ONLINE APPENDIX for "Part-Time Bayesians: Incentives and Behavioral Heterogeneity in Belief Updating"

Carlos Alós-Ferrer*[1] and Michele Garagnani[1]

[1]University of Zurich

March 21, 2022

A Unsupervised Approach to FMM

Our main FMM analysis relied on the supervised approach since our experiment was designed to disentangle pre-specified behavioral rules (Bayes' rule, model-free Reinforcement, Inertia, and Non-Updating), and an unsupervised approach lacks the ability to directly connect the obtained results to those behavioral rules. However, it is natural to ask what the results of the unsupervised FMM would be, and hence we present them here.

We followed an unsupervised FMM approach at the aggregate level and compared models with different numbers of components (2, 3, and 4). Specifically, we computed mixtures of logistic regressions (since the dependent variable is binary, i.e. betting on white or black for the second decision) with dummy regressors x_1, x_2, x_3 describing the first decision (ω): 4 balls (vs. 6), betting on white (vs. on black), and winning (vs. losing). Each FMM then classifies decisions in one of the pre-specified number of clusters and delivers the coefficients β_1, β_2, and β_3 for the dummies plus the constant β_0. We then computed the predicted probabilities for betting on white for each of the eight possible decision situations (dummy combinations; recall Figure 1(B)) as

$$p(b = 1|\omega) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3)}.$$

Table A.1 below presents the components as identified in this way. Rounding probabilities to the fourth decimal position already results in almost all the predicted values being zero or one, which we represent with the corresponding ball color as in Figure 1(B) to ease the exposition. The remaining values are rounded to the second decimal. Table A.2 presents the results of the estimation, i.e. component weights, standard deviations, log-likelihood, and the Akaike and Bayesian information criteria.

The best model according to the information criteria is the one with four components. The rules are almost precisely identified. The first component, which has the largest weight, prescribes to almost

*Corresponding author. Zurich Center for Neuroeconomics (ZNE), Department of Economics, University of Zurich. Blümlisalpstrasse 10, CH-8006 Zurich, Switzerland. E-mail: carlos.alos-ferrer@econ.uzh.ch

Table A.1: Results of the unsupervised FMM approach with 4, 3, and 2 components.

Input			4 Components				3 Components			2 Comp.	
	First Bet	Stimulus	C1	C2	C3	C4	C1	C2	C3	C1	C2
4 Balls	●	● Win	0.06	○	●	○	○	●	●	○	●
	●	○ Lose	0.33	○	●	●	○	●	●	○	●
	○	○ Win	●	○	○	○	○	●	○	○	●
	○	● Lose	●	○	●	○	○	●	○	○	●
6 Balls	●	● Win	0.07	○	●	○	○	●	●	○	●
	●	○ Lose	0.39	○	●	●	○	●	●	○	●
	○	○ Win	●	○	○	○	○	●	○	○	●
	○	● Lose	●	○	●	○	○	●	○	○	●

Table A.2: Estimation summary, unsupervised finite mixture model with different number of components. Std.Dev. in parenthesis.

Nr. of rules	C1	C2	C3	C4	Log Lik.	AIC	BIC
4	48.72	42.33	8.95	0.01	-10996.86	22005.73	22051.84
	(16.84)	(16.43)	(1.06)	(0.01)			
3	50.22	49.78	0.01		-11145.65	22301.31	22339.74
	(15.18)	(15.08)	(0.01)				
2	50.22	49.78			-11145.65	22293.31	22300.99
	(3.91)	(3.96)					

always bet black. Specifically, it dictates to bet black if the first bet was on white (independently of the outcome and task), and, if the first bet was on black, bet on black with a probability of around 94% if that bet won, and with a probability of around 65% if it lost. The second component, which captures most of the remaining weight, prescribes to blindly bet on white no matter what the decision is. The third component dictates to bet black unless the first bet was also on white and resulted in a win. The fourth, which has a negligible weight, prescribes to (unintuitively) bet white unless the first bet was on black and lost.

The results with three or two components are almost identical, with two components with almost equal weights corresponding to always betting on white and always betting on black, respectively. For three components, the third corresponds exactly to inertia, but has a negligible weight.

Essentially, the approach fits the data by postulating a dichotomy between "always bet on black" and "always bet of white." These results merely suggest that the unsupervised approach is not appropriate for our data. Essentially, the reason is that the appropriate level of analysis for the unsupervised approach is the aggregate level, because an application at the individual level would result in different components for each participant, which could in principle not be matched across participants (participant 1's first component is not participant 17's first component, and might actually not be any of her components). But the analysis at the aggregate level loses track of the appropriate identification unit, which is the individual. That is, if an individual uses a certain behavioral rule, this creates a perfect correlation across the choices of this individual for the eight different decision situations. This correlation, however, is lost in the unsupervised approach. Put differently, if data arose from 50% perfectly-Bayesian participants and 50% perfectly-anti-Bayesian ones (who always do the opposite of what Bayes rule prescribes), the unsupervised approach would effectively treat the dataset as purely random behavior. While one could start a discussion on different clustering methods avoiding these problems, in our case the supervised approach at the individual level completely sidesteps them.

B Estimation at the Aggregate Level

Our main FMM analysis in Section 4.1 was performed for each individual separately. In this section we show the results of the aggregate level of analysis, where we consider all choices in the dataset as deriving from identical agents. Hence, we estimate the proportions of choices classified as following each behavioral rule. This level of analysis shows which is the most common rule of behavior.

