Part-Time Bayesians: Incentives and Behavioral Heterogeneity in Belief Updating

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

[1]University of Zurich

March 23, 2022

Accepted in Management Science

Abstract

Decisions in management and finance rely on information that often includes win-lose feedback (e.g., gains and losses, success and failure). Simple reinforcement then suggests to blindly repeat choices if they led to success in the past and change them otherwise, which might conflict with Bayesian updating of beliefs. We use Finite Mixture Models and Hidden Markov Models, adapted from machine learning, to uncover behavioral heterogeneity in the reliance on difference behavioral rules across and within individuals in a belief-updating experiment. Most decision makers rely both on Bayesian updating and reinforcement. Paradoxically, an increase in incentives increases the reliance on reinforcement, because the win-lose cues become more salient. **JEL Classification:** G41 · D91 · C91

Keywords: Bayesian updating, Incentives, Reinforcement, Heterogeneity, Finite Mixture Models, Machine Learning

1 Introduction

Overwhelming evidence shows that human decision makers have a limited grasp of probabilities and, very especially, of how beliefs should be updated in the face of new information. Previous research has identified a veritable catalogue of deviations from Bayesian updating, most of them taking the form of heuristics and biases that become relevant when certain informational triggers are present (e.g., Kahneman Tversky, 1972; Grether, 1980; Camerer, 1987). Numerous studies have demonstrated that those deviations affect and distort financial decisions (e.g., Shleifer, 2000; Thaler, 2005; Shiller, 2005; Asparouhova et al., 2015; Frydman Camerer, 2016). In this work, we are interested in a kind of deviation from Bayesian updating which is particularly relevant for management and finance. Decisions in those fields often rely on information that includes win-lose feedback: gains and losses, success and failure, beating the competition or not, etc. Indeed, as shown by Knutson Bossaerts (2007), gains and losses are crucial to understand the neural foundations of financial decisions.

Whenever information carries a win-lose component, elementary reinforcement behavior (e.g., Thorndike, 1911; Sutton Barto, 1998; Schultz et al., 1997) dictates to repeat choices if they led

*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

to success in the past, and change them if they led to failure.[1] This kind of simple "win-stay, lose-shift" behavior (Thorndike's "law of effect") is in line with model-free reinforcement learning, which has been studied in recent contributions in neuroscience as opposed to model-based (reinforcement) learning, which itself can be construed as incorporating Bayesian behavior (e.g., Daw et al., 2005; Doll et al., 2012; Feher da Silva Hare, 2020). For example, Daw et al. (2011) show that, in a two-stage Markov decision task, subjects' choices seem to follow a mixture of model-based and model-free reinforcement learning. In a related task, Gläscher et al. (2010) find support for the coexistence of two distinct neural signatures corresponding to these two unique forms of learning. Crucially for our purposes, Beierholm et al. (2011) study switching between model-based and model-free reinforcement and estimate a high within-individual switching rate (of around one-third of the time).

We hypothesized that model-free reinforcement (win-stay, lose-shift behavior) is the main reason that financial actors often fail to conform with Bayesian updating in decision problems where new information allows to update prior beliefs and that information contains a win-lose component. In view of the neuroscience literature, we carried out a decision-making experiment to explore the idea that there is substantial heterogeneity *across individuals*, with some agents being "more Bayesian" than others, and also *within each given individual* (as suggested by Beierholm et al., 2011). Specifically, our hypothesis was that decision makers mostly follow either noisy versions of Bayesian updating or model-free reinforcement, but we did not exclude a priori the involvement of other candidates that the literature has proposed as playing a potential role.

Specifically, we consider two additional behavioral phenomena. The first is *conservatism bias* or non-updating behavior, which reflects a general failure to update beliefs, and, taken to the extreme, identifies the posterior with the prior (e.g., Edwards, 1968; Navon, 1978; El-Gamal Grether, 1995). For example, in financial markets, Barberis et al. (1998) argued that conservatism might explain underreaction to news. The second is *decision inertia*, which refers to the repetition of previous choices independently of feedback (Pitz Geller, 1970; Alós-Ferrer et al., 2016). For instance, it has been shown that a large fraction of stock account owners exhibit portfolio inertia (Bilias et al., 2010). While these phenomena represent broad behavioral tendencies, in the framework of our experiment they will correspond to precisely-defined behavioral rules, which we take as representative of behavior corresponding to neither Bayesian updating nor model-free reinforcement. We are particularly interested in these rules because they both capture feedback-independent deviations from Bayesian updating (hence contrasting with model-free reinforcement) involving low cognitive costs.

The second question we were interested in is whether, and if so, how our classifications would be affected by monetary incentives and their magnitude. On the one hand, increasing incentives should motivate decision makers to spend more cognitive effort on the task, thereby behaving more in accordance with the prescriptions of Bayes' rule. On the other hand, reinforcement tendencies are triggered by win-lose cues, and if incentives are increased, both wins and losses become larger in magnitude, hence making the triggers more salient. In particular, in an experiment on belief updating, Achtziger Alós-Ferrer (2014) showed that when the cue which triggers reinforcement is not paid, compared to when it is incentivised, participants rely less on win-stay/lose-shift. However,

[1] Charness Levin (2005) provide an intuitive example illustrating reinforcement behavior in managerial decisions. Imagine a rookie employee is sent to close a business deal because a more experienced negotiator is unavailable, and achieves good results. Next time a similar deal has to be closed, arguments like "never change a winning horse" might prompt the CEO to send the rookie again. This, however, neglects the possible informational content of the previous outcome, which might imply that a more experienced negotiator could achieve even better results.

Asparouhova et al. (2015) showed that, in a market situation, participants sometimes avoided belief updating under higher incentives, possibly for fear of making (costly) mistakes, relying on simpler behavioral rules instead. Thus, we will specifically ask how the reliance on different behavioral rules is affected by the magnitude of incentives.

To answer the questions posed above, we collected behavioral data using a novel experimental paradigm designed to disentangle four decision rules in belief updating (Bayes' rule, reinforcement, conservatism, and decision inertia) and varied the level of monetary incentives. To obtain a classification of decision makers and study its dependence on incentives, we follow two complementary approaches. First, we rely on *finite mixture models* from statistics (e.g., Frühwirth-Schnatter, 2006), which provide classifications on the basis of choice data accounting for heterogeneity, unobservable determinants of behavior, and rule-specific errors. These models have received increasing attention for the analysis of heterogeneity in economic behavior (e.g., Costa-Gomes et al., 2001; Bellemare et al., 2008; Bruhin et al., 2018; Barron, 2020). In agreement with our hypothesis, we find that roughly half of our sample follows non-Bayesian rules of behavior, with model-free reinforcers being the most numerous group among those. The other half is classified as mostly following Bayes' rule, but with significant error rates. Additionally, we find no performance improvement with higher incentives for Bayesian participants, but *non-Bayesian* decision makers do become significantly better. This suggests a heterogeneous effect of incentives. For half of our sample, low incentives suffice to spark cognitive deliberation and the use of noisy approximations of Bayes' rule. For these individuals, higher incentives do not result in additional improvements, in agreement with a ceiling effect. The rest of the sample, however, follows mostly model-free reinforcement, resulting in a lower performance. For these participants, higher incentives do result in increased reliance on Bayes' rule (but not enough to result in the subjects' reclassification as mostly Bayesians), and hence an increase in performance.

While this first strategy of analysis concentrates on heterogeneity across individuals, our second approach turns to heterogeneity *within* each individual, and specifically the temporal dynamics of the data. We ask the question of whether a given decision maker might rely on different decision rules over time and whether incentives do affect the balance. For this purpose, we implement an identification strategy adapted from the machine learning literature. Specifically, we rely on *Hidden Markov Models* (Rabiner, 1990; Frühwirth-Schnatter et al., 2019), which have been previously used in economics to, for example, identify switching among learning rules in repeated games (Ansari et al., 2012; Shachat et al., 2015) or to study bidding heuristics in auctions (Shachat Wei, 2012).

In our setting, the idea is that each behavioral rule corresponds to an unobservable (hence "hidden") state of an individual-level Markov chain whose transition probabilities capture the dynamics of behavior over time. Thus, we estimate the transition probabilities, which determine the long-run probabilities of the behavioral rules. We find that most subjects exhibit relatively large probabilities for both Bayes' rule and model-free reinforcement, as well as relatively large transition probabilities between those. Overall, a picture arises where some subjects rely mostly on Bayes' rule even if incentives are low, but occasionally follow model-free reinforcement and other rules, and either hit a ceiling when incentives are increased or even suffer detrimental effects due to increased reliance on model-free reinforcement as the win-lose cues become more salient. Other subjects rely predominantly on model-free reinforcement but occasionally follow Bayes' rule. For those, an increase in incentives is essentially detrimental. Last, some subjects mostly follow rules other than Bayes' rule and (model-free) reinforcement and generally achieve low levels of performance. For those, an increase in incentives is beneficial because it increases the transition probabilities toward Bayes' rule.

Our work follows upon previous literature on the classification of decision makers in terms of behavioral rules in the domain of belief updating. This includes both the distinction between model-free and model-based reinforcement learning (Daw et al., 2005; Doll et al., 2012) and previous work using finite mixture models. For example, in a belief-updating experiment, El-Gamal Grether (1995) showed that different decision makers favored different behavioral rules, but Bayes' rule was the most frequently used at the population level. That is, humans generally fail to perfectly follow Bayes' rule, but the latter is still a good representation of the behavior of a majority of people for a large proportion of the time. Of course, both observations are only compatible if there is heterogeneity in the reaction to new information and how beliefs are updated.

We contribute to the related literature on heterogeneity in belief updating in three ways. First, compared to existing paradigms (i.e., Charness Levin, 2005; Knutson Bossaerts, 2007), our new experimental task allows to explicitly consider and disentangle more than two behavioral rules (and, in particular, contrast model-free reinforcement learning to feedback-independent deviations from Bayesian updating which have been found to be relevant in the financial decision-making literature). Second, compared to previous classification contributions as El-Gamal Grether (1995), we take a step further in the identification of heterogeneity by investigating potential temporal dynamics, i.e., allowing for within-subject heterogeneity in the reliance on different behavioral rules within the course of the experiment. Third, we explicitly target the effect of incentives on heterogeneity, i.e., whether higher incentives affect reliance on one particular rule.

The most closely-related contribution to ours is Payzan-LeNestour Bossaerts (2015), which relied on a multi-armed bandit task. Participants mostly followed model-free reinforcement, but switched to Bayesian learning when nudged into paying attention to crucial statistics of the environment. We share with this work a focus on the comparison between Bayesian behavior and model-free reinforcement, and an interest in switching behavior. The main difference is that Payzan-LeNestour Bossaerts (2015) evaluate which of different candidate models describe the data better, while we consider within-individual heterogeneity, that is, we examine to which extent decision makers are both Bayesians and reinforcers. Other differences concern the specific implementation and task complexity. For instance, our experimental task belongs to the class of static settings typically used to study heuristics and biases, where priors are reset after each relevant decision and each trial is independent of (and equivalent to) the others. In contrast, as in many contributions contrasting model-free and model-based reinforcement, the task of Payzan-LeNestour Bossaerts (2015) is inherently dynamic, with the expected payoffs of the bandit's arms changing over time.

Our contribution is further related to a strand of papers using a different, simpler paradigm contrasting win-stay, lose-shift behavior and Bayesian updating, where, as in our case, information is endowed with a win-lose cue. In this paradigm, Charness Levin (2005) found very high error rates which would be consistent with model-free reinforcement behavior. By examining response times in the same paradigm, Achtziger Alós-Ferrer (2014) argued that the data could be well-explained by a dual-process model where simple reinforcement conflicts and interacts with more deliberative belief updating. Achtziger et al. (2015) examined neural evidence for simple reinforcement in this task using the electroencephalograph (EEG) and found that subjects with higher error rates under high incentives exhibited larger amplitudes in (extremely early) brain potentials linked to reinforcement learning. A possible interpretation in agreement with our results is that, even if larger monetary rewards increase effort, they also increase the salience of the win-lose cues which trigger reinforcement, hence creating a "reinforcement paradox" where higher incentives increase reliance on this alternative process instead of on Bayesian behavior. In line with this paradox, a pupil-dilation study (Alós-Ferrer et al., 2019) has recently shown that higher incentives in this paradigm do increase cognitive effort while failing to result in generally increased performance.

The paper is structured as follows. Section 2 discusses the experimental design and the behavioral rules. Section 3 provides a descriptive overview of the data. Section 4 applies finite mixture modeling to obtain a classification of decision makers, studies the effects of incentives on that classification, and briefly discusses response times. Section 5 applies a hidden Markov model to study the temporal dynamics, the effects of incentives, and the differences across behavioral types. Section 6 concludes.

