and in the 4 goods settings, agents of type (2, 3) use more the good 1 as a medium of exchange, ($U = 24.0, p < 0.001^*, n = 30$), as well as the agents of type (3, 4) ($U = 54.5, p = 0.048^*, n = 30$). These results indicate that an extended time could have allowed the human subjects to modify slowly their behavior towards the use of a medium of exchange.

Post-hoc simulations with more goods. To further explore the influence of agents' distribution on the use of a medium of exchange, we ran simulations with 5 and 6 goods in circulation. The Fig. S9 presents the results of these simulations.

The phase diagram in Fig. S9A summarizes the results of 10, 400 simulations with 5 goods. The number of agent types (5, 1) and (1, 2) is set at 50, and the number of agents of type (2, 3) is set a 100. The number of agents of type (3, 4) and (4, 5) varies between 10 and 200 (corresponding respectively to the values on the x -axis and y -axis). As for the initial figure (Fig. ??), each simulation lasted 100 time-steps, the exploration parameter (ϵ) was varied between 0.10 and 0.15, the learning rate (α) was varied between 0.10 and 0.25, the discount factor (β) was varied between 0.80 and 1.20, and the initial values of success rate estimates for all types of exchange and for all agents were set to 1.

Similarly, the phase diagram on Fig. S9B summarizes the results of 10, 400 simulations with 6 goods. The number of agent of types (6, 1) and (1, 2) is set at 50, and the number of agents of type (2, 3) and (3, 4) is set a 100. The number of agents of type (4, 5) and (5, 6) varies between 10 and 200 (corresponding respectively to the values on the x -axis and y -axis).

In a five goods economy as in a six goods economy, the highest frequency of indirect exchanges with good 1 observed is when the value $x_{G-2,G-1}$ as well the value of $x_{G-1,G}$ is nearly twice

that of x_{12} and $x_{G,1}$. This pattern is consistent with what has been observed with 4 goods in circulation (see Fig. ??).

Model comparison In order to further assess the dependence of our results to the model we used (and especially its implementation), we ran simulations with several alternative models.

For all these models, we represented the results as a phase diagram such as Fig. ?. As for the initial figure (Fig. ?), each simulation lasted 100 time-steps, and the initial values of success rate estimates for all types of exchange and for all agents were set to 1. If applicable, the exploration parameter (ϵ) was varied between 0.10 and 0.15. If applicable, the learning rate (α) was varied between 0.10 and 0.25. If applicable, the discount factor (β) was varied between 0.80 and 1.20.

For economies with 3 goods in circulation, the number of agent type (3, 1) is set at 50, and the number of agents of types (1, 2) and (2, 3) varies between 10 and 200.

For economies with 4 goods in circulation, the number of agent of types (4, 1) and (1, 2) is set at 50, and the number of agents of type (2, 3) and (3, 4) varies between 10 and 200.

We constructed a simple model with a learning rule à la Roth Erev (Roth and Erev, 1995), with only one free parameter that defines the stochasticity of the model.

The differences with the initial model are (1) the learning rule, that is the equation ?? in the Material & Methods section; (2) the discounting rule, that is the equation ?? in the Material & Methods section.

Let e_{ij} be the success rate estimation associated to the exchange of type (i, j) . The learning rule used for this new model is:

$$e_{ij}^{t+1} = \frac{n_{ij}^t - 1}{n_{ij}^t} e_{ij}^t + \frac{1}{n_{ij}^t} \cdot (s - e_{ij}^t) \quad (\text{S1})$$

with n_{ij} , being the number of attempts of a ij exchange at time t .

Let $v(ij)$ be the value associated to the choice ij (i.e. exchange i against j) and Δ_{ij} the estimation by the agent of the time that will be spent before consumption if he chooses ij . The new discounting rule is the following:

$$v(ij) = 1/\Delta_{ij} \quad (\text{S2})$$

In a three-goods setting, although a little more ‘noisy’, the results are following the same pattern as we previously observed (see Fig. S13). In a four-goods setting, we don’t observe any “hotspot” as it was the case with our model. One possible explanation for such a discrepancy in the results in the four goods settings is that this running averaging rule implicitly implies a dynamic learning rate (the impact of a new observation on the subjective values is higher in the beginning compared to the end) that is not optimal for coordination at 4 goods.

Also, we implemented another model with one softmax rule and a non-parametric discounting rule (which consists in dividing the utility by the expected delay to obtain it).

The differences with the initial model are (1) the discounting rule, that is the equation ?? in the Material & Methods section, and (2) the decision rule, that is the equation ?? in the Material & Methods section.

Let e_{ij} be the success rate estimation associated to the exchange of type (i, j) . The learning rule used for this new model is:

$$e_{ij}^{t+1} = e_{ij}^t + \alpha \cdot (s - e_{ij}^t) \quad (\text{S3})$$

with $\alpha \in [0, 1]$, a free parameter and s , a binary variable such as $s = 1$ if the agent succeeded in his exchange, 0 otherwise.

Let $v(ij)$ be the value associated to the choice ij (i.e. exchange i against j) and Δ_{ij} the estimation by the agent of the time that will be spent before consumption if he chooses ij :

$$v(ij) = 1/\Delta_{ij} \quad (\text{S4})$$

Let $p(ij)$ be the probability to choose to exchange i against j . The new decision rule is:

$$p(ij) = \frac{\exp(v(ij)_t/\gamma)}{\sum_k \exp(v(ik)_t/\gamma)} \quad (\text{S5})$$

with $\gamma \in [0, 1]$, a free parameter that describes the exploration-exploitation ratio.

In a three-goods setting as in the four goods setting, the results are following the same pattern as we previously observed (see Fig. S12). This result shows that the way to implement that has finally a limited impact on the results obtained.

In order to see the impact of the discount rule that we used (Osborne, 2016) on the results, we implemented two other models, one using an exponential discount rule and the other a hyperbolic discount rule, as defined in Green and al. Green, Fristoe, and Myerson, 1994.

The only difference with the initial model is the discounting rule, which is the equation ?? in the Material & Methods section.

Let $v(ij)$ be the value associated to the choice ij (i.e. exchange i against j) and Δ_{ij} the estimation by the agent of the time that will be spent before consumption if he chooses ij . We used as “exponential discount rule” the following:

$$v(ij) = \exp(-\beta\Delta_{ij}) \quad (\text{S6})$$

with $\beta > 0$, the discount factor parameter.

We used as “hyperbolic discounting rule” the following:

$$v(ij) = 1/(1 + \beta\Delta_{ij}) \quad (\text{S7})$$

with $\beta > 0$, the discount factor parameter.

In a three-goods setting as in the four goods setting, the results are following the same pattern as we previously observed. This result shows that the way to implement the value discounting rule has finally a limited impact on the results obtained (see Fig. S10 for the exponential discounting rule and Fig. S11 for the hyperbolic discounting rule).

Also, recent studies report the existence of asymmetry in value-update (Palminteri et al., 2017; Lefebvre et al., 2017): subjects exhibit a tendency to favor confirmatory information rather than disconfirmatory information. To investigate if our subjects present this type of learning bias, and to what extent it affects coordination in our environment, we consider an asymmetric model with two learning rates. The only difference with the initial model is then located in the learning rule, which is the equation ?? in the Material & Methods section.

At time step t , when an agent attempts to exchange i against j , it updates the success rate estimation associated to the exchange of type (i, j) , noted e_{ij} according to:

$$e_{ij}^{t+1} = e_{ij}^t + \begin{cases} \alpha^+ \cdot (s - e_{ij}^t), & \text{if } (s - e_{ij}^t) > 0 \\ \alpha^- \cdot (s - e_{ij}^t), & \text{if } (s - e_{ij}^t) < 0 \end{cases} \quad (\text{S8})$$

with both α^+ and α^- being free parameters belonging to the interval $[0, 1]$. If $\alpha^+ > \alpha^-$ we refer to an *optimistic bias*, if $\alpha^- > \alpha^+$ we refer to a *pessimistic bias*. The former implies that agents take more into account exchanges with better-than-expected outcomes. The latter implies that agents give priority to exchanges with worse-than-expected outcomes.

In a three good setting as in the four goods setting, the results are following the same pattern as we previously observed (see Fig. S14).

It seemed also interesting to see if these models could constitute a good explanation of our experimental results. For all models except the one without parameters, the BIC scores are not significantly different than the one of our initial model (see Fig. S19 and Table S10), showing that all models have the same explanatory power.

We also report in Fig. S18 the distribution of the best-fit parameter values for all the models.

Interestingly, the reinforcement model with an asymmetric learning rule shows a significant difference between the fitted learning rates associated to a negative prediction errors and the fitted learning rates associated to positive prediction errors ($U = 8321.0$, $p < 0.001$, $n = 166$), that is interpreted in the literature as an optimistic bias (Palminteri et al., 2017; Lefebvre et al., 2017). This is coherent with the literature, as our study implements an experience-based framework (agents have to learn the reward contingencies).