The estimation at the aggregate level treats all observations as equivalent. That is, each observation (ω, b) is assumed to be a realization of a distribution with

$$p(b|\omega) = \sum_{k=1}^{4} \eta_k \cdot p_k(b|\omega, \varepsilon_k)$$

and hence the likelihood of the dataset $D = \{(\omega_1, b_1), \ldots, (\omega_M, b_M)\}$ (where $M = 268 \times 60$) is

$$p(D) = \prod_{m=1}^{M} \sum_{k=1}^{4} \eta_k \cdot p_k(b_m|\omega_m, \varepsilon_k).$$

The estimation delivers the proportions η_k of choices made according to each rule and the overall error rates ε_k associated with the rules.

Rule	Est.weight	Est.error rate
Bayes	48.55	47.57
	(47.63)	(20.11)
Reinf.	27.36	47.77
	(42.68)	(23.59)
Inertia	18.50	51.12
	(33.16)	(13.39)
Non-upd.	5.60	48.81
	(13.36)	(23.51)
	N=16080	

Table B.1: Summary of aggregate estimation, Std.Dev. in parenethesis.

Table B.1 summarizes the distribution of the estimated parameters for each rule at the aggregate level. The number of observations for this level of analysis is the total number of choices collected from all subjects during the entire experiment. With 268 subjects and 60 repetitions of the paradigm for each subject, the total number of observations used in the aggregate estimation is N=16,080. The “Est.Weight” and “Est.Error rate” columns report the estimated weights and error rates for each behavioral rule, respectively.

As Figure B.1 shows the normative prescription is the most commonly followed behavioral rule, with 48.55% of choices classified as following Bayes’ rule. There is considerable support for reinforcement (27.36%) and inertia (18.50%), while just 5.60% of choices are classified as following non-updating. However, the classification is extremely inaccurate: the estimated error rates relative to each rule of behavior are very high, all close to 50%.

Such high error rates, which effectively turn the behavioral rules into random behavior, suggest that the analysis at the aggregate level is not appropriate for our data. This is not surprising. Our main analysis suggests strong heterogeneity across individuals. The analysis at the aggregate level,

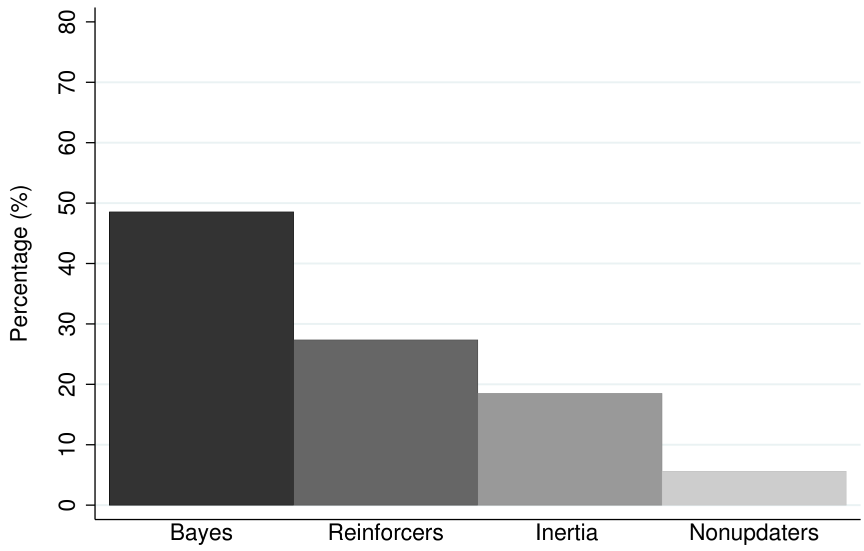


Figure B.1: Proportions of choices classified into each behavioral rule.

however, assumes that all choices come from the same or similar subjects. Given the high estimated error rates, we can refute this assumption.

C Finite Mixture Model Comparisons (Error-Free)

We present here a model comparison for an aggregate-level FMM estimation, where different number of rules are contemplated and where we consider mistake-free, deterministic rules. This reduces the number of free parameters and allows for a simple, conceptual robustness check of the FMM estimation in the main text.

Table C.1 reports the results of the model comparisons. The first column indicates the number of rules assumed in the estimation. Columns two to five contain the estimated weights for each rule. The next column reports log likelihood of the model (smaller in absolute value indicating a better fit), in order to compare model fit without penalising for the number of parameters estimated. The last two columns report AIC and BIC information criteria (lower score is better), which penalize for additional rules. Standard errors are reported in parenthesis and are computed using the Delta-method. The result of the estimation with four (mistake-free) rules, is qualitatively similar to our main results, with the difference that the weights on both inertia and non-updating are negligible. In general, Bayes' rule and (model-free) reinforcement are the rules with highest support across all different model specifications.

According to the log likelihood criterion, the best model is the one with four components. However, when we penalize for the number of parameters, AIC and BIC suggest that a model with only Bayes' rule and (model-free) reinforcement is the best fit to the data. Both AIC and BIC agree that the worst model is the one comprising only inertia and non-updating, but these criteria rank the other models differently, in agreement with the fact that BIC penalizes free parameters more strongly.

Qualitatively, these results provide a conceptual robustness check of our main results, since Bayes' rule and (model-free) reinforcement remain the most commonly-used rules of behavior. However, the analysis ignores heterogeneity across individuals and, as shown in Section B, might be inaccurate at the individual level.