2 Design and Procedures

2.1 Behavioral Rules and Experimental Design

We designed a novel belief-updating paradigm with the explicit objective of disentangling different behavioral rules when new information carries a win-lose component. Since reinforcement is relevant for a wide range of conceptually different decisions, we developed a frame-free, abstract paradigm in order to study behavioral heterogeneity free of potential confounds which could arise from particularities present in one application but absent in others. In view of our motivation, it was important to focus on binary decisions which would provide a simple win-lose feedback while giving an opportunity to update beliefs on an underlying state of the world. An additional concern was to develop a paradigm rich enough to allow disentangling Bayesian updating and (model-free) reinforcement from the other two rules of interest (conservatism and decision inertia). The paradigm we developed belongs to the larger class of urn tasks, which have been extensively used to study biases in belief updating, e.g. for the case of representativeness and conservatism (Grether, 1980, 1992) or for the comparison of Bayesian updating and reinforcement rules (Charness Levin, 2005; Achtziger Alós-Ferrer, 2014; Achtziger et al., 2015).[2]

The essence of the paradigm is as follows. Participants are presented with three covered urns, each containing balls of two possible colors (black or white) in different but known proportions. One of the three urns is chosen at random and a single ball is randomly extracted from it. Participants know the proportion of balls in each urn, that the actual urn is one of the three described ones, and that the urn has been selected randomly, with equal probabilities for each urn. Crucially, the ball is *not* replaced after the first extraction. Then, a second ball is extracted at random *from the same urn*. The participant's decisions are bets. Before each of the two extractions, participants bet on the *color* of the extracted ball. We are interested in the second betting choice, since the color of the first extracted ball allows updating the belief regarding the urn from which the balls are extracted from.[3]

[2]Received belief updating tasks from the decision-making literature are typically simpler and insufficient for our purposes. For example, in the task of Charness Levin (2005), Achtziger Alós-Ferrer (2014), and Achtziger et al. (2015), in half of the possible cases the Bayesian prescription is "obvious" (and error rates are extremely low) because, actually, it coincides with the prescription of reinforcement, while in the remaining cases the two rules conflict (and error rates are extremely high). Further, the task produces only four possible decision cases in total.

[3]The design is related to Asparouhova et al. (2015), who investigate belief updating based on sampling without replacement. As in our case, in that paradigm the first draw is of little interest in itself.

Figure 1: Experimental design. (A) Four-balls (top) and six-balls trials (bottom). (B) Prescriptions of the four candidate behavioral rules for the second bet, classified by first bet and received stimulus. (● = black ball extracted / bet on black ball, ○ = white ball extracted / bet on white ball).

(A)

● ○ ○ ○	● ● ○ ○	● ● ● ○
Prob. 1/3	Prob. 1/3	Prob. 1/3

● ○ ○ ○ ○ ○	● ● ● ○ ○ ○	● ● ● ● ● ○
Prob. 1/3	Prob. 1/3	Prob. 1/3

(B)

	First Bet	Stimulus	Bayesian	Reinforcement	Inertia	Non-updating
4 Balls	●	● Win	○	●	●	○
	●	○ Lose	●	○	●	●
	○	○ Win	●	○	○	●
	○	● Lose	○	●	○	○
6 Balls	●	● Win	●	●	●	○
	●	○ Lose	○	○	●	●
	○	○ Win	○	○	○	●
	○	● Lose	●	●	○	○

Incentives are straightforward. After each extraction, participants are paid a constant amount if and only if the color of the ball matches their bet. After two extractions the trial ends and all balls are replaced in the urns, before a new trial starts. Participants are aware that all trials are independent from each other, so the urn from which the two balls are extracted is randomly and independently determined according to the same uniform prior in each of the 60 repetitions.

It is important to note that our focus is on distinguishing behavioral rules and not on dynamic learning, in contrast to the literature investigating model-free and model-based reinforcement learning (e.g. Payzan-LeNestour Bossaerts, 2015). By design, there is limited scope for learning in our paradigm, as trials are independent, and hence the situation "resets" after each choice, which is very different from the two-stage or multi-armed-bandit tasks which are commonly used in that literature.

To be able to disentangle our candidate behavioral rules, there are two types of trials using two different urn compositions, as depicted in Figure 1(A). In one type, each urn contains four balls. In the other type, each urn contains six balls. In both cases, one of the urns contains exactly one black ball, the other contains exactly one white ball, and the third urn contains half each of black and white balls. Although the alternative urn compositions are superficially similar, and are both easy to understand, the prescriptions of the candidate rules (in particular, of Bayesian updating of beliefs) are quite different across the two kinds of trials, which in turn allows us to discriminate among them. In the experiment, participants were reminded of the composition of the urns and the number of balls in each urn at all times, as this information was prominently displayed in each trial.

The first of the rules we are interested in is optimization following Bayesian updating of the prior, or simply *Bayes' rule* for short. Bayes' rule captures the normatively correct way to integrate new information with prior beliefs, but it has been widely shown to perform poorly as a descriptive rule. Empirically-documented violations in experiments involving conditional probability judgments (e.g. Grether, 1980; Charness Levin, 2005) have shown that human beings are simply not Bayesian optimizers, although Bayes' rule is sometimes a reasonable approximation of behavior (El-Gamal Grether, 1995; Griffiths Tenenbaum, 2006). We use Bayes' rule as a benchmark, as it describes the normatively optimal behavior.

Straightforward computations show that a subject who used Bayes' rule in four-balls trials and chose black for the first bet should, for the second bet, shift to white if she won the first bet and stay with black if she lost the first best (and symmetrically if she bet on white the first time). In contrast, in six-balls trials Bayes' rule prescribes the exact opposite: stay with the same color for the second bet if she won the first, and shift if she lost.[4] The prescriptions of Bayes' rule are summarized graphically in the column "Bayesian" in Figure 1(B).

The main alternative rule we are interested in is (model-free) reinforcement learning. This is the natural candidate in any setting where information comes with a win-lose feedback. Reinforcement is a basic component of human behavior (Thorndike, 1911; Sutton Barto, 1998) and refers to the tendency to repeat whatever action yielded a positive result in the past and avoid those which led to failure. In this paradigm we assume the simplest form of reinforcement i.e.,"win-stay, lose-shift" behavior. This rule of thumb has been shown to explain deviations from Bayesian behavior in belief-updating paradigms where information is extracted from previous wins and losses, as in many economic settings (Charness Levin, 2005; Achtziger Alós-Ferrer, 2014).

The prescriptions of this rule are particularly simple (see column "Reinforcement" in Figure 1(B)). In case of success participants following (model-free) reinforcement would repeat the same choice, while they would shift to the other choice after failure. As a consequence, in our binary setting this rule always prescribes to place the second bet on the color of the actually-extracted first ball, that is, chasing after the previous winner. In particular, model-free reinforcement makes identical prescriptions for four-balls and six-balls trials, which means that it coincides with Bayes' rule for six-balls trials but is completely opposed to it for four-balls trials.

The third behavioral rule we are interested in is inertia, which is the tendency to repeat previous choices independently of the outcome (Pitz Geller, 1970; Akaishi et al., 2014; Alós-Ferrer et al., 2016), and has been linked to status-quo bias (Ritov Baron, 1992). For instance, Erev Haruvy (2016) argue that humans exhibit a strong tendency to simply repeat the most recent decision, and this tendency is sometimes stronger than the tendency to react optimally to the most recent outcome.

In our paradigm, inertia prescribes that participants ignore the results of the first bet and simply repeat the previous decision, i.e. bet black after betting black the first time and white after white (see column "Inertia" in Figure 1(B)). Again, the pattern is identical for four-balls and six-balls trials, and in both cases it differs from Bayes' rule *and* from reinforcement.

We term the last behavioral rule we focus on "non-updating." This rule postulates that the prior is not updated, and that the decision maker uses a posterior identical to the prior (hence uniform). However, in our setting there is a transparent change between the first and the second ball extraction,

[4] For example, after observing a white ball in the first draw in a four-ball trial, a Bayesian should update the probabilities of the three urns to 1/2, 1/3, and 1/6, respectively, and hence the probability of extracting a black ball in the second draw, given that there is one white ball less in the urn, is $(1/3) \cdot (1/2) + (2/3) \cdot (1/3) + 1 \cdot (1/6) = 5/9 > 1/2$, leading to an optimal bet on black. In contrast, if a white ball is extracted in the first draw of a six-ball trial, the updated probabilities of the urns are 5/9, 3/9, and 1/9, respectively, and the probability of a black ball in the second draw is $(1/5) \cdot (5/9) + (3/5) \cdot (3/9) + 1 \cdot (1/9) = 19/45 < 1/2$, leading to an optimal bet on white.

which is independent of probability updating: since there is no replacement, for the second bet the selected urn contains *one ball less* (three for four-balls trials and five for six-balls trials), namely the one extracted after the first bet. This poses an incorrect but simple optimization problem. That is, this rule prescribes that participants take into account the rather-obvious fact that after the first extraction one ball is missing, but they do not engage in belief updating at all.

Direct computations show that a subject who used the non-updating rule and won would shift to the opposite color for the second bet, but would stay with the same color if she lost the first bet (column "Non-updating" in Figure 1(B)). The prediction is the same for four-balls and six-balls trials. That is, by design, in our paradigm non-updaters would always behave in direct opposition to win-stay, lose-shift reinforcement, and hence the prescriptions of this rule coincide with those of Bayes' rule for four-balls trials and are the opposite for six-balls trials.

In summary, and as shown in Figure 1(B), by using just two binary choices and two urn compositions with three urns each we can fully disentangle four different behavioral rules. For four-balls trials Bayes' rule and non-updating prescribe the same decisions, which are in direct opposition to the prescriptions of (model-free) reinforcement. For six-balls trials Bayes' rule and (model-free) reinforcement prescribe the same decisions, which are in direct opposition to the prescriptions of non-updating. In both cases, inertia always coincides with exactly half of the prescriptions of each other rule and is opposed to it for the other half.

2.2 Procedures

A total of $N = 268$ university students (142 females; age range $18 - 43$, mean $= 24.07$) were recruited using ORSEE (Greiner, 2015) from the preregistered pool of the Cologne Laboratory for Economic Research (CLER), excluding students majoring in psychology or economics since they could have been familiar with similar paradigms. The experiment was programmed using z-Tree (Fischbacher, 2007). Each participant made decisions in 60 trials, 30 with the four-balls design and 30 with the six-balls design. To avoid order effects, half of the participants (randomly assigned) worked on four-balls trials first and six-ball trials later, and the remaining participants followed the inverse order.

Participants received a performance-based payment plus a show-up fee of 2.50 Euro. To study the effects of incentives in our classifications, there were two different treatments (with data collected in different sessions). Under Low Incentives, each successful bet was rewarded with 18 Euro-cents ($N = 128$). Under High Incentives, the payoff was 30 Euro-cents ($N = 140$) for each successful bet. The size of the incentives remained constant throughout the experiment and was common knowledge.[5] All (successful) decisions were paid. In our context, this payment mechanism is incentive compatible under mild assumptions on individual preferences, as shown by Azrieli et al. (2018).[6]

Participants received detailed written instructions and answered 9 control questions regarding the replacement of balls, states of the world, and trial independence. To ensure that participants

[5] We implemented a between-subject treatment of incentives as common in the literature (e.g., Barron, 2020) in order to avoid order effects and other confounds.

[6] Specifically, the pay-all mechanism is incentive compatible whenever subjects' preferences fulfill a condition called "no complementarities at the top" when evaluating bundles of outcomes (Azrieli et al., 2018). In our case, this assumption is immediately fulfilled as long as participants prefer more money over less (but the assumption might impose stronger restrictions in more complex environments). In the second decision within every trial, one option is dominated by the other and participants always have a strict incentive to choose the urn they believe to be more likely. Then, choosing the correct option in all trials dominates any other alternative bundle of choices, hence making this payment mechanism incentive-compatible according to Azrieli et al. (2018).

understood the task, those who got any control questions wrong were provided with further information by the experimenter, until a sufficient understanding of the task was reached. There were no practice trials. No time limit was imposed during the experiment; participants were free to use as much time as they needed. At the end of the experiment, participants completed some (non-incentivized) questionnaires and provided demographic information (gender, age, and field of studies). A session lasted about 90 minutes and average earnings were 17.93 Euros ($SD = 1.74$).

3 A Descriptive View of the Data

3.1 Choice Frequencies

The first bet is never of interest in itself, as given the prior both choices are equally likely to result in a payoff, and given the symmetry of the design, no behavioral rule makes a specific prescription. We are interested in the second bet within each trial. For those, rules generally make different prescriptions, which depend on the first bet and on the outcome of the first extraction.

Participants failed to consistently make optimal decisions (those prescribed by Bayes' rule). For the second bet within each trial, the average across individuals of the percentage of correct decisions was 61.23% (Median= 58.33%, SD=15.53, Minimum=20.00%, Maximum=100.00%), which was significantly different from random performance according to a Wilcoxon Signed-Rank (WSR) test ($N = 268$, $z = 14.193$, $p < .001$).

At the aggregate level, there were no significant differences in the percentage of correct answers across incentive levels according to a Mann-Whitney-Wilcoxon test (MWW; low incentives, average 62.54%; high incentives, 59.80%; $N = 268, z = 1.242, p = .214$). The percentage of correct answers in four-ball trials (average 58.22%) was lower than in six-ball trials (average 64.24%), although the difference did not reach significance (WSR test, within subject, $N = 268$, $z = -1.626$, $p = .104$). The (counterbalanced) order did not affect the percentage of correct answers (four-ball trials first, average 62.22%; six-ball trials first, average 57.25%; MWW test, $N = 268$, $z = -0.418$, $p = .676$).