In order to see if the presence of such bias could offer an interesting perspective on our data, we ran several post-hoc simulations.

The phase diagram in Fig. S15 summarizes the results of 900 simulations with $G = 3$ (3 goods in circulation), and 900 simulations with $G = 4$ (4 goods in circulation). For all these simulations, we used a non-uniform distribution promoting the use of money: the number of agent types (G , 1) and (1, 2) has been set to 50, and the number of agents of any other type has been set a 100. As for the initial figure (Fig. ??), each simulation lasted 100 time-steps, the exploration parameter (ϵ) was varied between 0.10 and 0.15, the learning rate (α) was varied between 0.10 and 0.25, the discount factor (β) was varied between 0.80 and 1.20, and the initial values of success rate estimates for all types of exchange and for all agents were set to 1.

This phase diagram doesn't present the influence of agent type distribution on the use of the good 1 as a medium of exchange, but the influence of the two learning rates (α^- and α^+). These results

seem to indicate that a low α^- (that is the tendency to ignore negative prediction errors) could prevent the use of a medium of exchange, for a three-goods setting economy as for a four-goods setting economy.

Also, we ran two additional simulations in the same conditions that the experiments I and II with agents presenting either an optimistic bias or a pessimistic bias. Regarding the 'optimistic bias' condition, we used the average best-fit parameters of the subjects (given that the subjects were presenting an optimistic bias), that is: $\beta = 1.083$, $\gamma = 0.271$, $\alpha^- = 0.288$ and $\alpha^+ = 0.467$. For the 'pessimistic bias condition', we swapped the values for the two α and let the other parameters unchanged ($\beta = 1.083$, $\gamma = 0.271$, $\alpha^- = 0.467$ and $\alpha^+ = 0.288$).

The results are depicted in the figures S16 and S17. The tables S8 summarizes the statistical results.

With three goods, we observe a significant difference between a uniform and non-uniform distribution of agents' type when agents present an optimistic bias ($U = 21.0$, $p < 0.001$, $n = 28$), while it's not the case with agents presenting a pessimistic bias ($U = 62.5$, $p = 0.098$, $n = 28$). With four goods in circulation, the difference between non-uniform and uniform is significant for both biases, but not for all agents' types. Agents of type (2, 3) present significant differences with an optimistic bias ($U = 54.5$, $p = 0.045$, $n = 30$) as well as agents of type (3, 4) ($U = 50.5$, $p = 0.031$, $n = 30$). Agents of type (2, 3) present significant differences with a pessimistic bias ($U = 42$, $p = 0.011$, $n = 30$) while agents of type (3, 4) do not ($U = 95$, $p = 0.843$, $n = 30$).

Unfortunately, the results are puzzling and do not confirm or infirm the hypothesis that an optimistic bias could preclude or promote the use of a medium of exchange. Similarly, the results with regards to a pessimistic bias are not informative enough to draw any conclusion.

References

- Armstrong, Richard A (2014). "When to use the Bonferroni correction." In: *Ophthalmic and Physiological Optics* 34.5, pp. 502–508.
- Roth, Alvin E and Ido Erev (1995). "Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term." In: *Games and economic behavior* 8.1, pp. 164–212.
- Osborne, Michael (2016). "Exponential versus hyperbolic discounting: a theoretical analysis." In: *Available at SSRN* 2518162.
- Green, Leonard, Nathanael Fristoe, and Joel Myerson (1994). "Temporal discounting and preference reversals in choice between delayed outcomes." In: *Psychonomic Bulletin & Review* 1.3, pp. 383–389.
- Palminteri, Stefano et al. (2017). "Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing." In: *PLoS computational biology* 13.8, e1005684.
- Lefebvre, Germain et al. (2017). "Behavioural and neural characterization of optimistic reinforcement learning." In: *Nature Human Behaviour* 1.4, p. 0067.

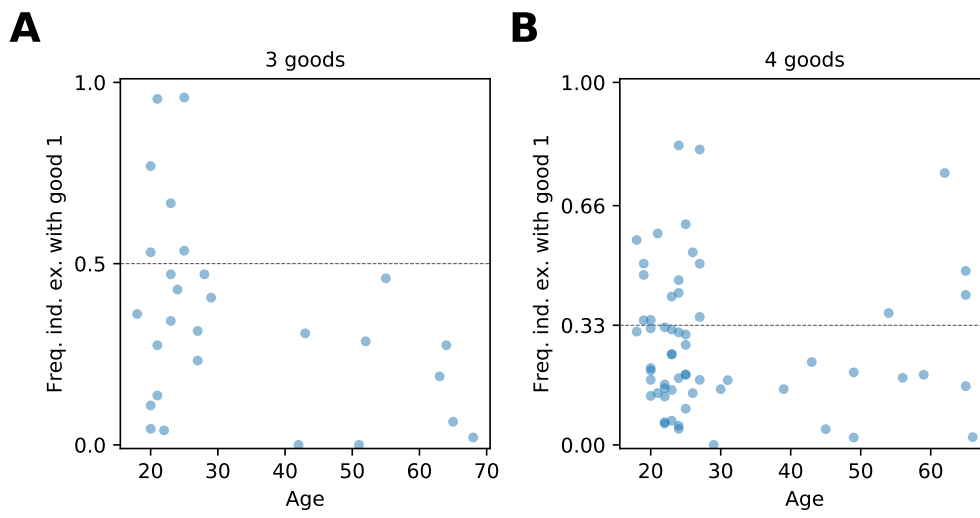


Figure S1. Use of a medium of exchange: age effect. **A.** Experiment I (3 goods in circulation). **B.** Experiment II (4 goods in circulation).

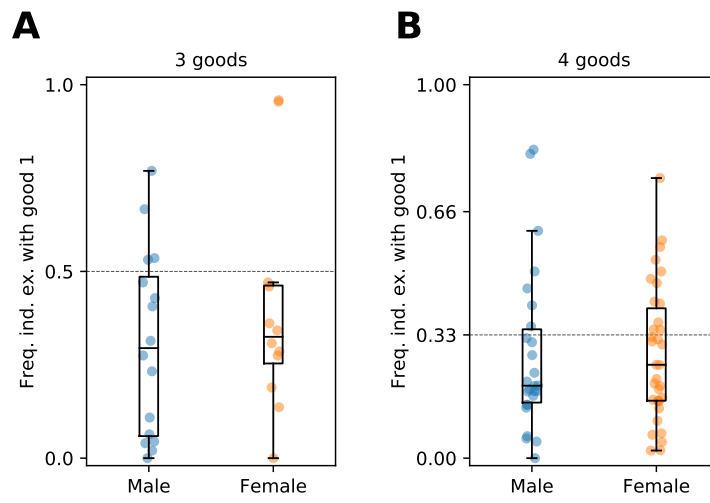


Figure S2. Use of a medium of exchange: gender effect. **A.** Experiment I (3 goods in circulation). **B.** Experiment II (4 goods in circulation).