Table C.1: Estimation summary, finite mixture model with different number of components. Std.Dev. in parenthesis.

Nr. of rules	Bayes	RL	Inertia	NonUp	Log Lik.	AIC	BIC
4	52.84	46.79	0.37	0.01	-11144.565	22297.130	22327.871
	(13.13)	(13.02)	(0.13)	(0.01)			
3	69.84	29.74	0.42		-11145.538	22297.076	22320.132
	(18.84)	(14.65)	(0.08)				
3	70.05	29.95		0.01	-11145.538	22297.076	22320.132
	(17.60)	(10.17)		(0.01)			
3	86.86		12.72	0.41	-11145.623	22297.246	22320.302
	(28.43)		(2.84)	(0.07)			
3		43.18	31.04	25.78	-11145.727	22297.454	22320.510
		(6.91)	(4.97)	(11.87)			
2	50.45	49.55			-11145.685	22295.370	22310.741
	(3.18)	(3.26)					
2	98.23		1.77		-11146.675	22297.350	22312.721
	(29.17)		(0.58)				
2	70.74			29.26	-11146.737	22297.474	22312.845
	(28.23)			(3.82)			
2		100.00	0.00		-11146.741	22297.482	22312.853
		(29.88)	(0.01)				
2		99.59		0.41	-11150.033	22304.066	22319.437
		(22.96)		(0.26)			
2			50.01	49.99	-11297.601	22599.202	22614.573
			(24.69)	(34.68)			

D Three-Component Finite Mixture Model

Because of the relatively high estimated error rates obtained by the finite mixture model for Inertia and Non-updating (Table 1), it is reasonable to ask whether subjects classified into these rules are actually simply randomizing. In order to investigate this hypothesis we estimated a different finite mixture model with only three behavioral rules. Specifically, we assumed that subjects followed either Bayes' rule, reinforcement, or a third rule (Random) which prescribes to uniformly randomise between the alternatives at each trial. The estimation procedure was the same as described in the main text (i.e., rules include error terms and the analysis is at the individual level), with the exception that for the third rule, the probability of an "error" was constrained to zero.

Table D.1 reports the results of the three-component finite mixture model in the same format as Table 1. We observe that most people are classified as reinforcers (49.24%), followed by Bayesians (34.85%), and randomizers (15.91%). This is in contrast with the four-components classification where Bayes' rule had the largest support. Conditioning on the most-used rule (column Cond.Mean) shows that the three-components classification performs worse than the four-components one. The conditional means are around 55% to 60%, indicating that the three-type classification is more ambiguous than the four-types one, where all conditional means were above 84%.

Table D.1: Estimation summary, finite mixture model with 3 components. Std.Dev. in parenthesis.

Behavioral rule	Est.Weight	Est.Error rate	Classified as	%	Cond.Mean	Cond.Error rate
Bayes	37.63	23.90	92	34.85	55.67	21.18
	(18.22)	(15.05)			(12.21)	(12.12)
Reinforcement	44.59	18.75	130	49.24	60.92	16.19
	(20.23)	(14.04)			(12.75)	(12.43)
Random	17.78	-	42	15.91	55.03	-
	(19.87)	-			(14.01)	-
N =264						

E Parameter Recovery for the HMM Estimation

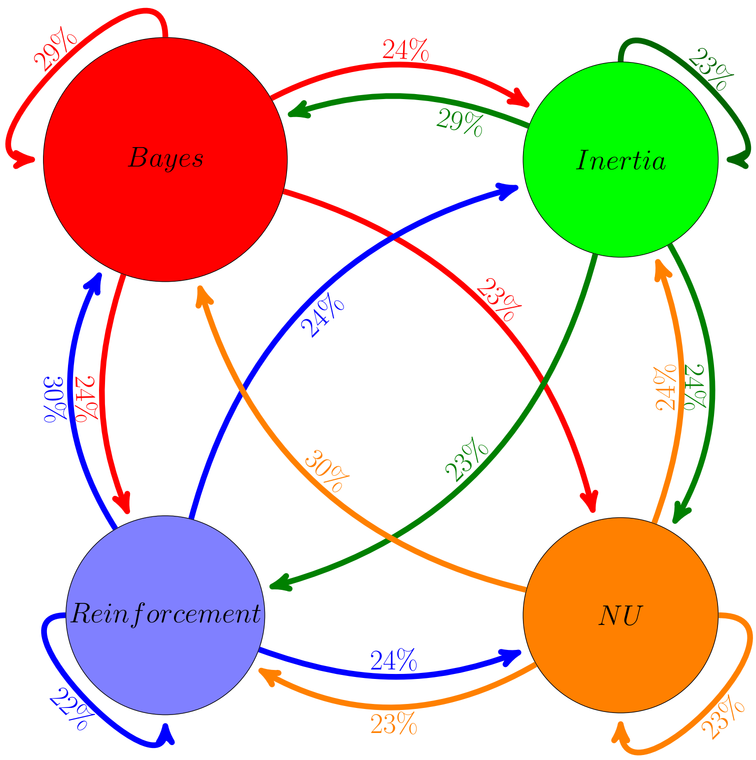
In this section we report the results of the HMM analysis for an artificial dataset where we know the data generating process. In particular, we mimic our dataset simulating 60 trials for $N = 268$ fictitious participants. Each participant follows each of the four rules with equal probability and switches among them also with equal probability. This represents a parameter recovery exercise, as a basic check of the estimation procedure (Palminteri et al., 2016).