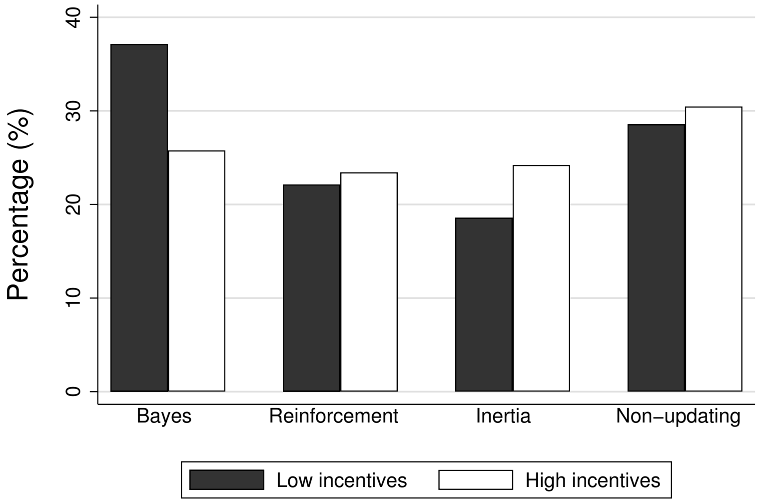
There is no evidence that the subjects "figured out" how to solve the task during the course of the experiment. In particular, there is no discernible trend for the low-incentive treatment (random effects Probit regression on the probability of a correct choice regressed against the trial number, Coef.= 0.001, $t = 0.50$, $p = .616$), indicating that there was no learning at the aggregate level. For the high-incentive treatment, we observe a marginally significant but very modest positive trend (random effects regression, Coef.= 0.002, $t = 1.87$, $p = .061$), which should not be overinterpreted.

3.2 A Naïve Classification

As an illustration, we start with a simple but naïve classification approach based on observable data, similar to previous approaches in the literature, before we turn to our finite-mixture and hidden-Markov analyses. We examine choices at the individual level and classify a subject as following a particular behavioral rule if his or her decisions coincide with the prescriptions of that rule more than the others (i.e., is she follows Bayes' rule more frequently than any of the alternative rules).

This simple procedure classifies 31.72% of subjects as following Bayes' rule, 22.76% as reinforcers, 29.48% as non-updaters, and only 21.27% as relying on inertia. Figure 2 displays the proportion of subjects classified as following each behavioral rule depending on the incentive condition. However, in

Figure 2: Naïve classification of subjects, by treatment.



this simple classification incentives seem to have limited or even detrimental effects, in the sense that there are no significant differences in the proportions assigned to each rule across incentive treatments for most rules (test of proportion; Reinforcers, $p = .801$; Inertia, $p = .259$; Non-updaters, $p = .733$) with the notable exception of Bayes' rule. There is a lower number of subjects classified as following the normative prescriptions in the higher (37.14%) compared to lower (25.78%) incentive treatment (test of proportions, $z = 1.996$, $p = .022$).

Unsurprisingly, subjects classified as following Bayes' rule earned more on average in the task (Mean 9.31) than other subjects (Mean 9.04; MWW, $N = 268$, $z = 2.759$, $p = .006$). These results are a first indication that, in this paradigm, being Bayesian actually paid off, but increasing incentives might actually have a negative effect on performance, at least for some people. The naïve classification provides a first glimpse at behavioral heterogeneity, but it is obviously of limited interest. The reason is that it adopts the unrealistic view that decision makers follow exactly one behavioral rule and remains silent on deviations with respect to a given rule, even though a subject is still classified as following a rule if a large part of his or her decisions deviate from it. Hence, in the following section we turn to a finite-mixture modeling approach.

4 Behavioral Types and Finite Mixture Modeling

We are interested in individual heterogeneity, and hence we now turn to a different method of classification of decision makers according to the postulated behavioral rules. The task of classifying the observed choices according to behavioral rules is not trivial, because which behavioral rule was followed by each participant for each decision is not directly observable. Hence, any inference on which rule determined which decision must rely on a comparison between observed choices and the prescriptions of the rules. This is complicated by the fact that different behavioral rules might prescribe the same choice part of the time. Further, in view of overwhelming evidence on decision errors and stochastic choice, it is unrealistic to assume that behavioral rules are purely deterministic.

To solve these problems, we explicitly introduce the possibility that subjects make errors during the experiment. In the model, an error occurs whenever a subject is following a specific behavioral rule but the actual choice deviates from the rule's prescriptions. For example, a subject might follow Bayes' rule but make a (maybe computational) mistake, and choose the incorrect option. In other words, we consider noisy versions of the behavioral rules of interest. Given strictly positive error probabilities, each and every choice might *a priori* have been generated by each and every candidate behavioral rule. For instance, in a situation where reinforcement and Bayes' rule prescribe different answers, a given (erroneous) answer might be due to the subject following reinforcement, or to the subject making a mistake while actually following Bayes' rule. To untangle behavior, we now turn to a finite mixture modeling approach, which accounts for unobservable determinants of behavior, potential heterogeneity among individuals, and the possibility of errors. In the subsections below we describe how we implement finite-mixture modeling for our data and report the actual estimation.

We adopt the *supervised approach* to finite-mixture modeling, where we specify the number and characteristics of types *ex ante*. This is appropriate in our case because we have clearly-defined candidates, as the experiment was designed in order to disentangle precisely the four behavioral rules we consider. In this sense we tie our hands *ex ante* to the set of relevant behavioral rules and examine whether the evidence supports that choice.[7]

4.1 A Finite Mixture Model for Noisy Behavioral Rules

Let $B = \{0,1\}$ denote the set of possible bets or outcomes, where 0 stands for "black" and 1 for "white." For each trial and participant in our experiment, the second bet $b \in B$ is made after the first, $a \in B$, and after observing the color of the first extracted ball, $x \in B$. Further, let $t \in T = \{4,6\}$ indicate whether the trial corresponds to the four-balls or the six-balls design. Before the decision b is made, the relevant characteristics of a trial are completely identified by $\omega = (t,a,x) \in \Omega = T \times B^2$. With this notation, a behavioral rule in our setting is a mapping

$$\beta_k : \Omega \mapsto B,$$

and the four behavioral rules $\beta_1, \ldots, \beta_4$ that we consider can be written down in this format simply by translating the corresponding columns in Figure 1(B). For example, if $k = 1$ is Bayes' rule, then $\beta_1(4,0,0) = 1, \beta_1(4,0,1) = 0, \ldots, \beta_1(6,1,0) = 0$.

Since the behavioral rules are deterministic and there are no ties in our design, the choice probabilities induced by each rule β_k are given by

$$p_k(b|\omega) = \begin{cases} 1 & \text{if } b = \beta_k(\omega) \\ 0 & \text{otherwise.} \end{cases}$$

The observations $m = 1, \ldots M$ in our dataset are of the form $(\omega, b) \in \Omega \times B$. A finite mixture model assumes that a distribution of types (in our case, behavioral rules) generates the actual observations, with η_k $(k = 1, \ldots, 4)$ being the probability of type k.

[7] The alternative *unsupervised approach* attempts to identify a pre-determined number of different types of individuals in the population. In Appendix A (see Section 4.5) we implement an unsupervised approach and show that this analysis is inappropriate for our data.

To take errors into account, we consider noisy behavioral rules by introducing an error or "trembling" probability ε_k, analogously to Bruhin et al. (2010). Since we deal with binary choice, the interpretation is simple: given ω and assuming that β_k generated the observed choice, $1-\varepsilon_k$ is the probability with which the choice prescribed by β_k actually obtains, while with probability ε_k the opposite choice is made. Formally,

$$p_k(b|\omega,\varepsilon_k) = \begin{cases} 1-\varepsilon_k & \text{if } b=\beta_k(\omega) \\ \varepsilon_k & \text{otherwise.} \end{cases} \tag{1}$$

Our estimation assumes that each individual $j=1,\ldots,268$ has independent probabilities η_k^j of relying on each rule, plus individual-level error rates ε_k^j $(k=1,\ldots,4)$. Hence, the likelihood of the sub-dataset containing j's decisions, $D_j=\{(\omega_1^j,b_1^j),\ldots,(\omega_{60}^j,b_{60}^j)\}$, is

$$\text{Prob}(D_j) = \prod_{m=1}^{60}\sum_{k=1}^{4}\eta_k^j\cdot p_k(b_m^j|\omega_m^j,\varepsilon_k^j)$$

and the likelihood of the entire dataset $D=\bigcup_{j=1}^{268} D_j$ is

$$\text{Prob}(D) = \prod_{j=1}^{268}\text{Prob}(D_j).$$

Since $\ln\text{Prob}(D) = \sum_{j=1}^{268}\ln\text{Prob}(D_j)$ and $\text{Prob}(D_j)$ depends only on the variables $\eta_1^j,\ldots,\eta_4^j$ and $\varepsilon_1^j,\ldots,\varepsilon_4^j$, which in turn play no role for $\text{Prob}(D_\ell)$ with $\ell\neq j$, maximum likelihood estimation can be done independently for each participant j. The estimation then delivers which type $k=1,\ldots,4$ each participant j is more likely to belong to, plus the individual error rate ε_k^j of the corresponding rule.

4.2 Estimation Results

We are interested in how many subjects consistently follow each (noisy) behavioral rule in this paradigm, and in how precise these rules are (i.e. how large the ε_k are). To answer this question, we estimate the finite mixture model described above separately for each subject. This type of individual-level analysis has been shown to be a useful tool in a number of other contexts (e.g., El-Gamal Grether, 1995; Costa-Gomes et al., 2001; Iriberri Rey-Biel, 2013; Bruhin et al., 2018).[8]

Following a common approach in the relevant statistics literature (e.g., Redner Walker, 1984), we adopt a maximum likelihood (ML) approach for parameter estimation in our finite mixture model. Specifically, we rely on the Davidon-Fletcher-Powell (DFP) algorithm (Fletcher Powell, 1963; Davidon, 1991).

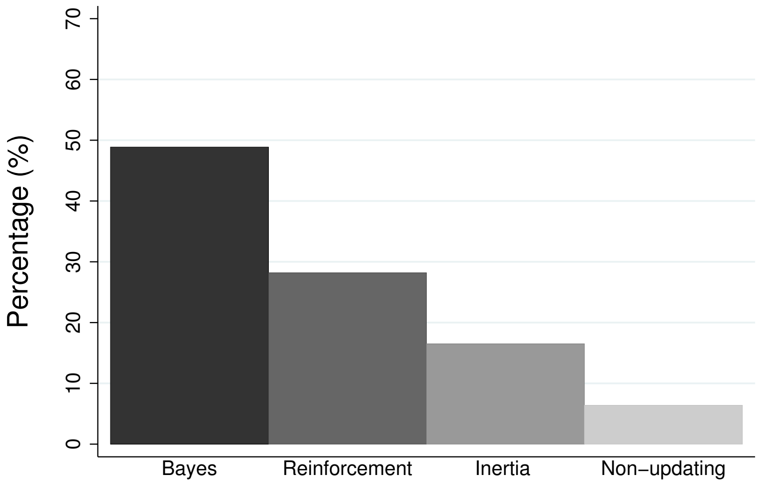
Table 1 summarizes the distribution of subjects classified as following each behavioral rule. ML estimation converged for $N=266$ of our 268 subjects. The first two columns of the table ("Est.Weight" and "Est.Error rate") report the average estimated parameters corresponding to the probability η_k^j of each rule and the relative error rate ε_k^j. "Classified as" and "%" report the number and percentage of subjects classified as most likely following each rule, that is, those subjects for which that rule had the highest probability. This classification is depicted in Figure 3, which shows that a slight majority of our subjects are classified as non-Bayesian, but Bayes' rule (at 49%) is the most frequently-followed

[8] Appendix A (see Section 4.5) reports an additional analysis at the aggregate level.

Table 1: Estimation summary, finite mixture model. Std.Dev. in parenthesis.

Behavioral rule	Est.Weight	Est.Error rate	Classified as	%	Cond.Mean	Cond.Error rate
Bayes	48.15	13.62	130	48.87	97.09	12.07
	(48.71)	(13.34)			(10.34)	(11.84)
Reinforcement	28.28	24.52	75	28.20	98.09	26.06
	(16.14)	(15.51)			(7.81)	(16.14)
Inertia	15.48	41.62	44	16.54	84.74	41.61
	(32.55)	(11.63)			(19.16)	(11.26)
Non-updating	8.09	47.45	17	6.39	88.35	35.16
	(23.26)	(23.32)			(16.72)	(19.29)
N =266						

Figure 3: Proportions of individuals classified into each behavioral rule.



rule, a result in alignment with previous observations by, e.g., El-Gamal Grether (1995). Confirming our basic hypothesis that reinforcement is the main determinant of deviations from Bayes' rule in our setting, this rule is the second highest in terms of number of participants mostly following it: around 28%. In contrast, inertia and non-updating describe just around 17% and 6% of our sample, respectively. This is in sharp contrast with the naïve classification reported above.

A good classification should be as unambiguous as possible. Ideally, to declare a decision maker as, say, a Bayesian or a reinforcer, one would like to have a probability η_k^j close to 1 for the corresponding rule. The columns "Cond.Mean" and "Cond.Error rate" of Table 1 report the estimated average parameters restricted to those subjects who are actually classified as most likely following the corresponding rule. The column "Cond.Mean" shows that, for our estimation, most estimated values of η_k^j are close to either 0 or 1 (the distributions are bimodal). In other words, almost all subjects are unambiguously classified as following one of the four behavioral rules, which is an indication of the goodness of the estimation. However, our assumption of noisy behavioral rules seems justified. In particular, the estimated error rates (conditional or not) differ across rules and are smallest for the two

Table 2: Treatment effect at the individual level.