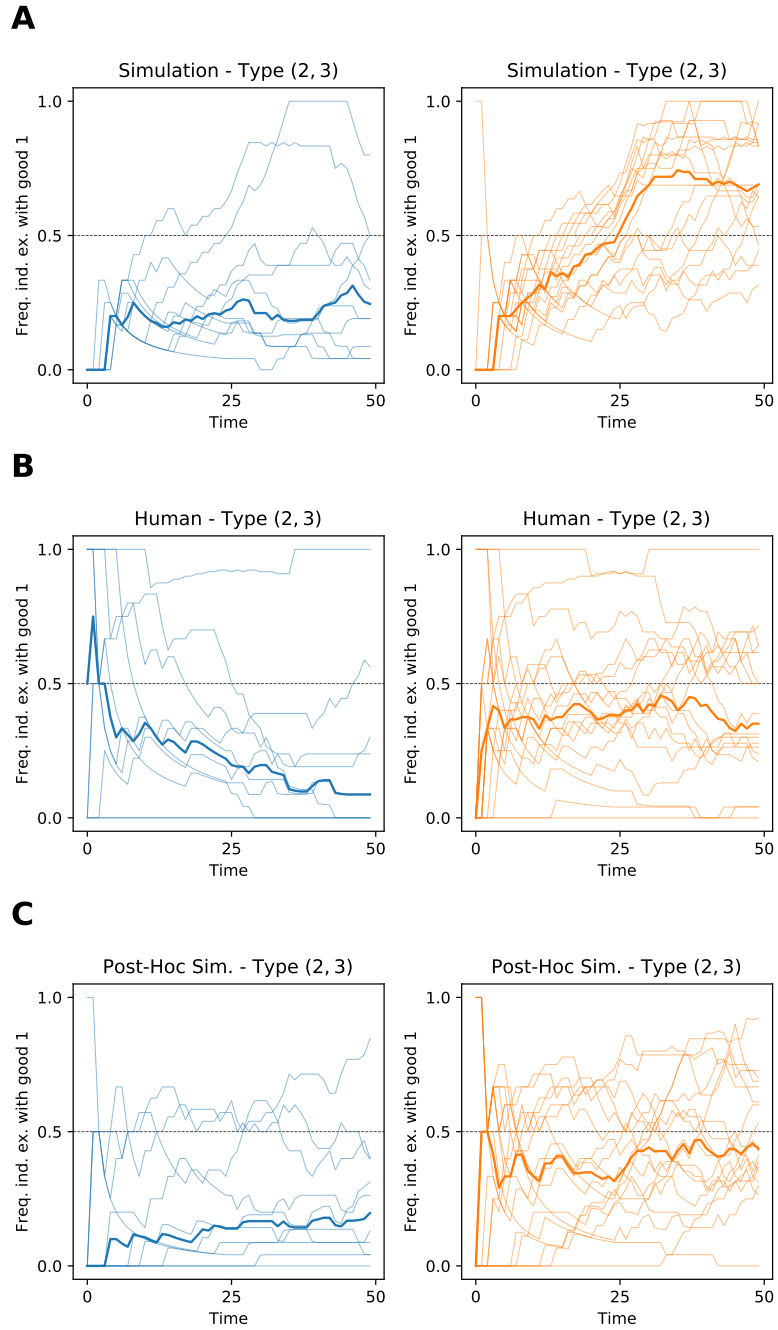


Figure S3. Experiment I: Individual behavior regarding the use of a medium of exchange. **A.** Simulation. **B.** Experimental results. **C.** Post-hoc simulation.

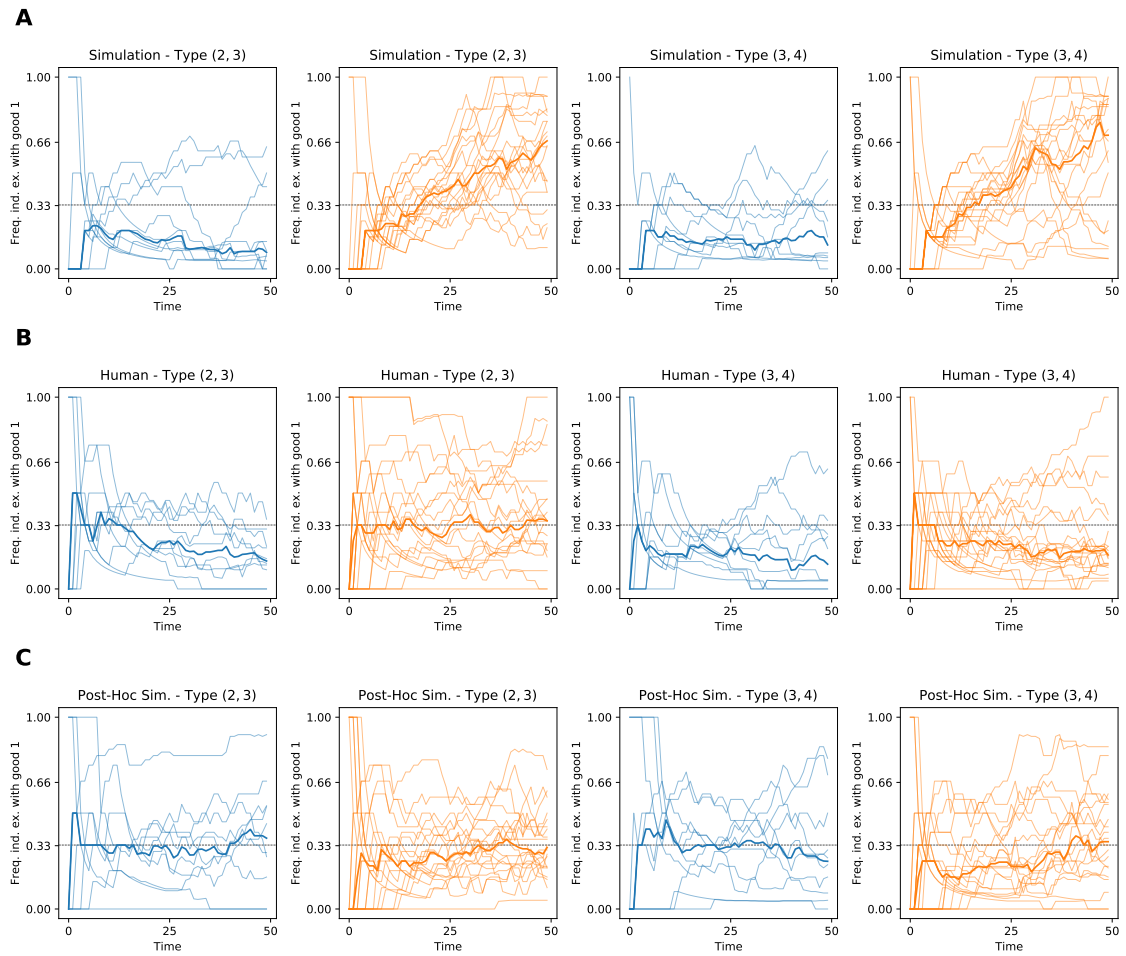


Figure S4. Experiment II: Individual behavior regarding the use of a medium of exchange. **A.** Simulation. **B.** Experimental results. **C.** Post-hoc simulation.

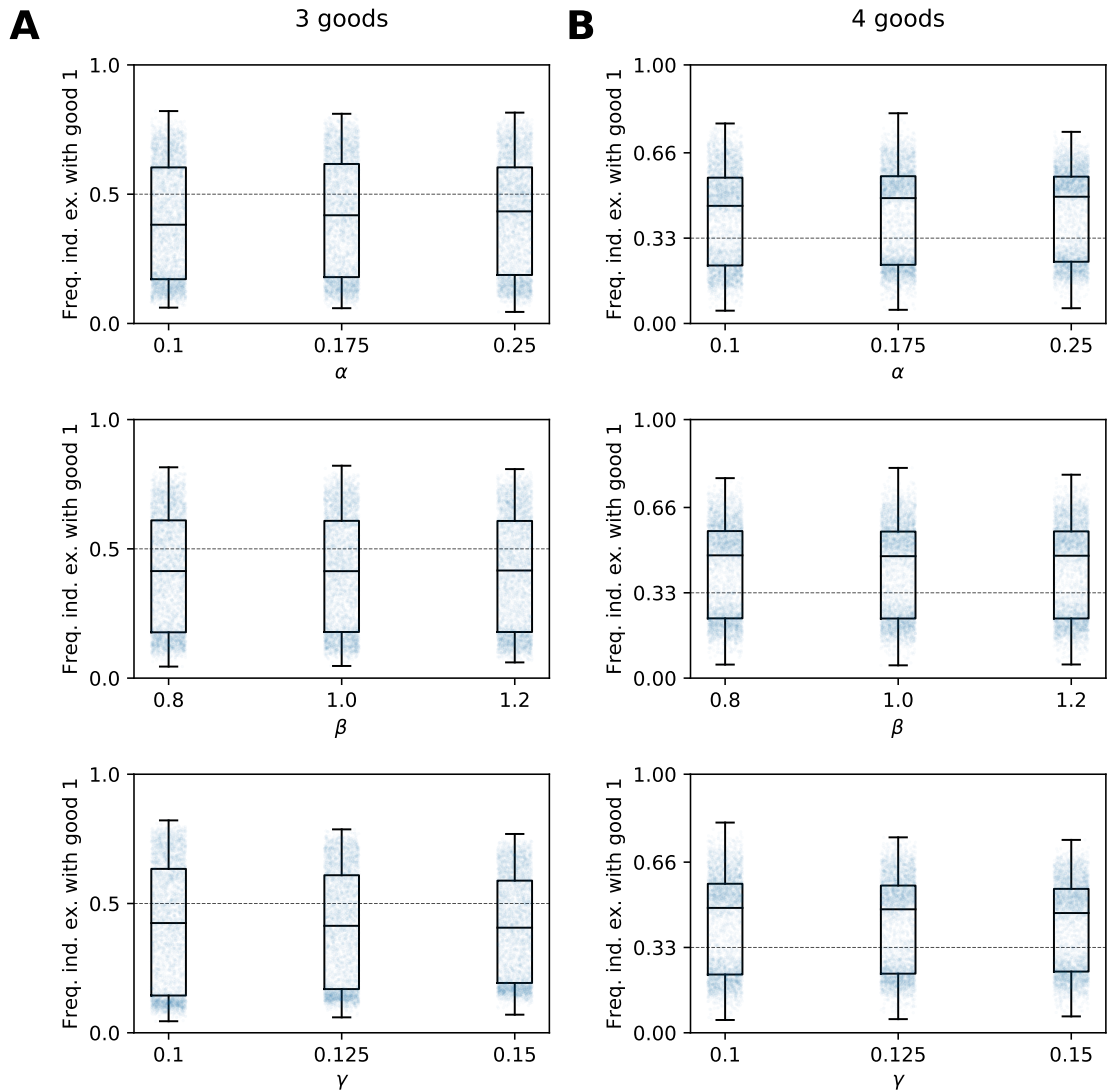


Figure S5. Sensitivity analysis of free parameters (α, β, γ). **A.** Economies with 3 goods in circulation. **B.** Economies with 4 goods in circulation.