Table E.1 and Figure E.1 are the analogues of Table 3 and Figure 6 in the main text, respectively. They display the average results of the estimation for this type of analysis, pooling across incentive treatments. The "From" column indicates the behavioral rule which was most likely followed in the previous trial. The columns under the "To" label indicate the behavioral rule to where the Markov process is most likely to go. The last column of the table ("Error") indicates the estimated probability of making an error at each state. The last row of the table ("Invariant") reports the probabilities in the corresponding invariant distribution. We observe that the recovery exercise closely mimics the data generating process, with only a minor overestimation of Bayesian behavior in the sample.

Table E.1: Temporal dynamics, individual averages with Std.Dev. in parentheses.

N=268	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	29.09%	23.84%	24.04%	23.05%	21.41%
	(13.26)	(11.96)	(11.98)	(11.21)	(14.00)
RL	29.61%	21.77%	24.22%	24.40%	25.41%
	(12.03)	(10.82)	(11.78)	(11.62)	(17.20)
Inertia	28.89%	23.56%	22.68%	24.88%	34.16%
	(12.31)	(10.86)	(11.08)	(11.85)	(13.38)
Non-up	30.20%	23.45%	23.68%	22.68%	28.95%
	(11.91)	(11.93)	(11.15)	(10.43)	(21.32)
Invariant:	29.47%	23.08%	23.68%	23.77%	

Figure E.1: Graphical depiction of the results of the HMM estimation for an artificial dataset where each participant follows each of the four rules with equal probability and switches among them with equal probability. Circle sizes are proportional to weights in the invariant distribution. Arrow thickness is proportional to transition probabilities among states. Probabilities themselves are rounded for graphical illustration.



F Heterogeneity in Transition Probability Matrices

For a dataset where each participant makes multiple decisions, a finite mixture model can be interpreted as making the assumption that, for each choice, a type (in our case a behavioral rule) is randomly selected to make the decision, following the corresponding probabilities (in our case η_k^j). That is, the dynamic interpretation is that the realization of types across time are i.i.d. In contrast, a hidden Markov model allows for specific dynamics where the probabilities of types at t depends on the actually-realized type at $t-1$. In this sense, HMMs encompass FMMs, since i.i.d. realizations are particular cases of Markov chains where the dynamics is trivial. A Markov chain is actually an i.i.d. stochastic process if all rows in the transition probability matrix are identical.

Examination of the average transition probability matrices in Tables 3, 4, and 6 shows that the matrix rows in each case are relatively close. This means that, at the average level, the dynamics can be described as being close to i.i.d., an observation which remains valid when conditioning by treatment or even when restricting to those participants classified as mostly Bayesian or mostly non-Bayesian. Of course, such a conclusion can only be reached by actually examining the dynamic model, and it would already be interesting to conclude that, allowing for temporal dynamics, the data reveal the model to be well-represented by an FMM. This observation, however, is not entirely accurate. The reason is that the matrices in Tables 3, 4, and 6 are *averages*. At the individual level, transition probability matrices do capture temporal dynamics which, in some cases, is far from an i.i.d. model.

To provide a quantitative idea of this variance, we computed the Euclidean distances across all four row vectors in each individual transition probability matrix, and then computed the maximum of those distances for each individual. Figure F.1 displays a histogram of the resulting distribution. Rather than being concentrated at or near zero, the distribution has a definite shape, with median 0.160 and average 0.193 ($SD = 0.110$).

Figure F.1: Histogram of the maximum distance across row vectors for individual transition probability matrices.

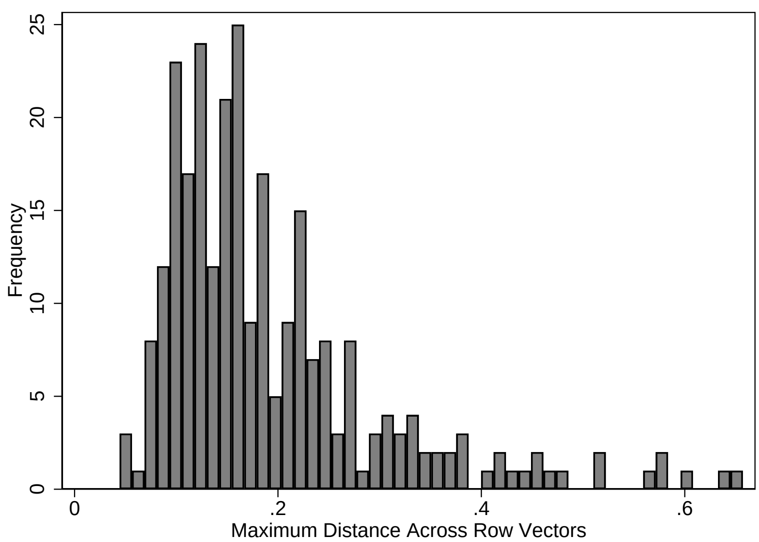


Table F.1: Summary of the analysis of the temporal dynamics for four selected individuals.