	Proportion				Est. Error Rate			
			Test of Prop.				MWW test	
Rule	Low	High	z	p-value	Low	High	z	p-value
Bayes	45.32	52.76	-1.212	0.113	13.05	11.14	0.553	0.580
Reinforcement	30.94	25.20	1.042	0.149	25.21	27.20	-0.397	0.692
Inertia	16.55	16.53	0.004	0.997	40.31	43.04	-0.200	0.841
Non-updating	7.19	5.51	0.563	0.574	35.77	34.28	0.098	0.922
N =266		N_{low}=126				N_{high}=140		

most-common ones. Specifically, Bayes' rule has the lowest estimated error rate, around 12%, followed by reinforcement at around 26%. In contrast, the two less-frequent rules, inertia and non-updating, appear to be associated with relatively large error rates.[9]

We can further explore possible correlations between the classification and demographic characteristics. Females are not more likely to be classified as following Bayes's rule or reinforcement than males (MWW tests: Bayes, $N = 266$, $z = 0.348$, $p = 0.728$; Reinforcement, $N = 266$, $z - 0.688$, $p = 0.491$). Moreover, there is no significant correlation between age and being classified into these types (Spearman's correlation: Bayes, $N = 266$, $\rho = -0.018$, $p = 0.766$; Reinforcement, $N = 266$, $\rho = -0.016$, $p = 0.798$).[10]

The bottom line of the classification is that over three quarters of our sample seem to have mostly (and consistently) followed either Bayes' rule or reinforcement during the entire experiment, once one accounts for occasional errors. That is, Bayes' rule explains part of the data, but reinforcement is an important driver of behavior and, in our classification, represents the most common deviation from Bayesian behavior. Our results further highlight that analyses of aggregate behavior can be potentially improved by considering the underlying heterogeneity in decision-making processes across individuals.

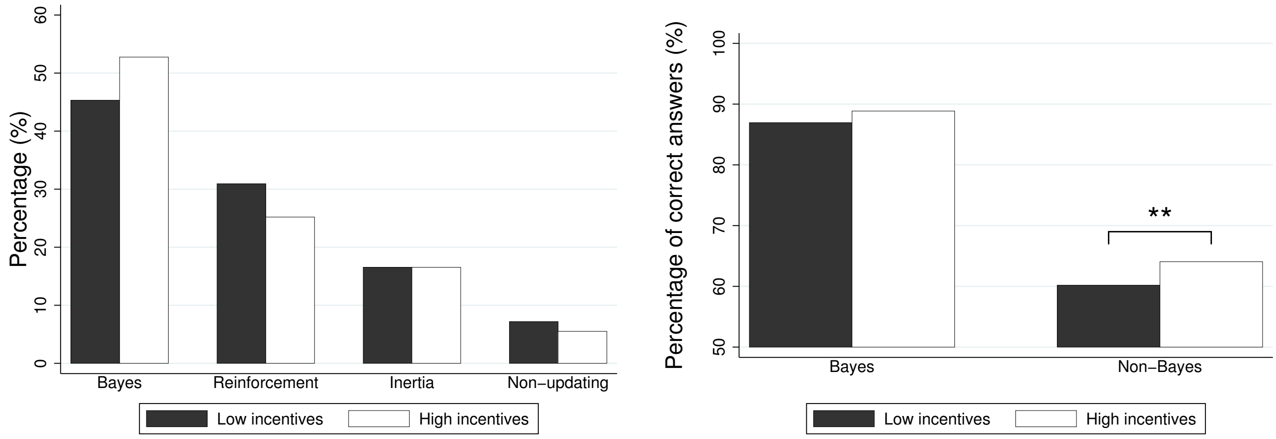
4.3 Behavioral Rules and Incentives

Participants in our study performed the belief-updating task for different incentive levels, with $N = 128$ participating under low incentives and $N = 140$ under high incentives. Since our estimation was conducted at the individual level and incentive levels were varied between subjects (and not within), it is possible to further differentiate the classification according to the level of incentives. Table 2 reports the proportion of individuals in each treatment classified as mostly following each behavioral rule (left-hand side), which are depicted in the left-hand panel of Figure 4. The right-hand side of Table 2 displays the average error rates for each behavioral rule, restricted to the individuals classified as mostly following it, and includes MWW tests for those error rates across treatments. As was the case for the naïve classification, there are no treatment effects for the percentage of individuals classified under each rule. Moreover, the (conditional) error rates for the different behavioral rules are not significantly different across treatments.

[9] Appendices C and D (see Section 4.5) report on the estimation of an FMM with only three (noisy) rules, and of FMMs at the aggregate level with varying numbers of mistake-free rules.

[10] Additionally, there is no correlation between the self-reported difficulty of the task and the classification (measured according to a Likert scale $[0, 10]$ with 10 being "very difficult;" mean 2.78, median 2; Spearman's correlation: Bayes, $N = 266$, $\rho = -0.011$, $p = 0.860$; Reinforcement, $N = 266$, $\rho = 0.017$, $p = 0.777$).

Figure 4: Left-hand panel: Individuals classified into each behavioral rule by treatment. Right-hand panel: Percentage of correct answers for subjects classified as following Bayes' rule and other behavioral rules by incentive treatment. Stars indicate MWW tests, $**p < .05$.

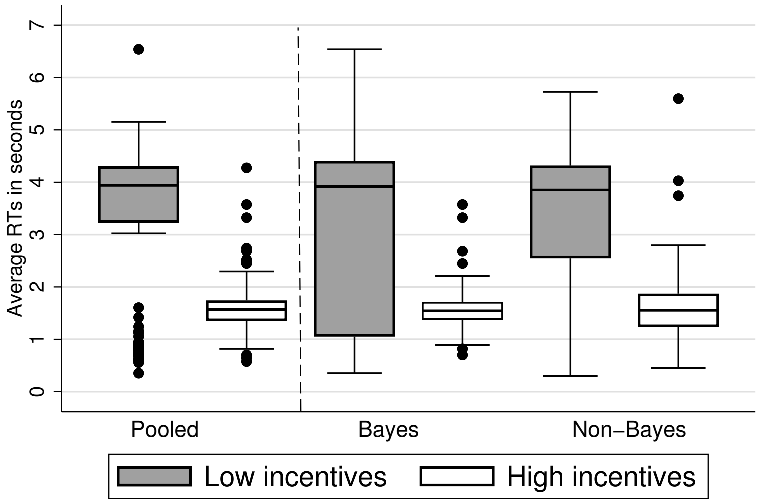


For participants classified as mostly following Bayes' rule, by definition, the error rates (ε_1^j) refer to deviations from normatively-optimal behavior. Hence, the first test on the right-hand side of Table 2 shows that Bayesian decision makers are not significantly better (in a normative sense) for higher incentives (88.86% correct answers under high incentives vs. 86.95% under low incentives; MWW, N=130, z=0.553, p=0.580). However, *non-Bayesian* decision makers do become significantly better in this sense. Specifically, as illustrated in the right-hand panel of Figure 4, if we pool all subjects classified as following either reinforcement, inertia, or non-updating into a non-Bayesian category, we find a significant increase in the percentage of normatively correct answers under high incentives (64.06%) compared to low incentives (60.19%; MWW, N=136, z=1.969, p=0.049).[11]

This result has a simple interpretation and speaks in favor of a heterogeneous effect of incentives. For part of the population (roughly half), low incentives suffice to engage in cognitive deliberation and decide following noisy, error-prone approximations of Bayes' rule. Higher incentives do not result in additional improvements for these individuals, and in particular error rates cannot be reduced, suggesting ceiling effects. Another large part of the population, however, follows non-Bayesian behavioral rules most of the time (mostly reinforcement) and accordingly achieves a lower performance. Taken as a whole, the latter subjects, however, do react to incentives and make more choices in alignment with Bayes' rule under high incentives.

The increase in Bayesian *choices* under high incentives for subjects classified as non-Bayesians, however, is not of sufficient magnitude to be reflected in a change in the classification, or even in a significant increase of the fraction of the time that those subjects are classified as relying on Bayes' rule (η_1^j), mainly because the latter is very close to zero. Indeed, the estimated average weight for Bayes' rule for non-Bayesian subjects is merely 0.74% (min= 0, max= 38.56%) under low incentives and 1.89% (min= 0, max= 44.14%) under high incentives (MWW, N=136, $z - 0.132$, $p = 0.895$).

Figure 5: Distribution of response times by treatment, pooled (left) and separately for subjects classified as Bayesians and non-Bayesians (right).



4.4 Response Times and Incentives

In our experiment, we recorded both choices and their response times. The analysis of the response times reveals a very large, highly significant difference between incentive treatments. As illustrated in Figure 5 (left), decisions are considerably *faster* under high incentives (avg. 1.627 seconds) compared to low ones (3.402 seconds; MWW, N=268, z=7.868, $p < .001$), with an average decrease of about 52%. This difference is independent of the classification. In particular, it holds both for subjects classified as most likely following Bayes' rule (Figure 5, right; high incentives, 1.592 s; low incentives, 3.183 s; MWW, N=130, z=3.957, $p < .001$) and for all those classified as non-Bayesian (high incentives, 1.594 s; low incentives 3.335 s, MWW, N=136, z=7.758, $p < .001$).

This effect demonstrates that participants did react to the different levels of incentives, i.e. there were differences in the perception of high vs. low incentives for our task. However, the interpretation of response times is nuanced, especially when different behavioral rules codetermine behavior (e.g., Achtziger Alós-Ferrer, 2014; Spiliopoulos Ortmann, 2018). Some authors have suggested an interpretation of response times as a proxy for cognitive effort in certain settings (Moffatt, 2005; Enke Zimmermann, 2019), e.g. iterative thinking in games (Alós-Ferrer Buckenmaier, 2020). However, this interpretation would be unwarranted in the simple decisions we study here.

On the contrary, we suggest that the interpretation of the effect of incentives on response times in our context is straightforward. One of the most stable and most firmly-established regularities involving response times is the *chronometric effect*, which boils down to the extremely robust observation that easier choice problems (where alternatives' evaluations show large differences) take less time to respond to than harder problems (e.g., Dashiell, 1937; Moyer Landauer, 1967). Hence, deliberation times are longer for alternatives that are more similar, either in terms of preference or along a prede-

[11] A similar result is obtained by comparing the estimated error rates for Bayes' rule, ε_1^j, for non-Bayesian subjects across incentive levels. The average estimated error rate is 16.48% for low incentives and 13.38% for high incentives (MWW, N=136, z=1.849, p=0.064).

fined scale. This effect has explicitly been shown in various economic settings such as intertemporal choice (Chabris et al., 2009), decisions under risk (Alós-Ferrer Garagnani, 2018), consumer choice (Krajbich et al., 2010), and dictator and ultimatum games (Krajbich et al., 2015).[12] In our case, increasing incentives makes the payoff difference between a correct and an incorrect choice larger. Hence, higher incentives result in a larger difference in payoffs, which, by the chronometric effect, result in shorter response times. This observation is enough to explain the effect we find.

4.5 Alternative Classifications

Our classification through an FMM follows the supervised approach, because our experiment was designed to disentangle four specific behavioral rules. It specifies noisy behavioral rules where error rates can be estimated, because those provide a measure of the accuracy of the classification. We conduct the estimation at the individual level, where the FMM is estimated separately for every participant, since we are especially interested in heterogeneity across and within individuals. Unfortunately, all three choices make the analysis computationally intensive.

The Online Appendices report on a number of additional, alternative classifications. First, we implemented an unsupervised approach, which is detailed in Appendix A. In this approach, we compared models with 2, 3, or 4 components without specifying their characteristics. The results show that the unsupervised approach is not appropriate for our data, because it loses track of the individual and essentially tries to fit the data through rules that always bet on black or always bet on white.[13]

Second, we estimated an FMM at the aggregate level, i.e. as if the entire dataset came from a representative decision maker. This analysis (assuming noisy behavioral rules as in our main analysis) is detailed in Appendix B. As in the model described above, this estimation puts the largest weights on Bayesian behavior and model-free reinforcement, with a sizeable weight on inertia (18.5%). However, the error rates are very high for all rules (close to 50%), which turns them into essentially random behavior. Again, this suggests that this kind of analysis is inappropriate for our data. The reason is that the aggregate-level analysis treats all observations as equivalent and effectively ignores heterogeneity across individuals.

Third, we conducted a number of model comparisons estimating the model at the aggregate level assuming mistake-free rules and varying the number of considered rules (2, 3, or 4; see Appendix C). The best model (according to criteria which penalize additional parameters) consists of only Bayes' rule and (model-free) Reinforcement. However, it is difficult to evaluate the fit of the classification in the absence of estimated error rates.

Fourth, as an additional model comparison, we estimated an FMM with noisy rules and at the individual level but only three types: Bayes' rule, reinforcement, and a pure-noise rule that simply randomizes between both alternatives. This estimation was very inaccurate, with conditional means between 55% and 61% (see Appendix D for details).