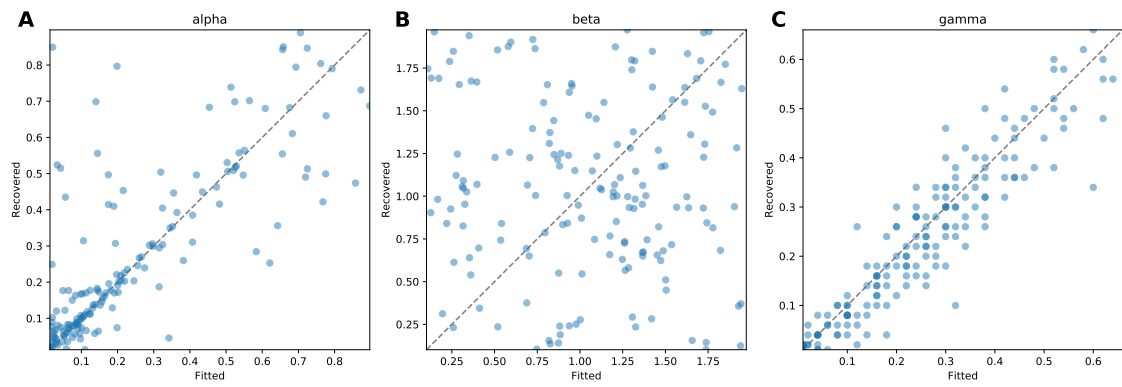


Figure S6. Parameter recovery. **A.** Learning rate (α). **B.** Discount factor (β). **C.** Exploration-exploitation ratio (γ).

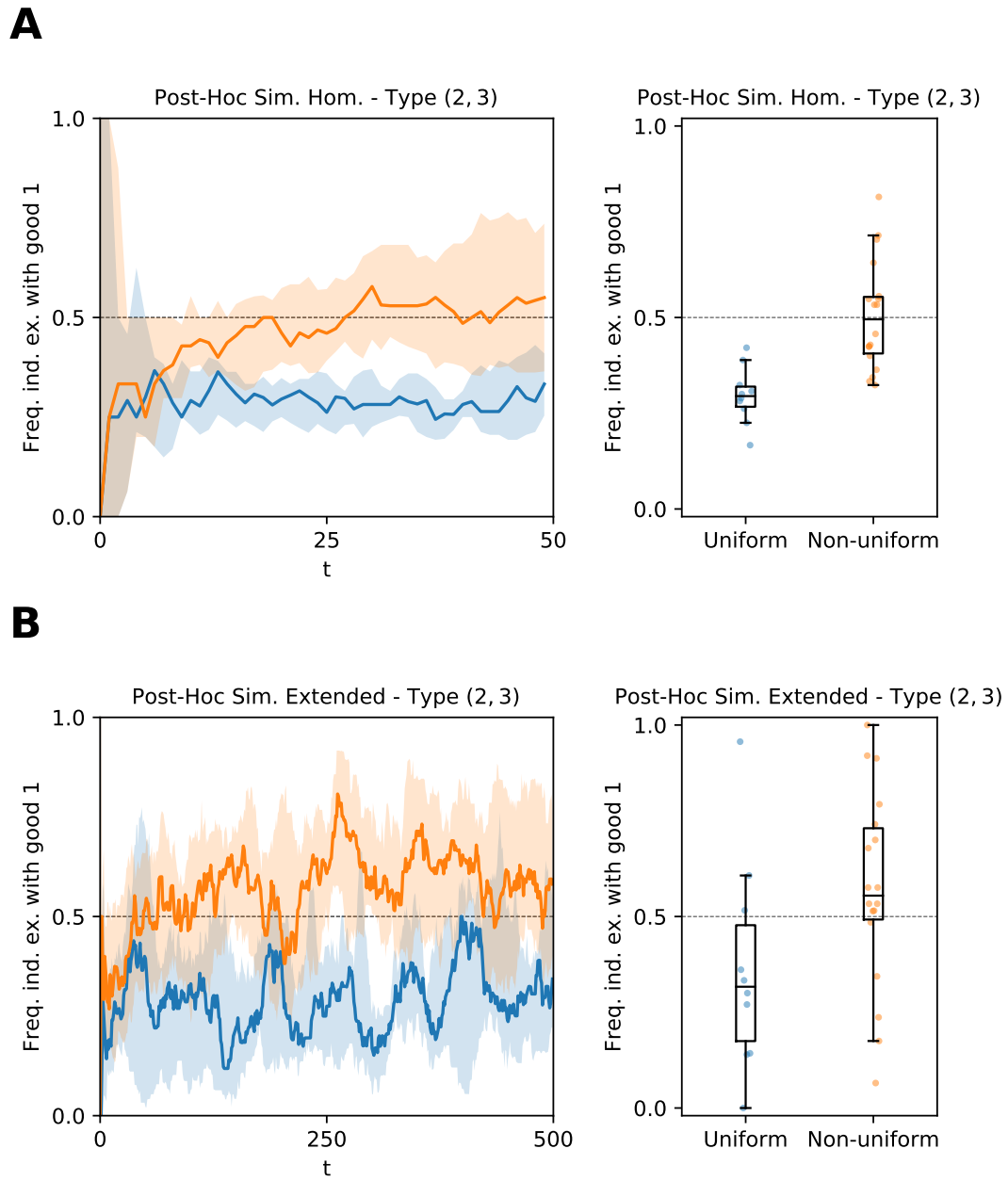


Figure S7. Experience I: Post-hoc simulations testing 'heterogeneity' and 'lack of time' hypothesis. **A.** Using the *average* best-value parameters (instead of a 'one agent-one subject' matching). **B.** With more time-steps (1000 instead of 50).

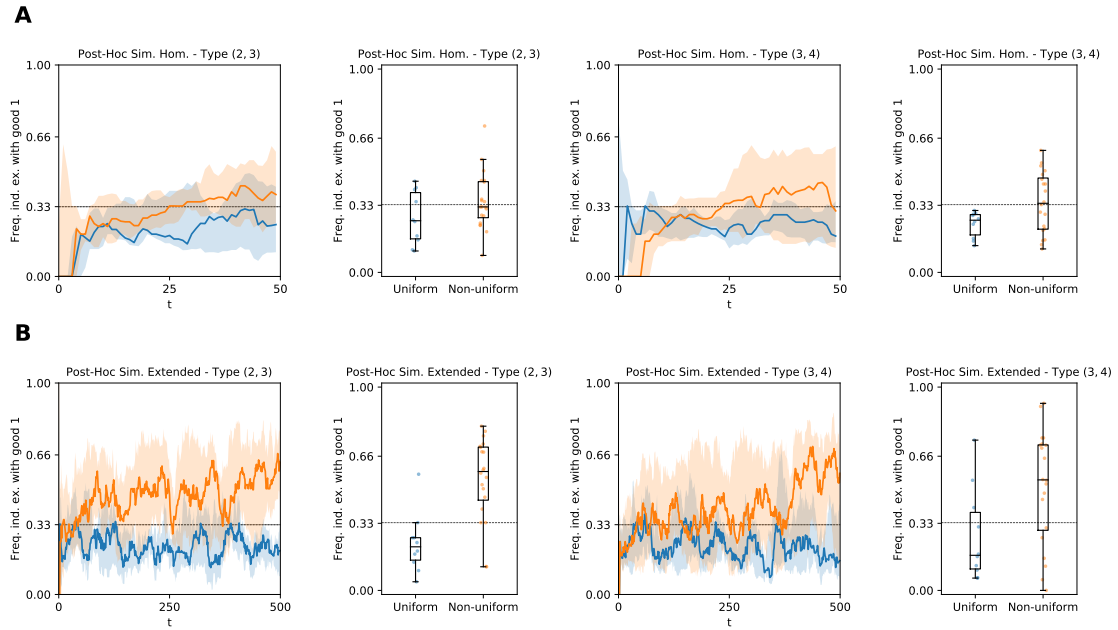


Figure S8. Experience II: Post-hoc simulations testing 'heterogeneity' and 'lack of time' hypothesis. **A.** Using the *average* best-value parameters (instead of a 'one agent-one subject' matching). **B.** With more time-steps (1000 instead of 50).