Maximum

	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	57.78%	37.20%	4.50%	0.52%	12.83%
RL	58.51%	37.25%	3.46%	0.80%	37.42%
Inertia	57.89%	39.47%	0.01%	2.63%	23.97%
Non-up	14.29%	85.71%	0.01%	0.01%	25.77%
Invariant:	57.76%	37.64%	3.90%	0.70%	0.657

75% quartile

	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	61.00%	35.33%	3.33%	3.33%	1.42%
RL	59.83%	36.01%	3.60%	0.55%	24.58%
Inertia	45.45%	51.52%	0.01%	3.03%	39.41%
Non-up	24.77%	59.75%	8.47%	7.01%	89.99%
Invariant:	59.02%	35.59%	3.19%	2.20%	0.228

Median

	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	15.74%	65.48%	5.58%	13.20%	4.75%
RL	21.27%	63.35%	4.66%	10.71%	19.71%
Inertia	11.11%	75.56%	4.44%	8.89%	51.85%
Non-up	21.24%	64.60%	1.77%	12.39%	42.66%
Invariant:	19.72%	64.46%	4.50%	11.31%	0.160

25% quartile

	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	8.05%	25.29%	4.60%	62.07%	32.07%
RL	10.47%	24.03%	5.42%	60.08%	25.23%
Inertia	9.61%	30.77%	7.69%	51.92%	32.86%
Non-up	8.14%	26.08%	4.98%	60.80%	27.11%
Invariant:	8.81%	25.73%	5.20%	60.26%	0.120

As a further illustration, Table F.1 gives four examples of actual individual transition probability matrices, with the associated estimated error rates and invariant distributions. The bottom-right cell in each table is the maximum distance across row vectors in the matrix. Those serve as an illustration for the meaning of the values (theoretically, the variable could range from 0 to $\sqrt{2} \simeq 1.414$, but higher values correspond to rather extreme cases). The four individuals correspond to the maximum value (top-left), 75%-quartile (top-right), median (bottom-left), and 25% quartile (bottom-right).

G Temporal Dynamics by Types

We report here the results of the HMM analysis conditional on subjects' classification, based on the most likely state in the individual invariant distribution. Table G.1 presents the average transition probability matrices, state-dependent error rates, and invariant distributions for subjects classified as mostly following Bayes' rule, reinforcement, inertia, and non-updating, respectively. Figure G.1 presents the corresponding graphical illustrations, including the tests across incentive conditions (low vs. high incentives).

Table G.1: Summary of the analysis of the temporal dynamics by type, individual averages with Std.Dev. in parentheses.

Bayesian subjects

N=130	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	64.85%	30.15%	2.97%	2.02%	10.07%
	(5.55)	(5.60)	(1.10)	(1.13)	(11.39)
RL	64.74%	30.35%	2.91%	2.00%	22.18%
	(6.24)	(6.16)	(13.44)	(1.18)	(14.95)
Inertia	63.41%	31.61%	3.21%	1.77%	39.44%
	(10.49)	(10.23)	(3.55)	(2.54)	(12.73)
Non-up	64.93%	29.90%	3.17%	2.01%	45.16%
	(15.44)	(15.21)	(5.19)	(3.37)	(23.92)
Invariant:	64.78%	30.25%	2.97%	2.01%	

Reinforcement subjects

N=75	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	25.96%	59.24%	7.93%	6.87%	12.80%
	(6.17)	(6.22)	(3.10)	(3.06)	(14.26)
RL	25.73%	59.25%	7.71%	7.30%	24.33%
	(5.15)	(5.39)	(2.28)	(2.68)	(16.17)
Inertia	26.44%	59.20%	7.48%	6.88%	38.67%
	(7.47)	(8.84)	(3.94)	(3.46)	(10.66)
Non-up	24.77%	59.75%	8.47%	7.01%	42.07%
	(6.87)	(7.22)	(4.12)	(4.33)	(25.04)
Invariant:	25.78%	59.28%	7.80%	7.14%	