[12] Alós-Ferrer et al. (2020) relies on the chronometric effect to obtain results on preference revelation in the absence of assumptions on underlying utility noise.

[13] We further applied the unsupervised approach to a simulated dataset where all subjects switch uniformly across our four rules (the dataset is described in Appendix E). Again, the results fit it to rules which always bet on the same color.

5 Temporal Dynamics and Hidden Markov Modeling

The finite mixture approach detailed in the previous section has shown that there is substantial heterogeneity across individuals, with half of our sample mostly following error-prone approximations of Bayes' rule, and the other half mostly following non-Bayesian rules. So far, however, we have ignored the temporal dimension. In particular, the previous analysis does not rule out possible dynamic relations across behavioral rules.[14] For example, so far we do not know whether a Bayesian choice might be more or less likely after having relied on the reinforcement rule. To answer these questions, this section investigates temporal effects by relying on a simple machine learning methodology: Hidden Markov modeling. While the previous section concentrated on identifying the behavioral rules that subjects most likely followed during the entire experiment (interindividual heterogeneity), in this section we focus on possible dynamic patterns where decision makers may switch from one behavioral rule to another over time (intraindividual heterogeneity).

5.1 Hidden Markov Models

To study the temporal interplay between behavioral rules, we need to classify each choice for each subject and trial, based on the rule the subject most likely followed *in that trial*, while conditioning on the most likely behavior in previous rounds of the experiment. With this aim, we propose a different model specification which goes beyond the finite mixture modelling approach used so far.

We rely on a simple machine learning algorithm known as *hidden Markov modelling* (HMM) (MacDonald Zucchini, 1997; Frühwirth-Schnatter et al., 2019; Ramaprasad Shigeyuki, 2004). The basic idea is as follows. Since we aim to investigate the dynamic dependence across different behavioral rules, we cannot assume that the latent variables are independently and identically distributed, as in finite mixture models. Rather, a hidden Markov model postulates that the latent variables are connected through a Markov chain. Specifically, in our setting, each state of the Markov chain represents a behavioral rule. Since we cannot directly observe which rule a participant actually uses in a given trial, these states are "hidden." Thus, instead of estimating static probabilities for each state, one needs to estimate transition probabilities across them.

Specifically, let the state of subject j's Markov Chain at time t be described by $X_t^j \in \{1, 2, 3, 4\}$, where $X_t^j = k$ means that, at time t, the subject follows the noisy behavioral rule p_k given by (1). That is, as in Section 4, we maintain our assumption that behavioral rules are noisy and mistakes are possible. In particular, a state does not correspond to a deterministic behavioral rule, but rather to its noisy counterpart, and thus we need to estimate the relative error probability for each state (which, in terms of HMM models, is the complementary of the *emission probability*).

For our data, Hidden Markov Modeling reduces to the estimation of the individual-level, state-dependent error rates ε_k^j, $k = 1, 2, 3, 4$, $j = 1, \dots, 268$, and the stationary transition probabilities

$$P^j(k, \ell) = \text{Prob}(X_{t+1}^j = \ell | X_t^j = k),$$

[14]Subjects classified as mostly Bayesian in the FMM made fewer correct choices (55.03%) for four-ball decisions, where reinforcement prescribes an error, than for six-ball decisions, where reinforcement points to the correct response (65.21%, WSR test $N = 130$, $z = -2.155$, $p = .031$). This would be compatible with those subjects relying on reinforcement as an alternative rule at least part of the time. The FMM classification, however, does not reflect this observation, in view of the large conditional means in Table 1.

Table 3: Temporal dynamics, individual averages with Std.Dev. in parentheses.

N=268	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	42.75%	34.73%	14.63%	7.88%	12.74%
	(22.53)	(17.70)	(21.60)	(13.02)	(13.02)
RL	42.63%	34.86%	14.56%	7.96%	22.85%
	(22.56)	(17.66)	(21.80)	(12.90)	(15.35)
Inertia	42.33%	35.52%	14.62%	7.53%	39.21%
	(22.52)	(18.70)	(21.69)	(12.54)	(11.64)
Non-up	42.64%	34.82%	14.88%	7.66%	42.93%
	(24.91)	(20.89)	(22.09)	(22.09)	(23.20)
Invariant:	42.64%	34.90%	14.62%	7.84%	

which form a 4×4 matrix P^j for each subject j. Given a deterministic initial condition X_0^j, or a probabilistic one $\left(\text{Prob}(X_0^j = k)\right)_{k=1}^4$, this transition probability matrix fully determines the probabilities of the events $X_t^j = k$ for all t and k by

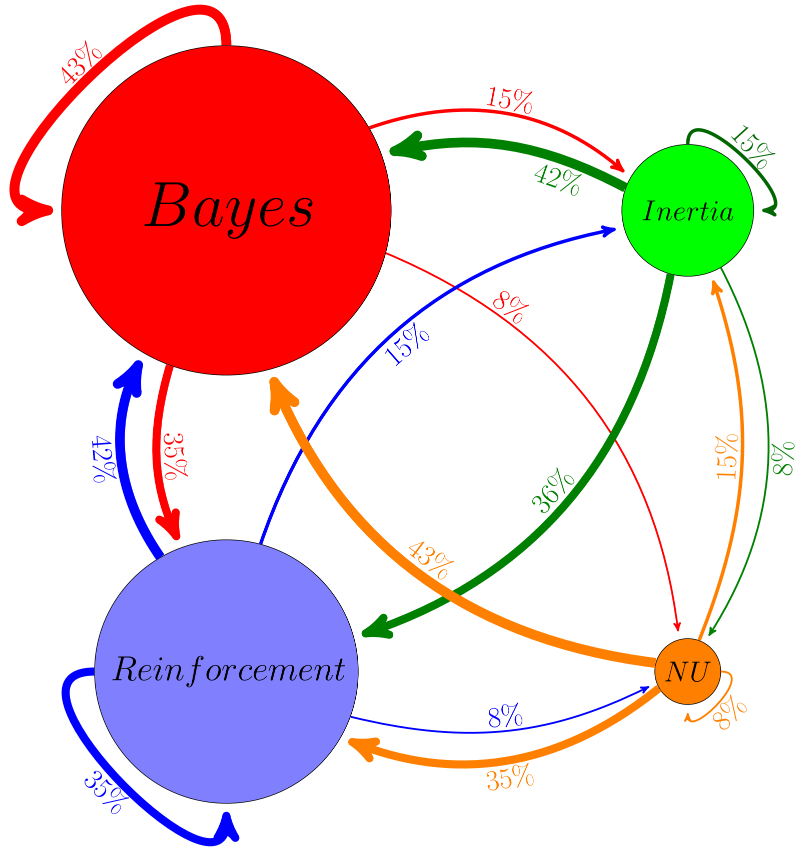
$$\left(\text{Prob}(X_t^j = k)\right)_{k=1}^4 = \left(\text{Prob}(X_0^j = k)\right)_{k=1}^4 \cdot \left[P^j\right]^t.$$

Following the HMM literature (Rabiner, 1990; Durbin et al., 1998), we adopt a standard approach to estimate the parameters of the HMM, namely a special case of expectation maximization (Dempster et al., 1977) known as the Baum-Welch algorithm (Baum et al., 1970). This algorithm is also known as the forward-backward (or α, β, γ) algorithm because it loops among three well-differentiated parts or tasks. Intuitively, the first task is to efficiently evaluate the probability of the observations given the model (following a top-down procedure, in our case from first to last trial). The second task is to identify the optimal state sequence, that is, which state sequence among all theoretically-possible ones is most likely to have occurred (now following a bottom-up or last-to-first procedure). The third task is to adjust the model's parameters given the most-likely sequence and the estimated probability of observations given the model, in a meaningful way. We refer the reader to Frühwirth-Schnatter (2006) or Frühwirth-Schnatter et al. (2019) for details on the algorithm. We declare that the algorithm has converged when the change in log likelihood, normalized by the total number of possible sequences and by sequence length, is smaller than 10^{-7} as standard in the literature (Ramaprasad Shigeyuki, 2004; Mamon Elliott, 2014). To determine the initial values for the algorithm, we used the values (η_k^j and ε_k^j) obtained in the FMM of Section 4.2 (and a flat prior for the 2 subjects for whom that estimation did not converge).

5.2 Temporal Dynamics: Estimation Results

We estimated the parameters of the HMM for all 268 participants, thereby obtaining a transition probability matrix and a vector of error terms for the hidden states for each individual. From these matrices, we computed all individual invariant distributions (μ such that $\mu = \mu \cdot P$) which, by the Fundamental Theorem of Markov Chains, summarize the long-run probabilities of the states, and,

Figure 6: Graphical depiction of the results of the HMM estimation. 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.



by the Ergodic Theorem, reflect the average time that the system spends at each state (e.g., Karlin Taylor, 1975).[15]

Table 3 displays 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. That is, the given matrices represent average transition probability matrices. The last column of the table ("Error") indicates the estimated probability of making an error (the complementary of the emission probability) at each state. The last row of the table ("Invariant") reports the probabilities in the corresponding invariant distribution.

Even at this level of aggregation, the results of the temporal dynamics analysis suggest an oscillatory pattern in behavior. This is made clear by Figure 6. In this and the following pictures, circled areas corresponds to the states (rules of behavior), and arrows to transitions. The areas of the circles and the thickness of the arrows are proportional to the weights in the corresponding invariant distri-

[15]Appendix E reports a parameter recovery exercise, where the data was simulated with a known generating process.

Table 4: Temporal dynamics by treatment, individual averages with Std.Dev. in parentheses.

Low Incentives

N=128	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	44.10%	31.68%	15.12%	9.10%	11.36%
	(24.73)	(16.58)	(22.55)	(14.99)	(12.39)
RL	44.11%	31.62%	14.91%	9.37%	23.30%
	(24.93)	(16.62)	(22.47)	(15.17)	(15.76)
Inertia	43.13%	32.79%	15.25%	8.83%	39.45%
	(24.41)	(17.36)	(22.67)	(14.72)	(13.11)
Non-up	44.28%	31.26%	15.49%	8.97%	43.71%
	(26.60)	(18.57)	(23.25)	(15.46)	(24.15)
Invariant:	43.97%	31.79%	15.11%	9.13%	

High incentives

N=140	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	41.29%	38.07%	14.09%	6.56%	14.01%
	(19.84)	(18.33)	(20.54)	(10.34)	(13.49)
RL	41.00%	38.40%	14.17%	6.42%	22.44%
	(19.60)	(18.13)	(21.13)	(9.66)	(15.01)
Inertia	41.45%	38.49%	13.93%	6.12%	38.98%
	(20.32)	(19.70)	(20.64)	(9.46)	(10.15)
Non-up	40.85%	38.72%	14.22%	6.22%	42.21%
	(22.89)	(21.81)	(20.81)	(9.83)	(22.36)
Invariant	41.17%	38.30%	14.11%	6.42%	

butions and to the transition probabilities, respectively. Participants gravitate mostly toward Bayes' rules and reinforcement and continue using those rules in most cases, but transition probabilities across rules are typically large, very especially between Bayes' rule and reinforcement. We observe that, at the aggregate level, the rows of the transition probability matrix are similar to each other, indicating that the average dynamics might be close to an i.i.d. process. This, however, is not necessarily true at the individual level. Individual transition probability matrices display considerable variance, and are often quite far from reflecting i.i.d. processes. Appendix F provides a more detailed examination of this point, including a measure of heterogeneity in transition probability matrices (Figure F.1) and several examples at the individual level (Table F.1).