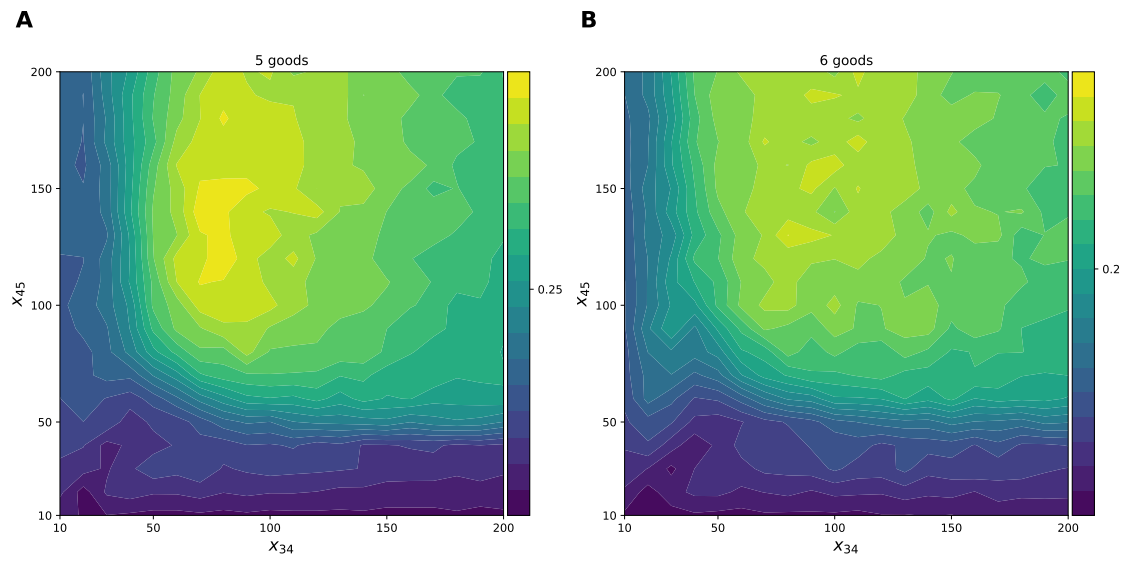


Figure S9. Influence of agent distribution on the use of a medium of exchange: post-hoc simulations with 5 and 6 goods in circulation. **A.** Economies with 5 goods in circulation. **B.** Economies with 6 goods in circulation.

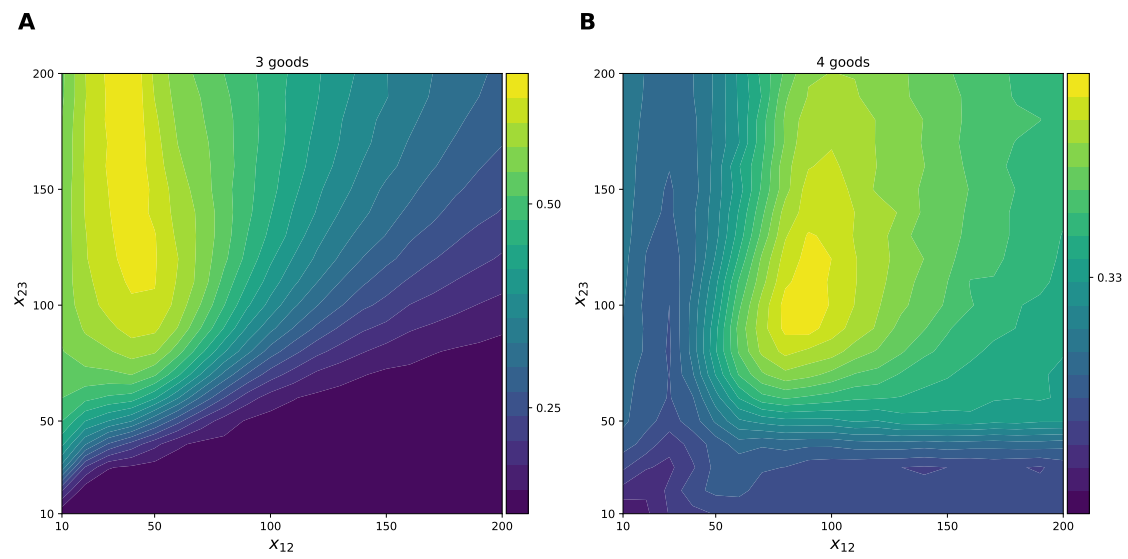


Figure S10. Influence of agent distribution on the use of a medium of exchange: post-hoc simulations with an exponential discounting rule. The hotter the color, the higher the indirect exchange frequency involving good 1 as a medium of exchange. **A.** Economies with 3 goods in circulation. **B.** Economies with 4 goods in circulation.

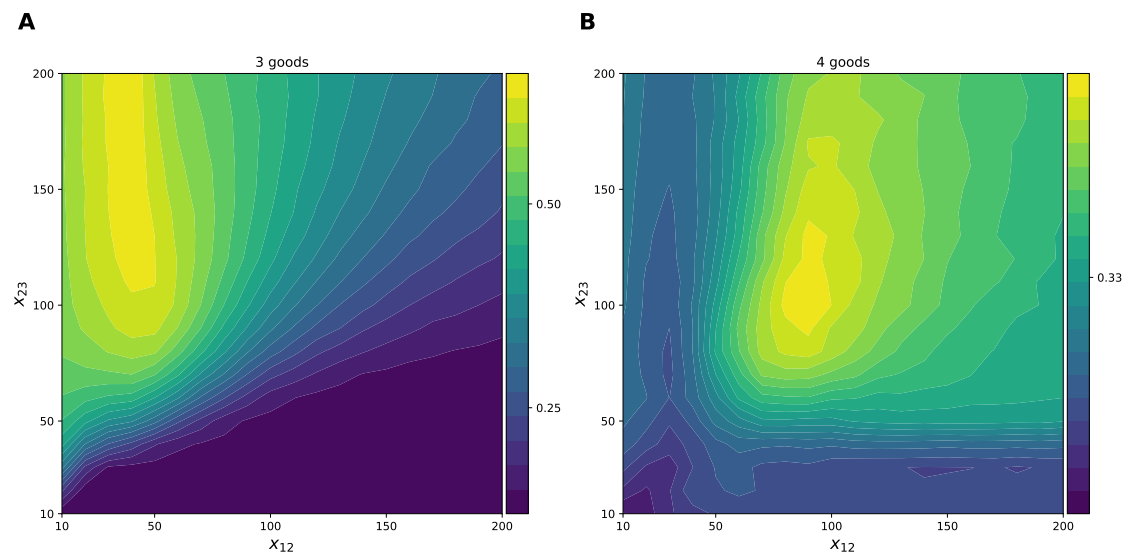


Figure S11. Influence of agent distribution on the use of a medium of exchange: post-hoc simulations with an hyperbolic discounting rule. The hotter the color, the higher the indirect exchange frequency involving good 1 as a medium of exchange. **A.** Economies with 3 goods in circulation. **B.** Economies with 4 goods in circulation.

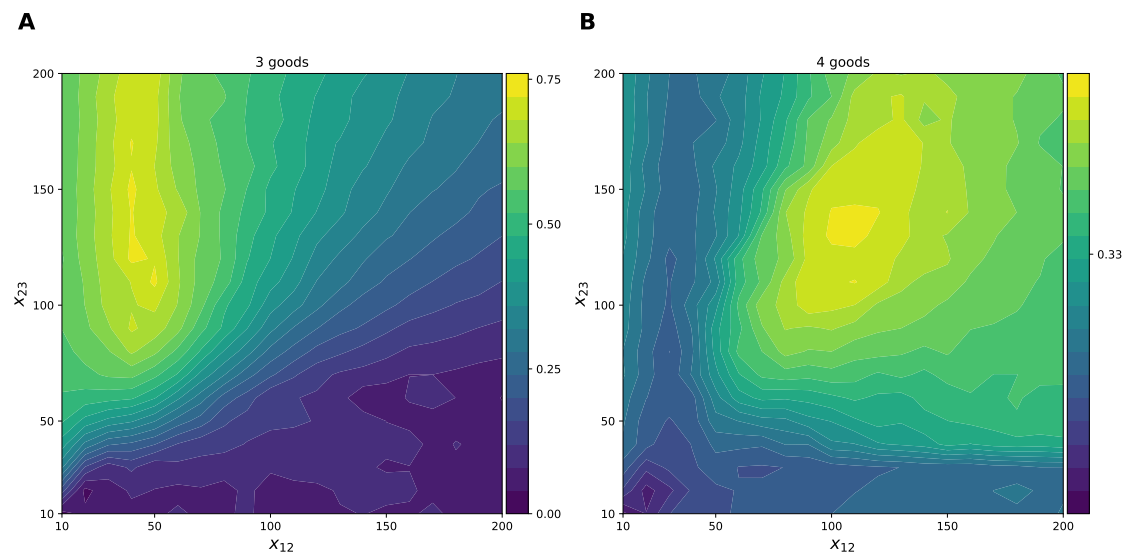


Figure S12. Influence of agent distribution on the use of a medium of exchange: post-hoc simulations with a softmax decision rule. The hotter the color, the higher the indirect exchange frequency involving good 1 as a medium of exchange. **A.** Economies with 3 goods in circulation. **B.** Economies with 4 goods in circulation.