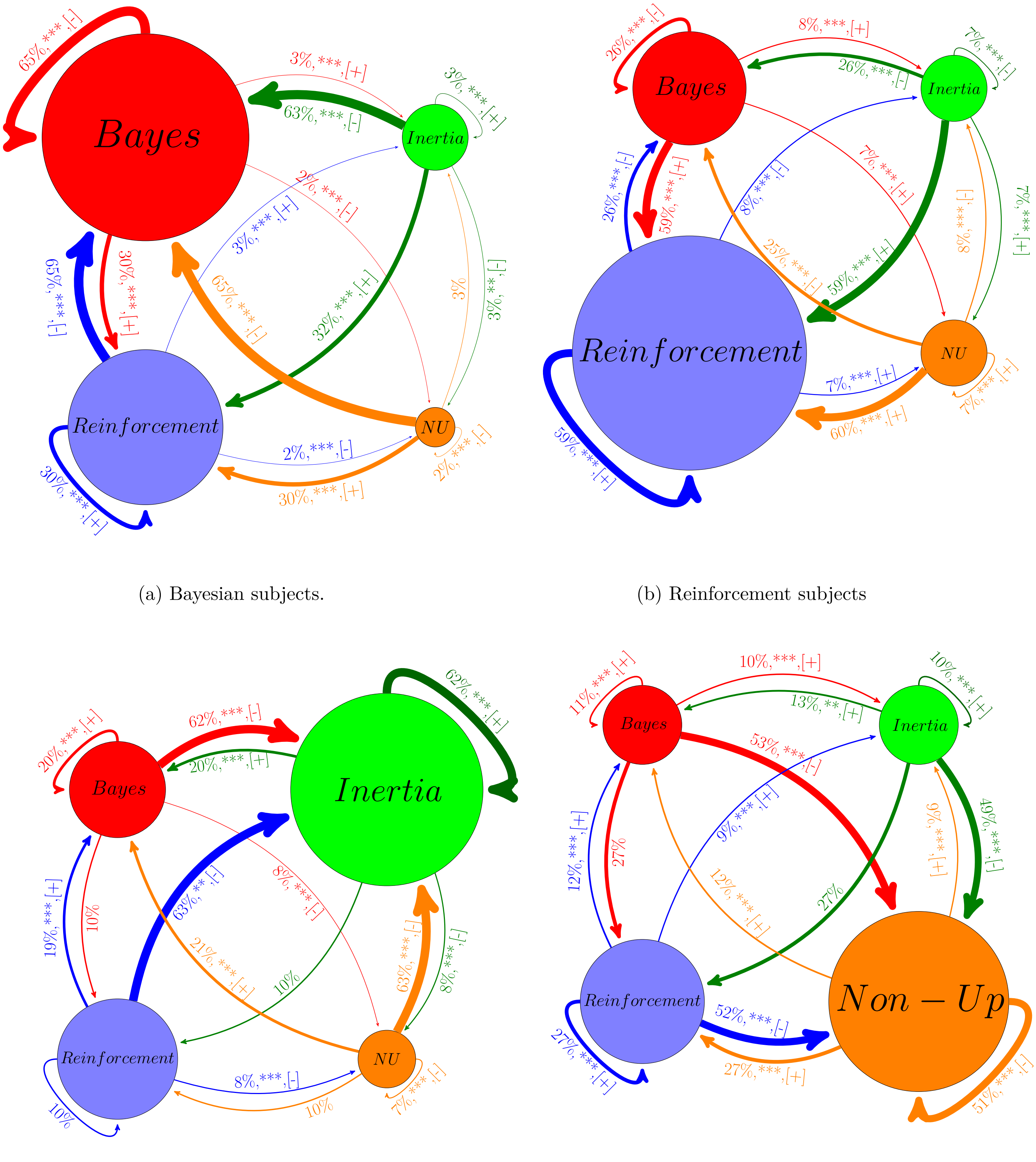
Inertia subjects

N=44	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	20.00%	9.88%	62.45%	7.65%	17.22%
	(5.64)	(2.32)	(4.48)	(3.02)	(15.46)
RL	19.27%	10.06%	62.89%	7.79%	20.14%
	(6.87)	(3.29)	(5.31)	(3.90)	(13.31)
Inertia	19.83%	10.17%	62.44%	7.56%	39.18%
	(5.51)	(1.15)	(3.56)	(2.48)	(10.98)
Non-up	20.52%	10.08%	62.84%	6.56%	42.42%
	(8.73)	(2.95)	(7.32)	(3.37)	(19.51)
Invariant:	19.86%	10.10%	62.52%	7.53%	

Non-updating subjects

N=19	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	10.57%	26.86%	10.04%	52.52%	13.64%
	(4.07)	(4.66)	(6.39)	(9.40)	(10.00)
RL	12.09%	26.80%	9.35%	51.76%	28.50%
	(2.78)	(3.23)	(5.10)	(10.47)	(18.61)
Inertia	12.89%	27.45%	10.17%	49.48%	39.89%
	(4.79)	(5.26)	(5.84)	(11.03)	(9.17)
Non-up	11.90%	27.39%	9.34%	51.37%	30.94%
	(2.87)	(3.58)	(5.28)	(10.87)	(18.51)
Invariant:	11.89%	27.17%	9.50%	51.44%	

Figure G.1: Average probabilities of switching from one behavioral rule to another conditional on types. Stars indicate MWW between low and high incentives, $*** = p < .01$, $** = p < .05$, $* = p < .10$. [+/−] indicates a significant increase from low to high [+] or vice versa [−]. Arrow thickness indicates the probability to transition from one state to another. Size of the circle indicates the ergodic distribution.



(a) Bayesian subjects.

(b) Reinforcement subjects

(c) Inertia subjects

(d) Non-updating subjects

H Translated Experimental Instructions

The original instructions were in German. In the following instructions, white balls were associated with the key “F” (this was counterbalanced in the experiment). Also, these instructions correspond to the participants who made decisions first for four-ball urns, then for six-ball urns. The order was reversed for half of the participants.

Text in brackets [...] was not displayed to subjects.

[General Instructions]

The experiment consists of two parts. In the first part you will play a decision-making game, in which balls are extracted from an urn. Your payment (which you will receive at the end of the experiment) depends on your decisions and on chance. This will be explained below. Additionally, and in any case, you will receive 2.50 Euro for showing up punctually for this experiment.

The second part of the experiment is a questionnaire.

At the end of the experiment, the total amount will be paid in cash and anonymously (payment for the first part, plus Euro 2.5 for participation in the experiment). Please read now the instructions for the first part carefully.

[Instructions for the Task]

In this game, a container (an “urn”) is presented to you on the screen. The urn contains black and white balls. The computer extracts balls from this urn. The aim of the game is to correctly predict as often as possible whether a white or a black ball is drawn. The following instructions will first explain the elements in the screen and the operation through the keyboard. then the exact rules of the game will be explained.

[Figure H.1 was displayed here, with numbers (1), (2), (3) identifying the areas. This figure displays an actual screenshot from the experimental task.]

(1) The urn is displayed in the center of the screen. There are four balls in this urn at the start of a run. There are black and white balls. They are displayed in blue on the screen as long as they are “hidden” in the urn. Below the urn you will later see which balls have already been drawn in this round.