5.3 Temporal Dynamics and Incentives

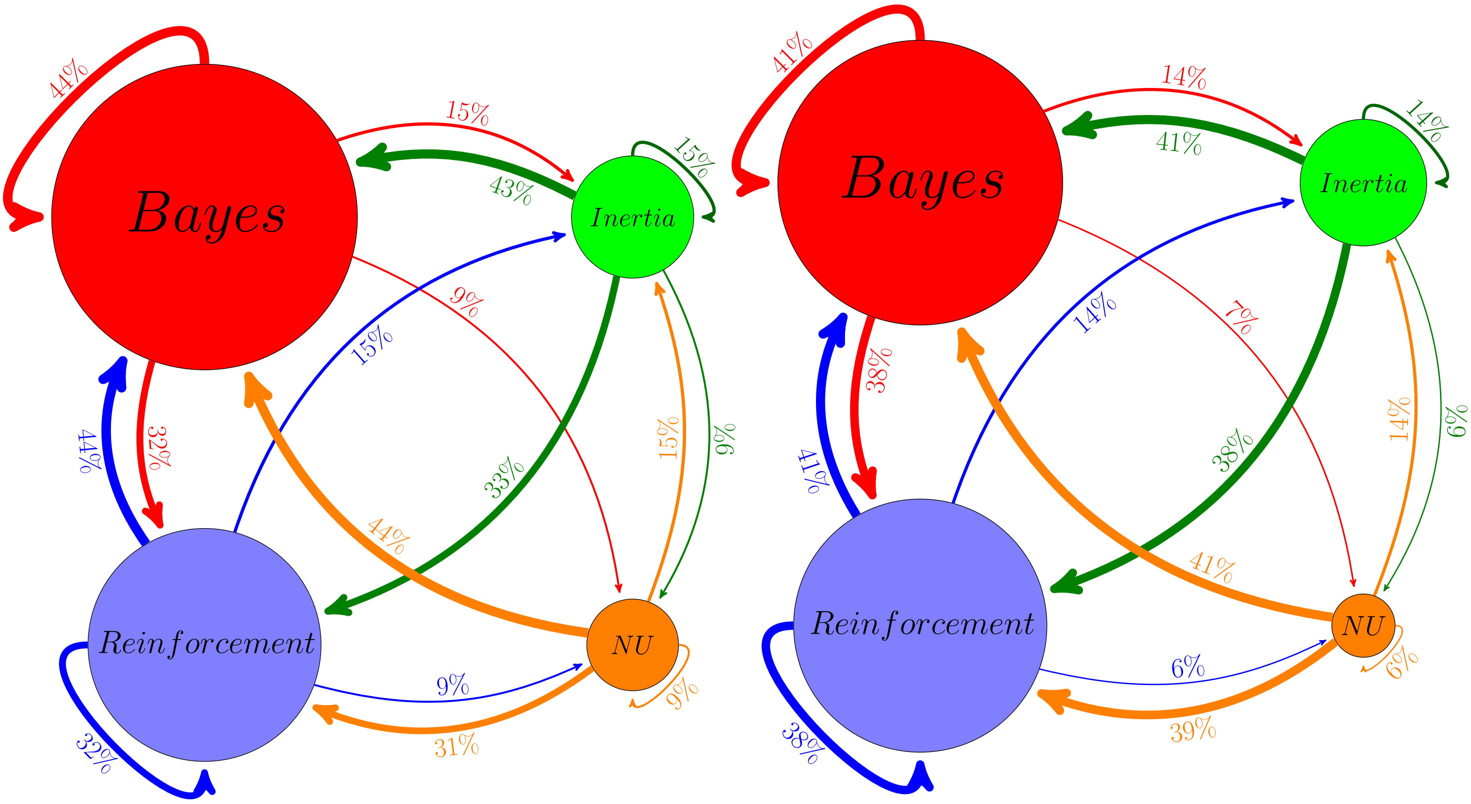
Table 4 displays the average results of the HMM estimation (transition probability matrices, state-dependent error rates, and invariant distributions) separately by treatment (low vs. high incentives), following the conventions in Table 3. Figure 7 depicts the results, again separately by treatment (conventions are as in Figure 6). The overall pattern is confirmed, with participants mostly relying on Bayes' rules and reinforcement but displaying relatively large transition probabilities. However, the effects of incentives are now more clear.

Specifically, comparing the two conditions, we observe different dynamics among behavioral rules. The most striking result is that increasing incentives shifts behavior *toward reinforcement and away from Bayes' rule.* After using Bayes' rule, under high incentives (compared to low incentives), it is less likely to stay with it and use it again (Low 44.10%, High 41.29%; MWW, N=268, z=2.959, p=.003) and more likely to switch away to reinforcement (Low 31.68%, High 38.07%; MWW, N=268, z=-4.329, $p < .0001$). In contrast, after using reinforcement, under high incentives it is more likely to stay with this rule (Low 31.62%, High 38.40%; MWW, N=268, z=-4.749, $p < .0001$) and less likely to switch away to Bayes' rule (Low 44.11%, High 41.00%; MWW, N=268, z=2.927, $p = .003$).[16]

In other words, we observe a significant treatment effect in the temporal dynamics in the form of a clear shift toward reinforcement and away from Bayes' rule with larger incentives. This effect can only be observed clearly by examining the temporal dynamics. Comparing individual invariant distributions, we observe that, in accordance with the results above, the overall probability of relying on reinforcement is larger under high incentives (Low 31.79%, High 38.30%; MWW, N=268, z=-4.756,

[16] Increasing incentives decreases the probability of staying with non-updating (Low 8.97%, High 6.22%; MWW, N=268, z=2.141, p =.0323) but not with inertia (Low 15.25%, High 13.93%; MWW, N=268, z=0.801, p=.423).

Figure 7: Graphical depiction of the results of the HMM estimation, by treatment. 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.



(a) Low incentives

(b) High incentives

$p < .0001$). The long-run probability of relying on Bayes' rule is smaller under high incentives, but the difference is not significant (Low 43.97%, High 41.17%; MWW, N=268, z=1.018, $p = .310$).

These results have a twofold interpretation. On the one hand, they speak in favor of a ceiling effect of incentives where a higher reward for each correct choice does not decrease the error rates and does not increase the overall reliance on Bayes' rule. On the other hand, higher incentives seem to increase the appeal of simple reinforcement. The latter is compatible with the electroencephalography (EEG) study by Achtziger et al. (2015), which pointed out that increasing incentives increases the salience of the received feedback and makes simple reinforcement processes more prominent. This is simply because if monetary incentives are increased, the win/loss cues which lead to an activation of reinforcement processes become more salient, and hence reliance on reinforcement increases.

We remark that these results can only be obtained once we rely on the HMM analysis. The previous FMM analysis neglects the temporal dynamics, and it is only by allowing for the latter that a more clear picture of the effects of incentives on actual behavior starts to emerge. We will further examine heterogeneity in the reaction to incentives in this context in Section 5.5 below.

5.4 A Dynamic Classification

Our FMM analysis in Section 4.2 allowed us to classify participants according to the rule they mostly followed under an i.i.d. assumption (Figure 3). Since our HMM analysis delivers the individual invari-

Table 5: Estimation summary, HMM conditional on subjects classified as mostly following the relative behavioral rule. Std.Dev. in parenthesis.

Behavioral rule	Classified as	%	Cond.Mean	Cond.Error rate
Bayes	130	48.51	64.78	10.07
			(5.51)	(11.39)
Reinforcement	75	27.99	59.28	24.33
			(5.25)	(16.17)
Inertia	44	16.42	62.52	39.18
			(3.31)	(10.98)
Non-updating	19	7.09	51.44	35.16
			(10.14)	(18.51)
N =268				

ant distributions, which summarize the long-run proportion of time spent in each state, i.e. using each rule, we can derive an alternative classification from those invariant distributions dispensing with the i.i.d. assumption. Specifically, we classify individuals on the basis of the state displaying the largest probability in their respective individual invariant distributions. That is, similarly to Subsection 4.2, a subject is considered Bayesian if Bayes' rule is given the largest probability in the invariant distribution derived from her individual transition probability matrix.

The result of this classification turns out to be empirically identical to the one obtained in Section 4.2 and depicted in Figure 3. That is, although individual weights for the hidden states are quite different from the rule weights obtained in the FMM model, every single individual retains the same classification in the HMM as in the FMM case, except for the two unclassified subjects (who are re-classified as non-updaters). That is, the previous classification according to the most-used behavioral rules turns out to be stable. Although the objective of the HMM analysis is to examine the temporal dynamics and not this derived classification, we view the fact that the latter agrees with the ones derived in the previous section as a validation of the approach.

Table 5 summarizes the results of the HMM classification. Although the classification in terms of which rule is followed most of the time is almost identical to the one derived from the FMM, the conditional means are quite different. In the FMM case, the conditional means were above 97% for Bayesians and reinforcers, and above 84% for non-updaters and subjects relying mostly on inertia. In contrast, in the HMM classification conditional means have a different interpretation, since they are derived from invariant distributions. As can be seen in the third column of Table 5, in this case the conditional means are between 51% and 65%. That is, when one neglects the possibility of dynamic dependence among behavioral rules, the FMM classification reliably identifies the most-used rule for each individual, but does not offer a good explanation for deviations from that rule at the individual level. In contrast, taking into account the possible temporal dynamics, the HMM classification uncovers heterogeneity within each individual, in the sense that subjects who mostly follow a behavioral rule actually also rely on different rules a significant proportion of the time, as reflected by the weights in the invariant distribution. Thus, the conditional means in the HMM case reveal that within-individual heterogeneity in behavioral rules is of a sizeable magnitude.

Table 6: Temporal dynamics distinguishing subjects classified as most likely following Bayes' rule from others; 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%	

Non-Bayesian subjects

N=138	To				
From	Bayes	RL	Inertia	Non-up	Error
Bayes	21.94%	39.05%	25.61%	13.41%	14.38%
	(7.79)	(23.29)	(25.65)	(16.30)	(14.27)
RL	21.79%	39.10%	25.53%	13.58%	23.50%
	(7.33)	(23.12)	(25.98)	(16.05)	(15.75)
Inertia	22.47%	39.20%	25.37%	12.96%	38.98%
	(8.14)	(23.56)	(25.79)	(15.46)	(10.53)
Non-up	21.64%	39.46%	25.92%	12.98%	40.79%
	(8.33)	(23.57)	(25.93)	(16.32)	(22.79)
Invariant	21.98%	39.16%	25.56%	13.30%	

5.5 Heterogeneity in the Effects of Incentives

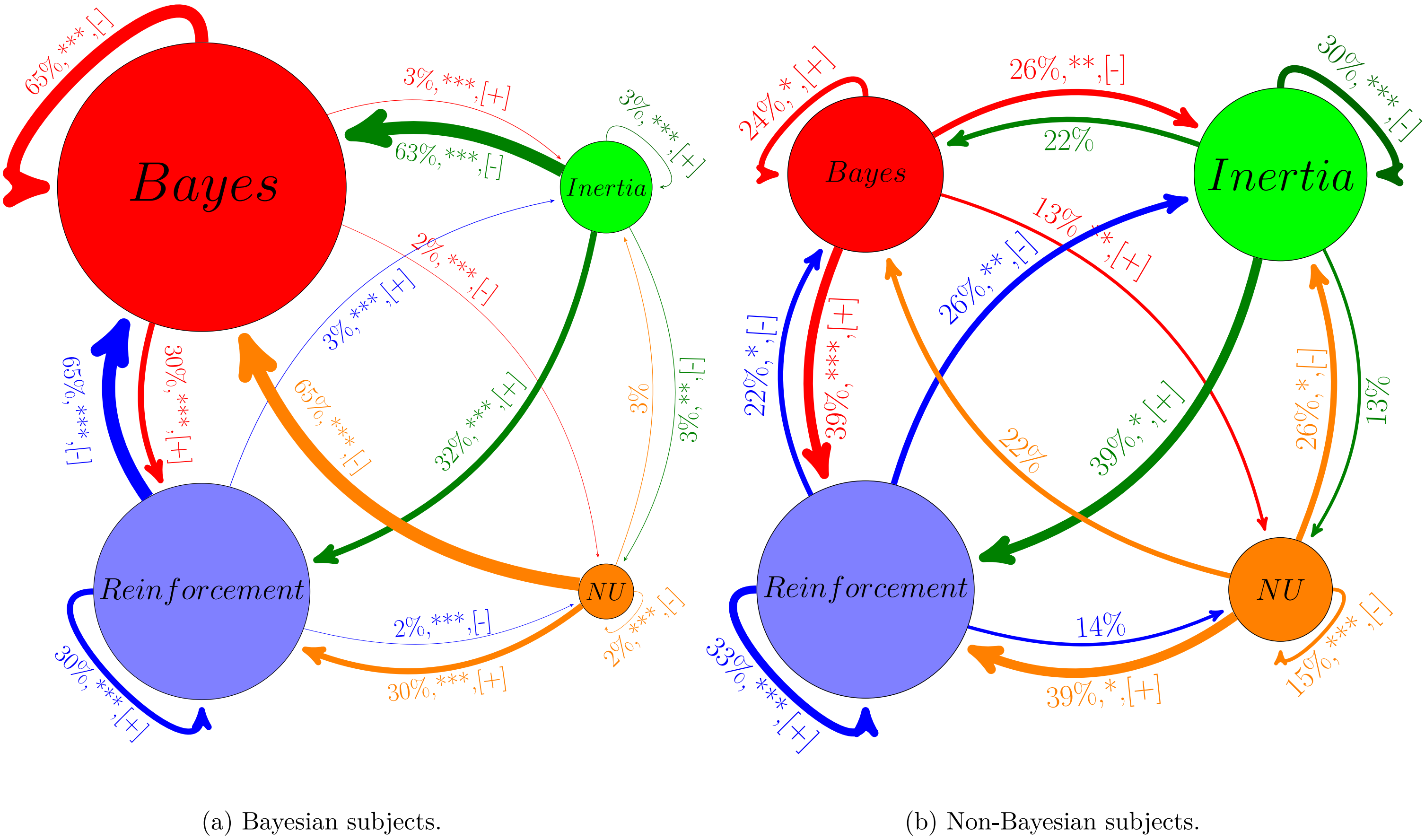
We now proceed to refine the results on the effects of incentives (Section 5.3) by accounting for heterogeneity, relying on the classification obtained above. Table 6 displays the average transition probability matrices conditional on subjects classified as Bayesians (132) or Non-Bayesians (136), respectively. As in Table 4, the right-most columns list the estimated error rates, and the bottom rows detail the (average) invariant distributions. The two panels of Figure 8 give a graphical representation of the results. Each picture is analogous to the one in Figure 7, but restricted to subjects classified as Bayesians or Non-Bayesians, respectively. The significance of tests for each transition probability (high vs. low incentives), as well as the direction of the significant change (+/−), is highlighted.