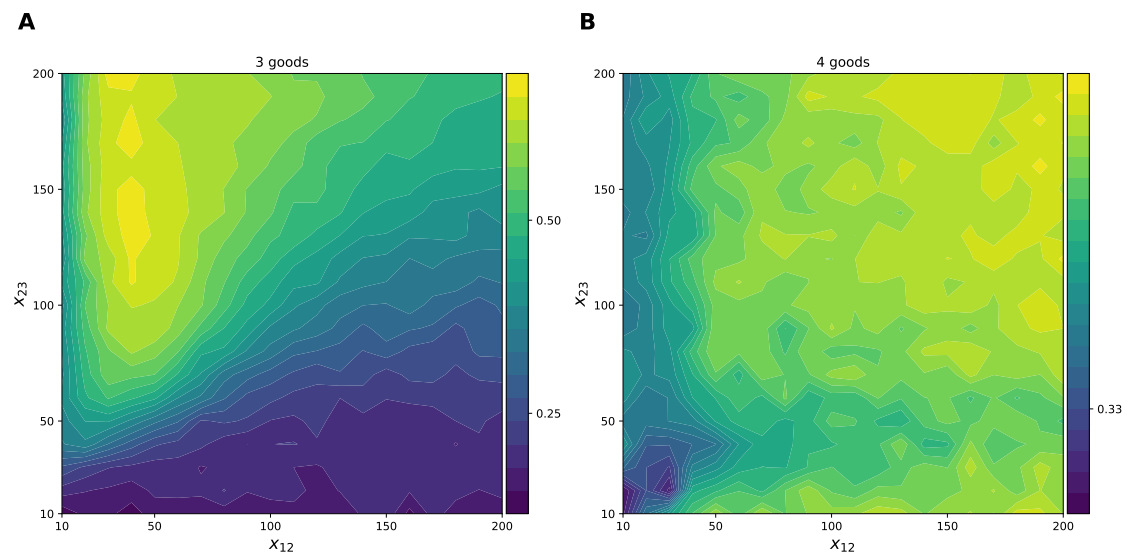


Figure S13. Influence of agent distribution on the use of a medium of exchange: post-hoc simulations with a non-parametric learning rule. The hotter the color, the higher the indirect exchange frequency involving good 1 as a medium of exchange. **A.** Economies with 3 goods in circulation. **B.** Economies with 4 goods in circulation.

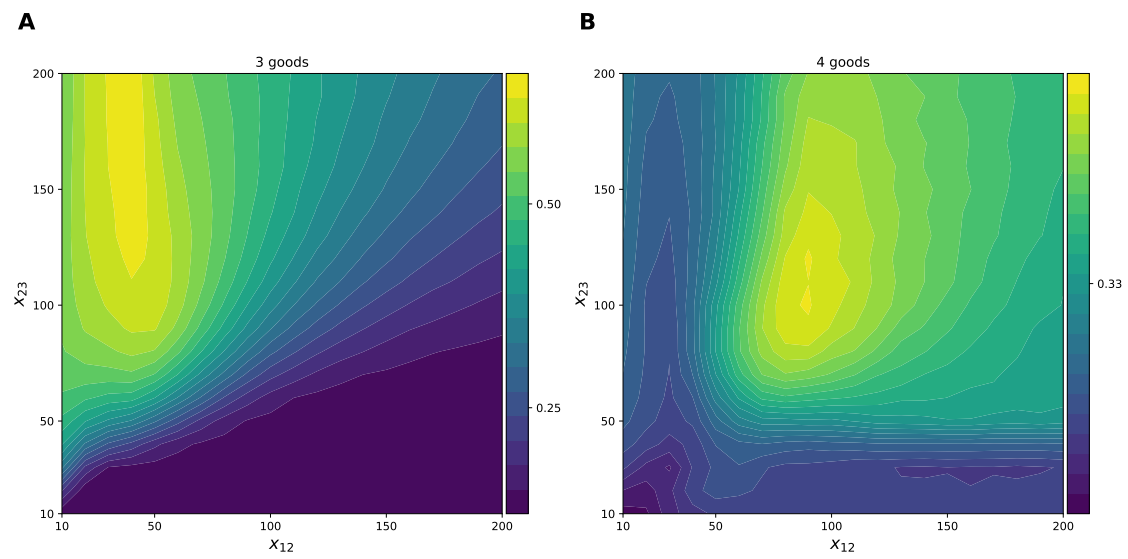


Figure S14. Influence of agent distribution on the use of a medium of exchange: post-hoc simulations with an asymmetric learning rule. The hotter the color, the higher the indirect exchange frequency involving good 1 as a medium of exchange. **A.** Economies with 3 goods in circulation. **B.** Economies with 4 goods in circulation.

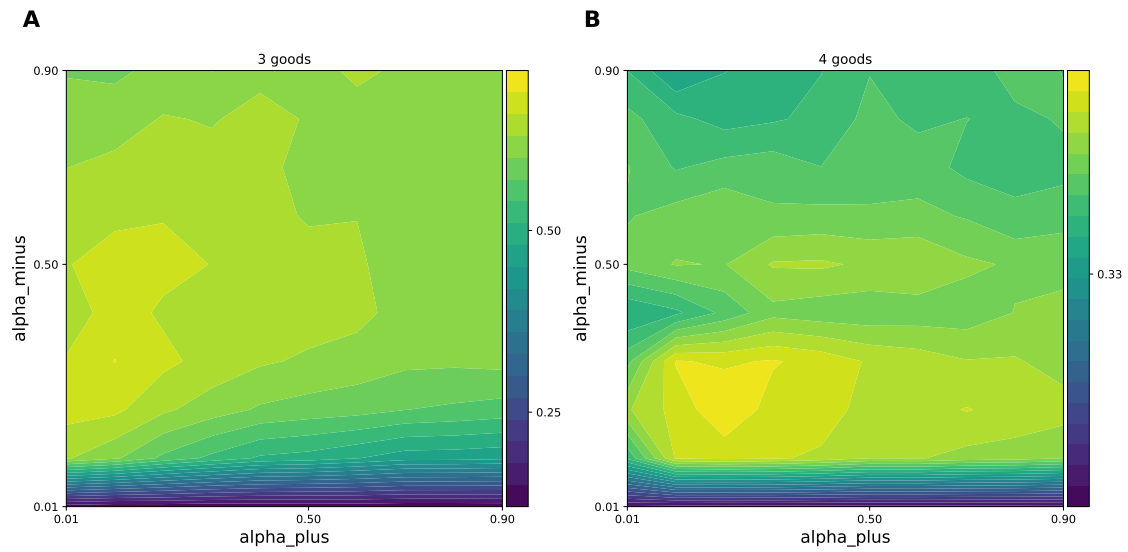
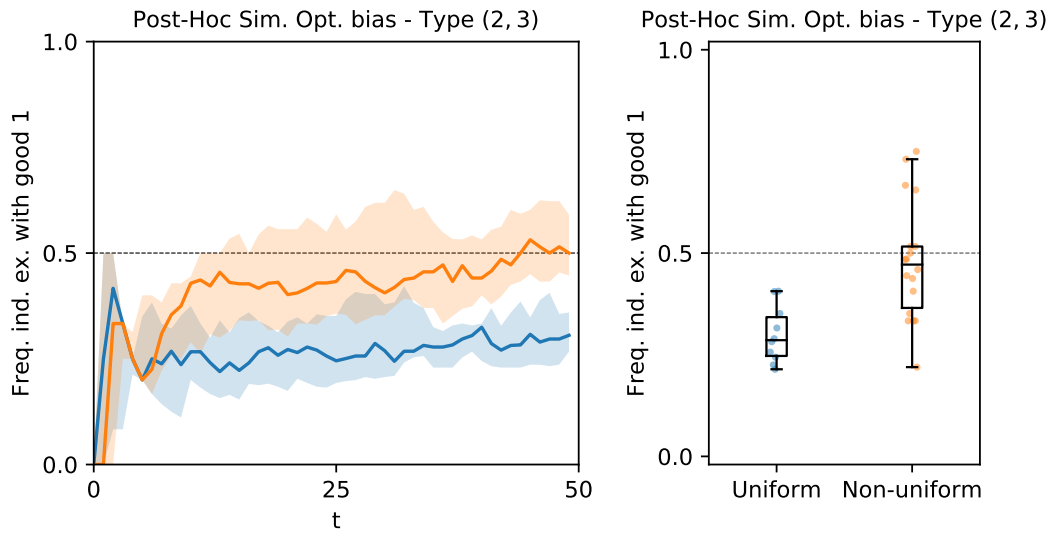


Figure S15. Influence of the asymmetry of the learning rates on the use of a medium of exchange: post-hoc simulations with an asymmetric learning rule. The hotter the color, the higher the indirect exchange frequency involving good 1 as a medium of exchange. **A.** Economies with 3 goods in circulation. **B.** Economies with 4 goods in circulation.