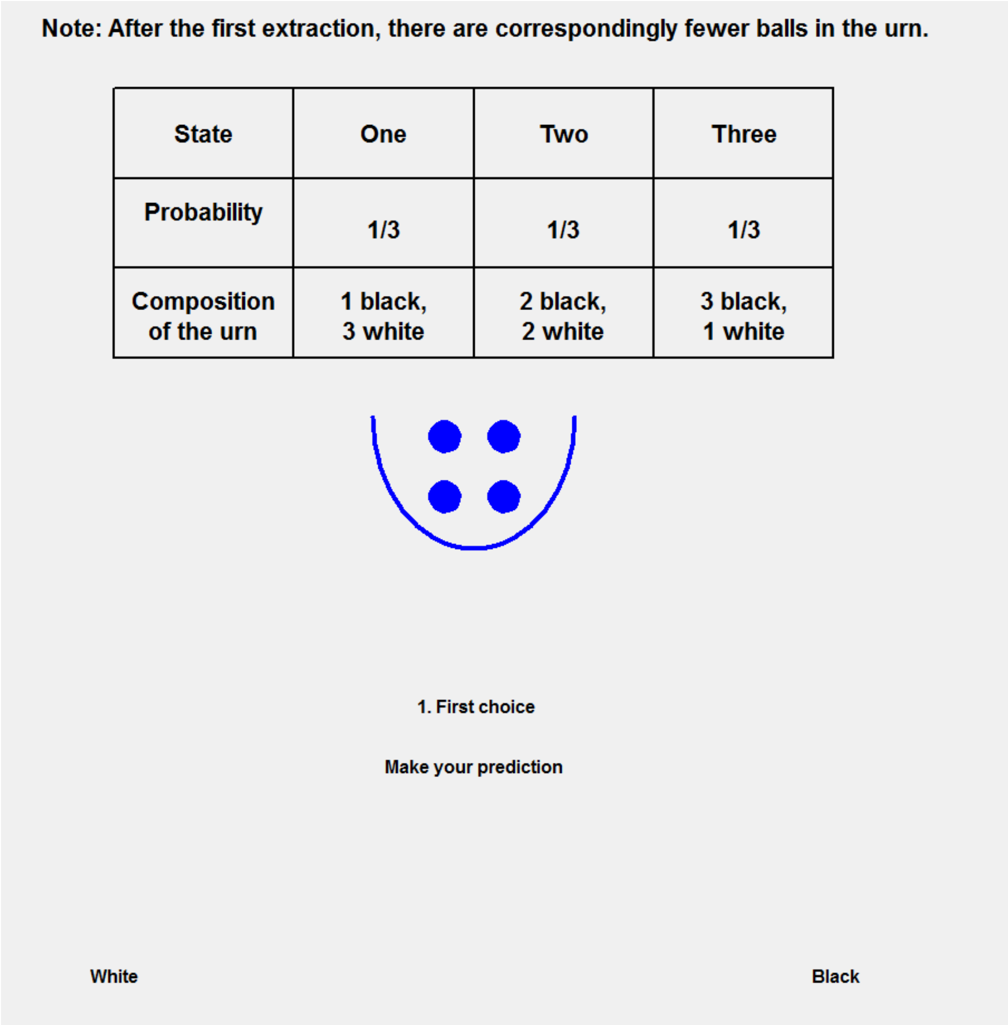
(2) At the bottom (center) you will be informed how many trials you have already completed.

(3) In the upper area of the screen, important information regarding the game rules is summarized.

Operation

The task involves only three keys. Two keys are marked in yellow in your keyboard. With the left (“F”) you make a prediction for a white ball and with the right (“J”) you make a prediction for a black ball. Use the spacebar to start a new run when prompted to do so on screen.

Figure H.1: Example screen of a trial in the experiment (4 balls).



State	One	Two	Three
Probability	1/3	1/3	1/3
Composition of the urn	1 black, 3 white	2 black, 2 white	3 black, 1 white

Rules of the Task

There will be a total of 60 trials, each of which consists of two predictions and two extractions. At the start of each trial, you predict the color of the ball (black or white) that will be drawn first from the urn. Chance determines which of the balls is drawn from the urn. The result (black or white) is displayed under the urn. You can then make your prediction for the color of the next ball to be drawn. The result of the second extraction is also displayed under the urn.

Note: An extracted ball is not put back in the urn. Rather, it is thrown away! After the first extraction, there will be only three balls left in the urn. Only in the next trial the urn will be refilled with four balls.

Earnings

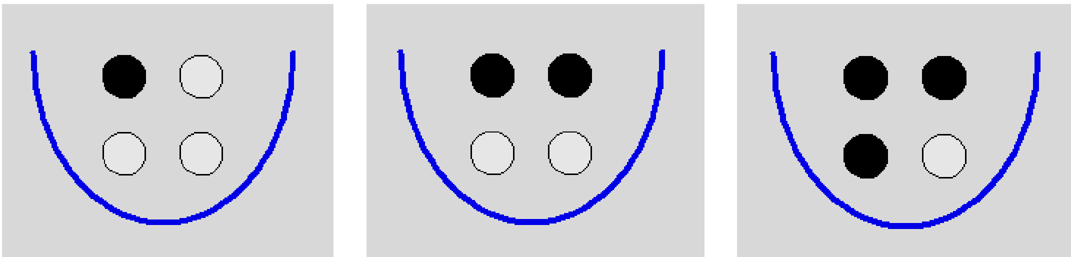
For every correct prediction you receive 18 cents. *[30 cents for the high-incentives treatment]* This means that you will receive this amount if you have predicted a black ball and a black ball is actually drawn, as well as if you have predicted a white ball and a white ball is drawn. If you predicted the wrong color, you will receive nothing.