Increasing incentives has different effects for different subjects. For those relying mostly on Bayes' rule, higher incentives seem to be slightly detrimental, in the sense that the long-run probability of Bayes' rule (in the invariant distribution) is smaller under high incentives (Low 70.24%, High 59.65%; MWW, N=130, $z = 9.833$, $p < .001$). This is mostly due to a decrease in the probability to stay with Bayes' rule after using it (Low 70.17%, High 59.84%; N=130, $z = 9.832$, $p < .001$) and an increase in the probability of transition from Bayes' rule to reinforcement (Low 24.75%, High 35.24%, MWW, N=130, z= -9.832, $p < .001$), in line with the increased appeal of reinforcement under high incentives mentioned above. This is further supported by a similarly-large increase in the probability of staying with the reinforcement rule after using it (Low 24.62%, High 35.74%, MWW, N=130, z= -9.813, $p < .001$), and a reduction in the probability of transition from reinforcement to Bayes' rule (Low 70.50%, High 59.32%, MWW, N=130, z= 9.813, $p < .001$). We remark that the changes in probabilities across incentives are substantial, especially compared to those at the aggregate level.

For subjects classified as following mostly some non-Bayesian rule, the long-run probability of Bayes' rule is slightly larger under high incentives than under low ones (Low 21.13%, High 22.67%; N=138, $z = -2.816$, $p = .005$). This results from a small increase in the probability to stay with Bayes' rule after using it (Low 20.90%, High 22.77%; N=138, z= -1.814, $p = .070$). However, there is a significant increase in the probability of switching from Bayes' rule to reinforcement (Low 37.36%, High 41.17%; N=138, z= -2.596, $p = .009$). Further, in line with the increased appeal of reinforcement under higher incentives, the probability of staying with this rule after using it is larger under high incentives than under low ones (Low 37.33%, High 41.33%, MWW, N=138, z= 2.673, $p = .007$).

Aggregating all subjects classified as non-Bayesians together (138), however, again loses valuable information. Appendix G reports the temporal dynamics (average transition matrices and invariant distributions, Table G.1 and Figure G.1) for all four types. Subjects classified as following a specific

Figure 8: Average probabilities of switching from one behavioral rule to another separating subjects classified as mostly following Bayes' rule from others. Stars indicate MWW tests 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. Numbers are approximated to the nearest integer for graphical illustration.



behavioral rule are in general quite consistent and exhibit large probabilities for returning to that rule after a deviation. However, the effects of incentives are quite different across types, and in particular when comparing reinforcers and (non-Bayesian) non-reinforcers.

The majority of non-Bayesians are classified as reinforcers (75). For those, increasing incentives results in a large increase in the long-run probability of model-free reinforcement (Low 54.89%, High 65.03%; MWW, N=75, $z = -9.834$, $p < .001$) and a sizeable decrease in the probability of using Bayes' rule (Low 30.08%, High 20.09%; MWW, N=75, $z = 7.371, p < .001$). This is accompanied by a large increase in the probability of switching from Bayes' rule to reinforcement (Low 54.70%, High 65.34%; N=75, z= -7.220, $p < .001$) and a large reduction in the probability of transition from reinforcement to Bayes' rule (Low 29.87%, High 20.18%, MWW, N=75, z= 7.370, $p < .001$).

For non-Bayesian subjects classified as mostly following inertia (44) or non-updating (19), results are quite different. Higher incentives lead to an increase in the long-run probability of Bayes' rule (Inertia: Low 14.95%, High 25.11%; MWW, N=44, $z = -5.677$, $p < .001$; Non-updating: Low 9.81%, High 14.86%; MWW, N=19, $z = -3.636$, $p < .001$) accompanied by a decrease in the long-run probability of the own, respective rule (Inertia: Low 65.16%, High 59.72%; MWW, N=44, $z = 5.631$, $p < .001$; Non-updating: Low 59.95%, High 40.07%; MWW, N=19, $z = 3.638$, $p < .001$).[17]

These results clarify the mechanisms linking incentives to performance in our paradigm while taking into account behavioral heterogeneity. The effect of increasing incentives is double-edged. On the one hand, higher incentives do have a positive, presumably motivational effect for some subjects, leading to a higher reliance on Bayes' rule. On the other hand, higher incentives seem to generally increase the reliance on reinforcement, in agreement with a resulting, increased salience of the win-lose cues that activate that behavioral rule (recall Section 5.3).

The overall picture is hence as follows. For subjects classified as mostly Bayesian, low incentives suffice to spark the use of Bayes' rule, presumably because those are sufficient motivation to overcome their cognitive costs. For those, an increase in incentives produces no improvement, suggesting a ceiling effect. The effects might even be slightly detrimental, in line with the interpretation that higher incentives make model-free reinforcement more appealing. For subjects classified as mostly reinforcers, the win-lose monetary cues spark reliance on reinforcement already for low incentives, and an increase in incentives simply enhances those cues and is hence mostly detrimental. For subjects classified neither as Bayesians nor as reinforcers, low incentives do not suffice to trigger the use of Bayes' rule, but increasing incentives creates a significant shift toward it, which is of a large magnitude in relative terms. Although their performance is generally worse than that of Bayesians, these subjects do respond to incentives as standard economic theory would suggest.

Thus, the analysis of the temporal dynamics casts light on the mechanisms undergoing the effects of incentives on performance for belief updating tasks, and specifically shows that those effects are highly heterogeneous. Further, our analysis through Hidden Markov Modeling shows that the underlying heterogeneity is well-captured by the view that reinforcement is the main driver behind deviations from Bayesian updating, with alternative rules playing a comparatively small role.

6 Conclusion

We designed a novel belief-updating experimental paradigm to disentangle alternative rules of behavior when an existing prior can be updated in the face of new information, and that information carries a win-lose feedback, as is often the case for financial and managerial decisions. The analysis shows that, in such cases, model-free reinforcement is the main driver of deviations from Bayesian updating, with alternative rules as decision inertia or non-updating playing a smaller role.

We employ a multi-layered identification strategy. First, by applying finite mixture models, we find large levels of heterogeneity, with around half of the population relying mostly on Bayesian updating, over a quarter relying on reinforcement, and the rest on the remaining rules. At this level of analysis, we find that increasing incentives results in a performance increase for subjects classified as non-Bayesians, while evidence for Bayesians is compatible with ceiling effects.

Second, we use hidden Markov models to examine the temporal dynamics across different behavioral rules. We find considerable heterogeneity *within* individuals, with most of them relying on several, different behavioral rules over time, especially Bayes' rule and model-free reinforcement. Thus, even subjects classified as Bayesians are only "part-time Bayesians." The analysis of these patterns shows

[17] For subjects relying mostly on inertia, there was no significant difference in the probability of reinforcement (Low 10.21%, High 9.03%; MWW, N=44, $z = 0.530$, $p = .604$). For subjects relying mostly on non-updating, higher incentives increased the reliance on reinforcement (Low 25.00%, High 29.78%; MWW, N=19, $z = -3.635$, $p < .001$).

significant effects of incentives with relatively large probability shifts. Subjects who rely mostly on reinforcement increase the frequency of this rule under higher incentives. Subjects who rely mostly on Bayes' rule experience detrimental effects of incentives, due to an analogously-increased frequency of reinforcement. These results might reflect a "reinforcement paradox" where model-free reinforcement is triggered more often when the monetary values attached to the win-lose cues increases. Last, subjects who rely on either inertia or non-updating benefit from increased incentives, as they react by shifting toward Bayes' rule, improving their performance.

Our results go beyond the well-known fact that human decision makers are not Bayesian. On the one hand, decision makers can be fruitfully classified according to the decision rules they mostly rely on. Further, the resulting heterogeneity extends to the effects of financial incentives. While some decision makers experience ceiling effects or even detrimental consequences, others do react positively to incentives. On the other hand, decision makers exhibit non-trivial temporal patterns and rely on different behavioral rules over time. When it comes to belief updating, it is not only that one size does not fit all, but rather, that one size might not even fit one.

References

Achtziger A, Alós-Ferrer C (2014) Fast or Rational? A Response-Times Study of Bayesian Updating. *Management Science* 60(4):923–938.

Achtziger A, Alós-Ferrer C, Hügelschäfer S, Steinhauser M (2015) Higher Incentives Can Impair Performance: Neural Evidence on Reinforcement and Rationality. *Social Cognitive and Affective Neuroscience* 10(11):1477–1483.

Akaishi R, Umeda K, Nagase A, Sakai K (2014) Autonomous Mechanism of Internal Choice Estimate Underlies Decision Inertia. *Neuron* 81(1):195–206.

Alós-Ferrer C, Buckenmaier J (2020) Cognitive Sophistication and Deliberation Times. *Experimental Economics* forthcoming.

Alós-Ferrer C, Fehr E, Netzer N (2020) Time Will Tell: Recovering Preferences when Choices are Noisy. *Journal of Political Economy* forthcoming.

Alós-Ferrer C, Garagnani M (2018) Strength of Preference and Decisions Under Risk, Working Paper, University of Zurich.

Alós-Ferrer C, Hügelschäfer S, Li J (2016) Inertia and Decision Making. *Frontiers in Psychology* 7 (169):1–9.

Alós-Ferrer C, Jaudas A, Ritschel A (2019) Effortful Bayesian Updating: A Pupil-dilation Study, Working Paper, University of Zurich.

Ansari A, Montoya R, Netzer O (2012) Dynamic Learning in Behavioral Games: A Hidden Markov Mixture of Experts Approach. *Quantitative Marketing and Economics* 10(4):475–503.

Asparouhova E, Bossaerts P, Eguia J, Zame W (2015) Asset Pricing and Asymmetric Reasoning. *Journal of Political Economy* 123(1):66–122.

Azrieli Y, Chambers CP, Healy PJ (2018) Incentives in Experiments: A Theoretical Analysis. *Journal of Political Economy* 126(4):1472–1503.

Barberis N, Shleifer A, Vishny R (1998) A Model of Investor Sentiment. *Journal of Financial Economics* 49(3):307–343.

Barron K (2020) Belief Updating: Does the Good-News, Bad-News Asymmetry Extend to Purely Financial Domains? *Experimental Economics* forthcoming.

Baum LE, Petrie T, Soules G, Weiss N (1970) A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics* 41(1):164–171.

Beierholm UR, Anen C, Quartz S, Bossaerts P (2011) Separate Encoding of Model-Based and Model-Free Valuations in the Human Brain. *Neuroimage* 58(3):955–962.

Bellemare C, Kröger S, Van Soest A (2008) Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities. *Econometrica* 76(4):815–839.

Bilias Y, Georgarakos D, Haliassos M (2010) Portfolio Inertia and Stock Market Fluctuations. *Journal of Money, Credit and Banking* 42(4):715–742.

Bruhin A, Fehr E, Schunk D (2018) The Many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences. *Journal of the European Economic Association* 17(4):1025–1069.

Bruhin A, Fehr-Duda H, Epper T (2010) Risk and Rationality: Uncovering Heterogeneity in Probability Distortion. *Econometrica* 78(4):1375–1412.

Camerer CF (1987) Do Biases in Probability Judgment Matter in Markets? Experimental Evidence. *American Economic Review* 77(5):981–997.

Chabris CF, Morris CL, Taubinsky D, Laibson D, Schuldt JP (2009) The Allocation of Time in Decision-Making. *Journal of the European Economic Association* 7(2-3):628–637.

Charness G, Levin D (2005) When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect. *American Economic Review* 95(4):1300–1309.

Costa-Gomes MA, Crawford VP, Broseta B (2001) Cognition and Behavior in Normal-Form Games: An Experimental Study. *Econometrica* 69(5):1193–1235.

Dashiell JF (1937) Affective Value-Distances as a Determinant of Aesthetic Judgment-Times. *American Journal of Psychology* 50:57–67.

Davidon WC (1991) Variable Metric Method for Minimization. *SIAM Journal on Optimization* 1(1):1–17.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69(6):1204–1215.

Daw ND, Niv Y, Dayan P (2005) Uncertainty-Based Competition between Prefrontal and Dorsolateral Striatal Systems for Behavioral Control. *Nature Neuroscience* 8(12):1704–1711.

Dempster AP, Laird NM, Rubin DB (1977) Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society Series B* 39(1):1–38.

Doll BB, Simon DA, Daw ND (2012) The Ubiquity of Model-Based Reinforcement Learning. *Current Opinion in Neurobiology* 22(6):1075–1081.

Durbin R, Eddy SR, Krogh A, Mitchison G (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, volume 5 (Cambridge, UK: Cambridge University Press).

Edwards W (1968) Conservatism in Human Information Processing. Kleinmuntz B, ed., *Formal Representation of Human Judgment*, 17–52 (New York, NY: Wiley).

El-Gamal MA, Grether DM (1995) Are People Bayesian? Uncovering Behavioral Strategies. *Journal of the American Statistical Association* 90(432):1137–1145.

Enke B, Zimmermann F (2019) Correlation Neglect in Belief Formation. *Review of Economic Studies* 86(1):313–332.

Erev I, Haruvy E (2016) Learning and the Economics of Small Decisions. Kagel J, Roth AE, eds., *Handbook of Experimental Economics, Volume 2*, 638–716 (Princeton University Press).

Feher da Silva C, Hare TA (2020) Humans Primarily Use Model-Based Inference in the Two-Stage Task. *Nature Human Behaviour* 4(10):1053–1066.

Fischbacher U (2007) z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics* 10(2):171–178.

Fletcher R, Powell MJ (1963) A Rapidly Convergent Descent Method for Minimization. *The Computer Journal* 6(2):163–168.