A



B

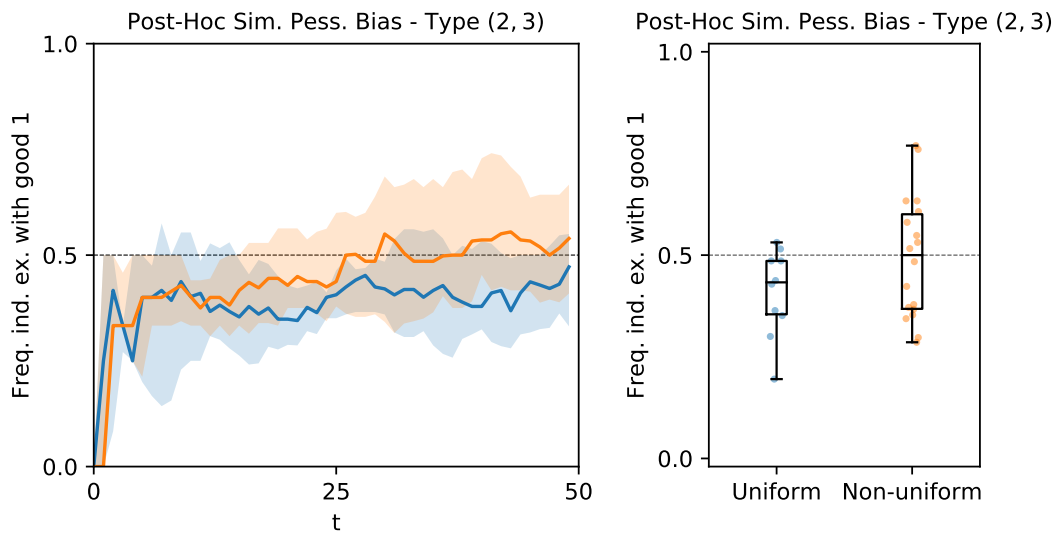


Figure S16. Experience I: Post-hoc simulations with an asymmetric learning rule. **A.** Optimistic bias. **B.** Pessimist bias.

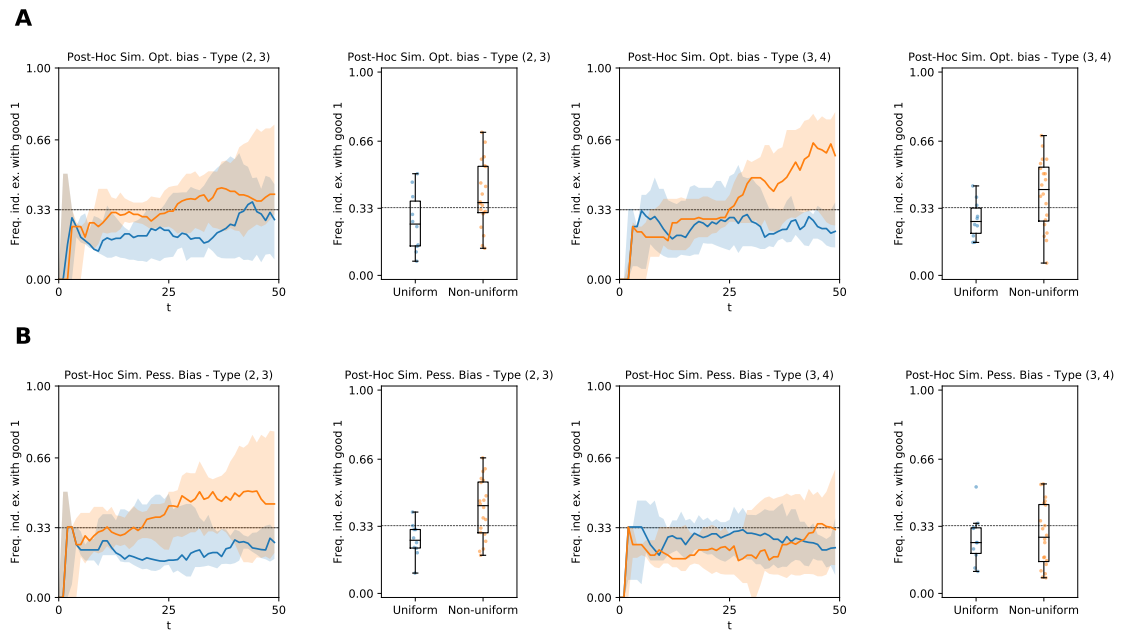


Figure S17. Experience II: Post-hoc simulations with an asymmetric learning rule. **A.** Optimistic bias. **B.** Pessimist bias.

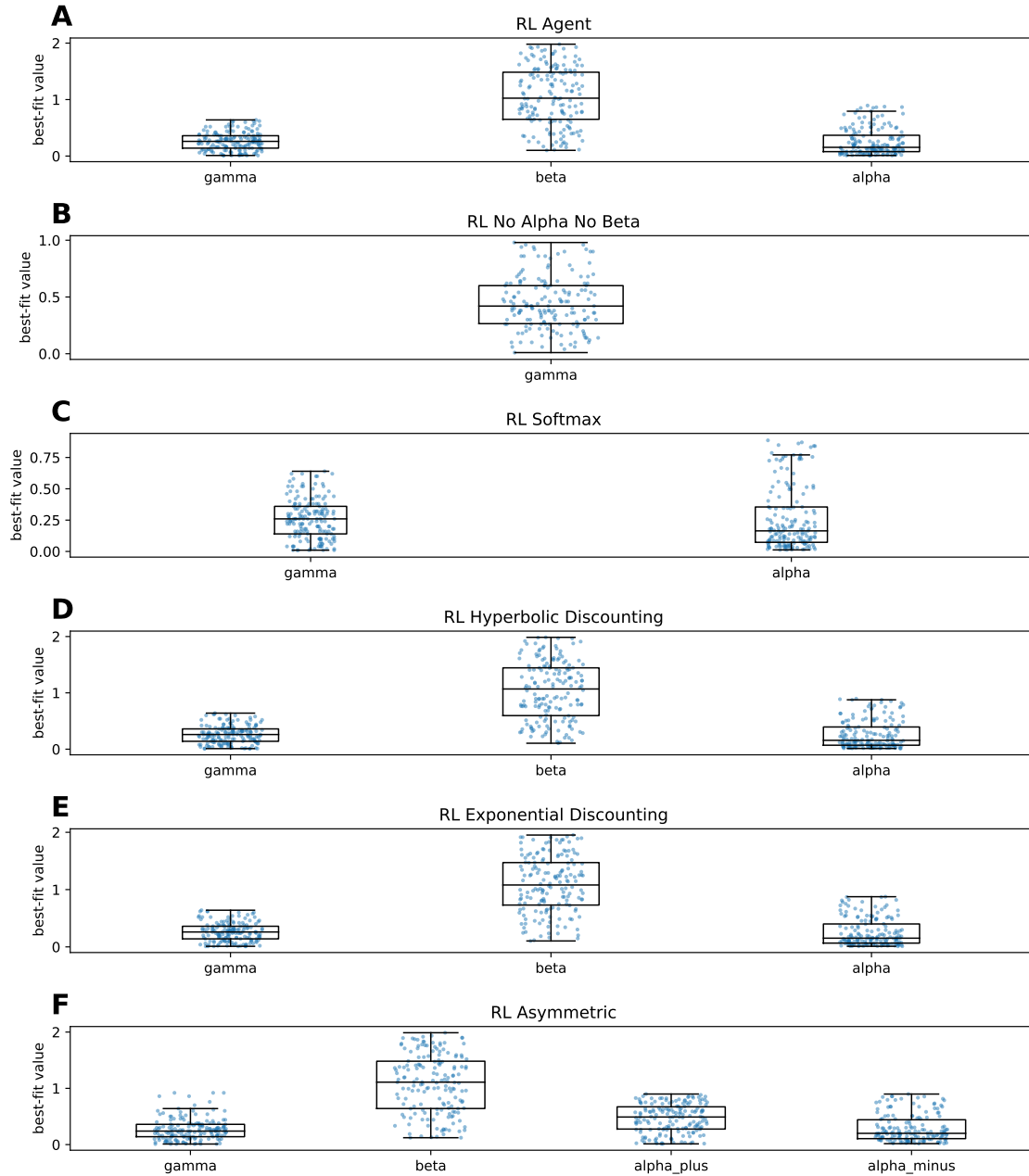


Figure S18. Distribution of the best-fit parameter values for the different models. **A.** Reinforcement model described in the Methods & Material section. **B.** Reinforcement model with a non-parametric learning rule. **C.** Reinforcement model with a softmax decision rule. **D.** Reinforcement model with an hyperbolic discounting rule. **E.** Reinforcement model with an exponential discounting rule. **F.** Reinforcement model with an asymmetric learning rule.