States of the World

The most important thing about this game is that you understand how many black and white balls are in the urn. In this experiment, a distinction is made between three possible states of the world:

1. In the first state of the world (left) the urn contains 1 black and 3 white balls.
2. In the second state of the world (center) the urn contains 2 black and 2 white balls.
3. In the third state of the world (right) the urn contains 3 black and 1 white ball.

Table H.1: Explanation of the possible urn compositions.



Note: The pictures show only the number of black and white balls in the different states of the world, not the exact positions of the balls. The balls are mixed in the urn.

As mentioned earlier, in each trial you will first make a prediction for a color, after which a ball will be drawn from the urn. The drawn ball will be thrown away. Then you will make another prediction and the next ball will be drawn. However, you do not know whether the first, second or third state of the world prevails for the respective run. This is randomly determined by the computer for each individual trial. The chance of finding the first, second or third state of the world in one run is one third each time.

Please note: The state of the world does not change during a trial!

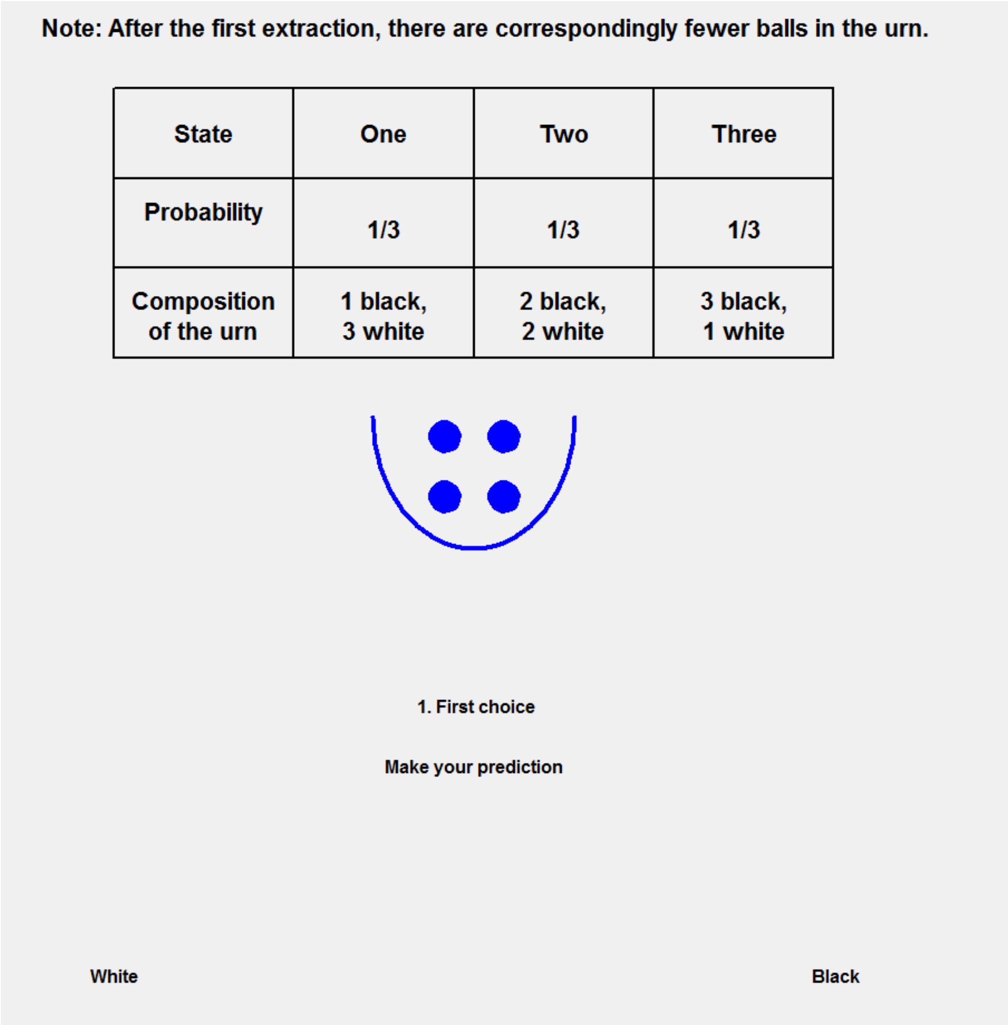
The following table summarizes again the distribution of black and white balls in the different states of the world.

[A table stating "1, 2, 3," then "1/3, 1/3, 1/3", and the three pictures of urns above was included here.]

In order to make as many correct predictions as possible, and hence earn as much money as possible, it is important that you have understood this table. If something is not clear or if you have any questions about the experiment, please raise your hand and wait until an experimenter approaches you.

The table above is valid for the first half of the 60 trials. After 30 trials, the number of black and white balls in the urn will change. When the time comes, you will receive the corresponding information about the new distribution of the balls.

Figure H.2: Example screen of a trial in the experiment (6 balls).



[After 30 trials, participants were given a similar set of written instructions explaining the task with urns containing 6 balls instead of 4. See Figure H.2].

References

Palminteri S, Wyart V, Koechlin E (2016) Computational Cognitive Neuroscience: Model Fitting Should not Replace Model Simulation.