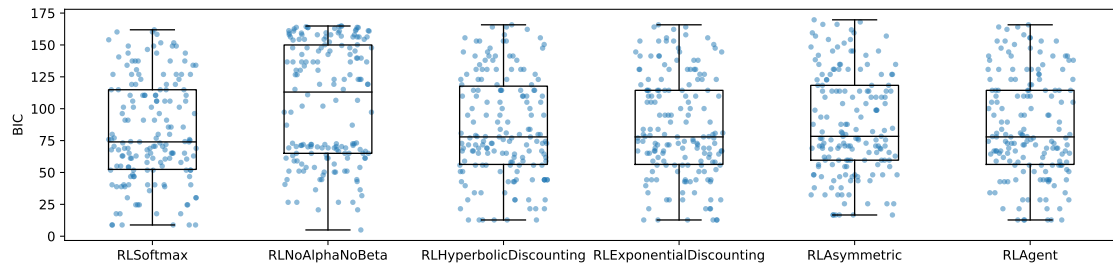


Figure S19. BIC (Bayesian Information Criterion) scores. Lower values indicate better description of the data by the model.

Condition	G	x_{31}	x_{12}	x_{23}	n_{eco}	n_t
Uniform	3	10	10	10	1	50
Non-uniform	3	9	9	18	1	50

Table S1. Experiment I: Parameters.

Condition	G	x_{41}	x_{12}	x_{23}	x_{34}	n_{eco}	n_t
Uniform	4	10	10	10	10	1	50
Non-uniform	4	10	10	20	20	1	50

Table S2. Experiment II: Parameters.

Exp.	Factor	Pop. of interest	Measure	Test	Value	Raw <i>p</i> -value	Corr. <i>p</i> -value	<i>n</i> _{obs}
I	Agent type dist.	Art. - Type (2, 3)	Ind. good 1	<i>U</i>	21.0	< 0.001	< 0.001*	28
I	Agent type dist.	Hum. - Type (2, 3)	Ind. good 1	<i>U</i>	50.5	0.031	0.031*	28
I	Agent type dist.	Post-hoc art. - Type (2, 3)	Ind. good 1	<i>U</i>	48.0	0.023	0.023*	28
II	Agent type dist.	Art. - Type (2, 3)	Ind. good 1	<i>U</i>	21.0	< 0.001	< 0.001*	30
II	Agent type dist.	Art. - Type (3, 4)	Ind. good 1	<i>U</i>	28.0	< 0.001	0.002*	30
II	Agent type dist.	Hum. - Type (2, 3)	Ind. good 1	<i>U</i>	56.0	0.028	0.056	30
II	Agent type dist.	Hum. - Type (3, 4)	Ind. good 1	<i>U</i>	77.5	0.167	0.333	30
II	Agent type dist.	Post-hoc art. - Type (2, 3)	Ind. good 1	<i>U</i>	99.0	0.491	0.982	30
II	Agent type dist.	Post-hoc art. - Type (3, 4)	Ind. good 1	<i>U</i>	78.5	0.178	0.355	30

Table S3. Experiences I & II: Statistics summary. ‘*’ indicates a significant result to the threshold of 5%.

Exp.	Factor	Measure	Test	Value	<i>p</i> -value	<i>n</i> _{obs}
I	Gender	Ind. good 1	<i>U</i>	103.0	0.762	28
I	Age	Ind. good 1	<i>r</i> _{pearson}	0.16	0.206	66
II	Gender	Ind. good 1	<i>U</i>	402.5	0.560	60
II	Age	Ind. good 1	<i>r</i> _{pearson}	0.03	0.777	100

Table S4. Gender and age effect: Statistics summary. No result is significant to the threshold of 5%.

G	Factor	Measure	Test	Value	Raw p -value	Corr. p -value	n_{obs}
3 goods	α -value	Ind. good 1	H	76.17	< 0.001	< 0.001*	10800
3 goods	β -value	Ind. good 1	H	0.01	0.997	1.000	10800
3 goods	γ -value	Ind. good 1	H	19.71	< 0.001	< 0.001*	10800
4 goods	α -value	Ind. good 1	H	45.50	< 0.001	< 0.001*	10800
4 goods	β -value	Ind. good 1	H	0.02	0.991	1.000	10800
4 goods	γ -value	Ind. good 1	H	0.69	0.709	1.000	10800

Table S5. Sensitivity analysis of free parameters: Statistics summary. “*” indicates a significant result to the threshold of 5%.

Parameter	Test	Value	<i>p</i> -value	<i>n_{obs}</i>
α	$r_{pearson}$	0.88	< 0.001*	166
β	$r_{pearson}$	0.01	0.920	166
γ	$r_{pearson}$	0.91	< 0.001*	166

Table S6. Parameter recovery. ‘**’ indicates a significant result to the threshold of 5%.

Manip.	Exp.	Factor	Pop. of interest	Measure	Test	Value	Raw p -value	Corr. p -value	n_{obs}
Hom. pop.	I	Agent type dist.	Type (2, 3)	Ind. good 1	U	9.5	< 0.001	< 0.001*	28
Extended time	I	Agent type dist.	Type (2, 3)	Ind. good 1	U	50.0	0.029	0.029*	28
Hom. pop.	II	Agent type dist.	Type (2, 3)	Ind. good 1	U	68.5	0.086	0.172	30
Hom. pop.	II	Agent type dist.	Type (3, 4)	Ind. good 1	U	59.5	0.039	0.078	30
Extended time	II	Agent type dist.	Type (2, 3)	Ind. good 1	U	24.0	< 0.001	< 0.001*	30
Extended time	II	Agent type dist.	Type (3, 4)	Ind. good 1	U	54.5	0.024	0.048*	30

Table S7. Post hoc simulations with *averaged* best-fit parameter values (homogeneous pop.) or extended time: Statistics summary. ‘*’ indicates a significant result to the threshold of 5%.

Manip.	Exp.	Factor	Pop. of interest	Measure	Test	Value	Raw <i>p</i> -value	Corr. <i>p</i> -value	<i>n</i> _{obs}
Opt. bias	I	Agent type dist.	Type (2, 3)	Ind. good 1	<i>U</i>	21.0	< 0.001	< 0.001*	28
Pess. bias	I	Agent type dist.	Type (2, 3)	Ind. good 1	<i>U</i>	62.5	0.098	0.098	28
Opt. bias	II	Agent type dist.	Type (2, 3)	Ind. good 1	<i>U</i>	54.0	0.023	0.045*	30
Opt. bias	II	Agent type dist.	Type (3, 4)	Ind. good 1	<i>U</i>	50.5	0.016	0.031*	30
Pess. bias	II	Agent type dist.	Type (2, 3)	Ind. good 1	<i>U</i>	42.0	0.006	0.011*	30
Pess. bias	II	Agent type dist.	Type (3, 4)	Ind. good 1	<i>U</i>	95.0	0.421	0.843	30

Table S8. Post hoc simulations with a optimistic or pessimistic bias: Statistics summary. “*” indicates a significant result to the threshold of 5%.

Model	BIC	n_{param}	n_{obs}
RL Agent	85.77 ± 39.85 STD	3	166
RL No Alpha No Beta	104.29 ± 45.86 STD	1	166
RL Softmax	81.91 ± 39.89 STD	2	166
RL Hyperbolic Discounting	85.84 ± 39.84 STD	3	166
RL Exponential Discounting	85.79 ± 39.87 STD	3	166
RL Asymmetric	87.35 ± 39.46 STD	4	166

Table S9. Average BIC scores.

Comparison	Test	Value	Raw p -value	Corrected p -value	n_{obs}
RL Agent - RL No Alpha No Beta	U	10620.0	< 0.001	0.002*	332
RL Agent - RL Softmax	U	12981.0	0.181	1.000	332
RL Agent - RL Hyperbolic Discounting	U	13757.0	0.491	1.000	332
RL Agent - RL Exponential Discounting	U	13778.0	0.500	1.000	332
RL Agent - RL Asymmetric	U	13529.0	0.388	1.000	332
RL No Alpha No Beta - RL Softmax	U	9899.0	< 0.001	< 0.001*	332
RL No Alpha No Beta - RL Hyperbolic Discounting	U	10646.0	< 0.001	0.003*	332
RL No Alpha No Beta - RL Exponential Discounting	U	10623.0	< 0.001	0.002*	332
RL No Alpha No Beta - RL Asymmetric	U	10879.0	< 0.001	0.007*	332
RL Softmax - RL Hyperbolic Discounting	U	12966.0	0.177	1.000	332
RL Softmax - RL Exponential Discounting	U	12978.0	0.180	1.000	332
RL Softmax - RL Asymmetric	U	12746.0	0.119	1.000	332
RL Hyperbolic Discounting - RL Exponential Discounting	U	13761.0	0.492	1.000	332
RL Hyperbolic Discounting - RL Asymmetric	U	13546.0	0.396	1.000	332
RL Exponential Discounting - RL Asymmetric	U	13531.0	0.389	1.000	332

Table S10. BIC scores comparison: Statistics summary. ‘*’ indicates a significant result to the threshold of 5%.