Frühwirth-Schnatter S (2006) *Finite Mixture and Markov Switching Models* (New York: Springer Series in Statistics).

Frühwirth-Schnatter S, Celeux G, Robert CP (2019) *Handbook of Mixture Analysis* (New York: Chapman and Hall).

Frydman C, Camerer CF (2016) The Psychology and Neuroscience of Financial Decision Making. *Trends in Cognitive Sciences* 20(9):661–675.

Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States Versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning. *Neuron* 66(4):585–595.

Greiner B (2015) Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association* 1:114–125.

Grether DM (1980) Bayes Rule as a Descriptive Model: The Representativeness Heuristic. *Quarterly Journal of Economics* 95:537–557.

Grether DM (1992) Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence. *Journal of Economic Behavior and Organization* 17:31–57.

Griffiths TL, Tenenbaum JB (2006) Optimal Predictions in Everyday Cognition. *Psychological Science* 17(9):767–773.

Iriberri N, Rey-Biel P (2013) Elicited Beliefs and Social Information in Modified Dictator Games: What Do Dictators Believe Other Dictators Do? *Quantitative Economics* 4(3):515–547.

Kahneman D, Tversky A (1972) Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology* 3:430–454.

Karlin S, Taylor HM (1975) *A First Course in Stochastic Processes, 2nd Ed.* (San Diego: Academic Press).

Knutson B, Bossaerts P (2007) Neural Antecedents of Financial Decisions. *Journal of Neuroscience* 27(31):8174–8177.

Krajbich I, Armel C, Rangel A (2010) Visual Fixations and the Computation and Comparison of Value in Simple Choice. *Nature Neuroscience* 13(10):1292–1298.

Krajbich I, Bartling B, Hare T, Fehr E (2015) Rethinking Fast and Slow Based on a Critique of Reaction-Time Reverse Inference. *Nature Communications* 6(7455):1–9.

MacDonald IL, Zucchini W (1997) *Hidden Markov and Other Models for Discrete-Valued Time Series* (Boca Raton, FL: Chapman & Hall/CRC Press).

Mamon RS, Elliott RJ (2014) *Hidden Markov Models in Finance: Further Developments and Applications (Volume II)* (New York: Springer).

Moffatt PG (2005) Stochastic Choice and the Allocation of Cognitive Effort. *Experimental Economics* 8(4):369–388.

Moyer RS, Landauer TK (1967) Time Required for Judgements of Numerical Inequality. *Nature* 215(5109):1519–1520.

Navon D (1978) The Importance of Being Conservative: Some Reflections on Human Bayesian Behaviour. *British Journal of Mathematical and Statistical Psychology* 31:33–48.

Payzan-LeNestour E, Bossaerts P (2015) Learning about Unstable, Publicly Unobservable Payoffs. *The Review of Financial Studies* 28(7):1874–1913.

Pitz GF, Geller ES (1970) Revision of Opinion and Decision Times in an Information-Seeking Task. *Journal of Experimental Psychology* 83(3):400–405.

Rabiner LR (1990) A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Setti G, ed., *Readings in Speech Recognition*, 267–296 (Amsterdam, Netherlands: Elsevier).

Ramaprasad B, Shigeyuki H (2004) *Hidden Markov Models: Applications to Financial Economics* (Berlin, Germany: Springer Science & Business Media).

Redner RA, Walker HF (1984) Mixture Densities, Maximum Likelihood and the EM Algorithm. *SIAM Review* 26(2):195–239.

Ritov I, Baron J (1992) Status-Quo and Omission Biases. *Journal of Risk and Uncertainty* 5(1):49–61.

Schultz W, Dayan P, Montague PR (1997) A Neural Substrate of Prediction and Reward. *Science* 275(5306):1593–1599.

Shachat J, Swarthout JT, Wei L (2015) A Hidden Markov Model for the Detection of Pure and Mixed Strategy Play in Games. *Econometric Theory* 31(4):729–752.

Shachat J, Wei L (2012) Procuring Commodities: First-Price Sealed-Bid or English Auctions? *Marketing Science* 31(2):317–333.

Shiller RJ (2005) *Irrational Exuberance (Second Edition)* (Princeton, Oxford: Princeton University Press).

Shleifer A (2000) *Inefficient Markets: An Introduction to Behavioural Finance* (Oxford, UK: OUP Oxford).

Spiliopoulos L, Ortmann A (2018) The BCD of Response Time Analysis in Experimental Economics. *Experimental Economics* 21(2):383–433.

Sutton RS, Barto AG (1998) *Reinforcement Learning: An Introduction* (Cambridge, MA: MIT Press).

Thaler RH (2005) *Advances in Behavioral Finance*, volume 2 (New Jersey: Princeton University Press).

Thorndike EL (1911) *Animal Intelligence: Experimental Studies* (NY: MacMillan (see also Hafner Publishing Co., 1970)).

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 $D = \{(\omega_1, b_1), \dots, (\omega_M, b_M)\}$ (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,

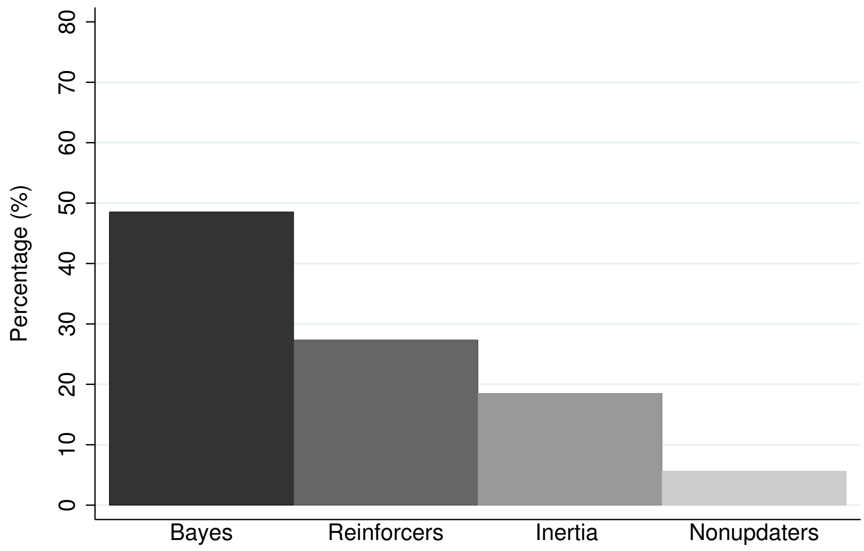


Figure B.1: Proportions of choices classified into each behavioral rule.

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

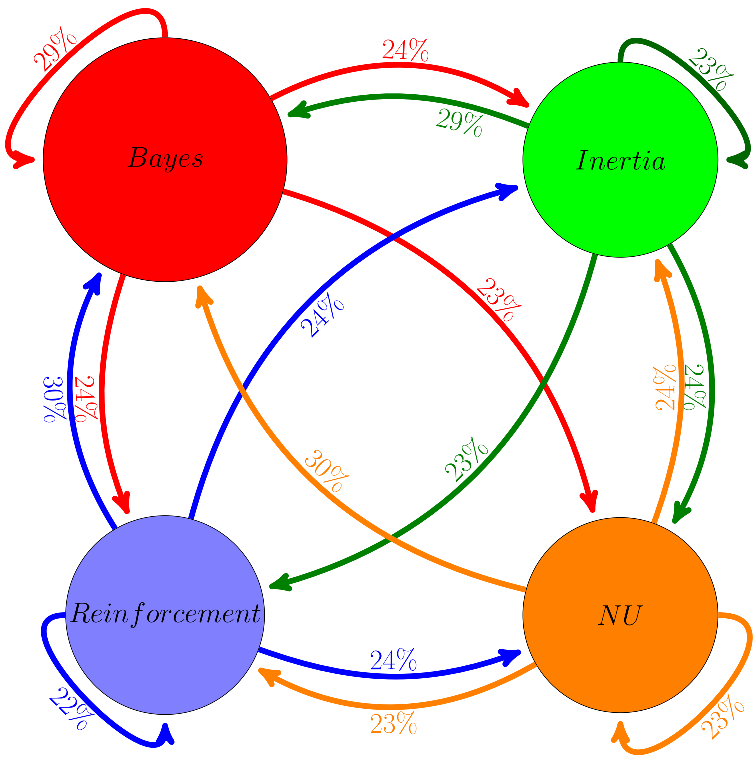
In this section we report the results of the HMM analysis for an artificial dataset where we know the data generating process. In particular, we mimic our dataset simulating 60 trials for $N = 268$ fictitious participants. Each participant follows each of the four rules with equal probability and switches 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.

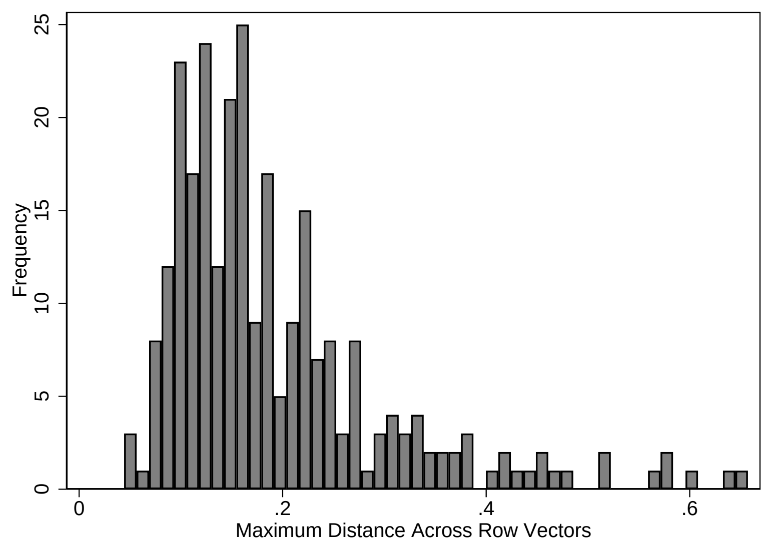


Table F.1: Summary of the analysis of the temporal dynamics for four selected individuals.

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%	

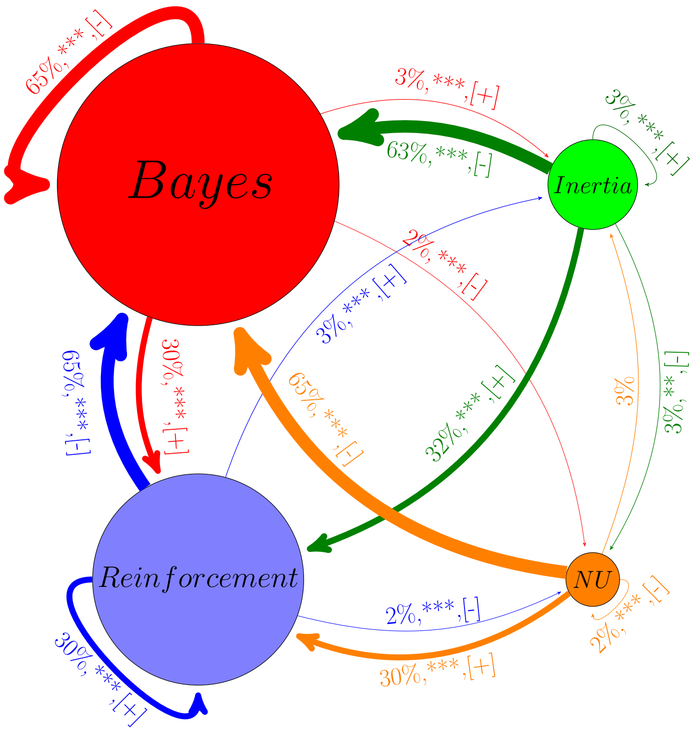
Inertia subjects

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%	

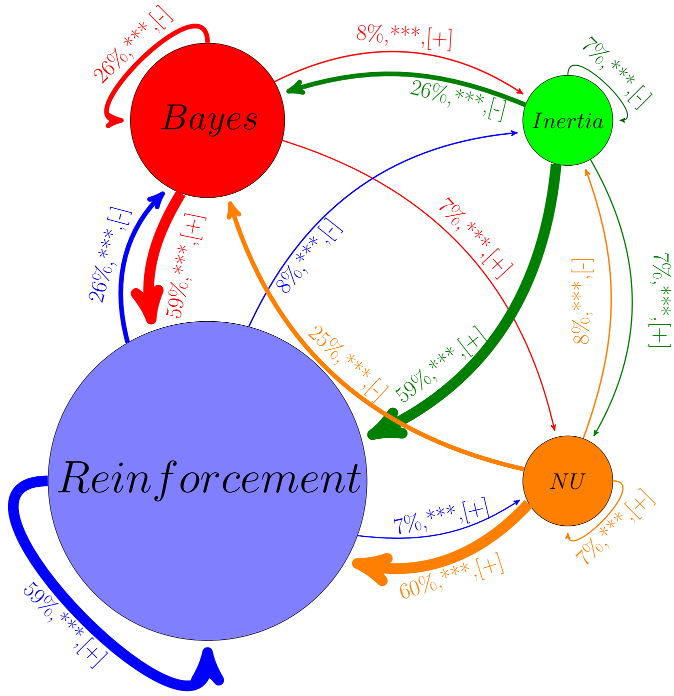
Non-updating subjects

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%	

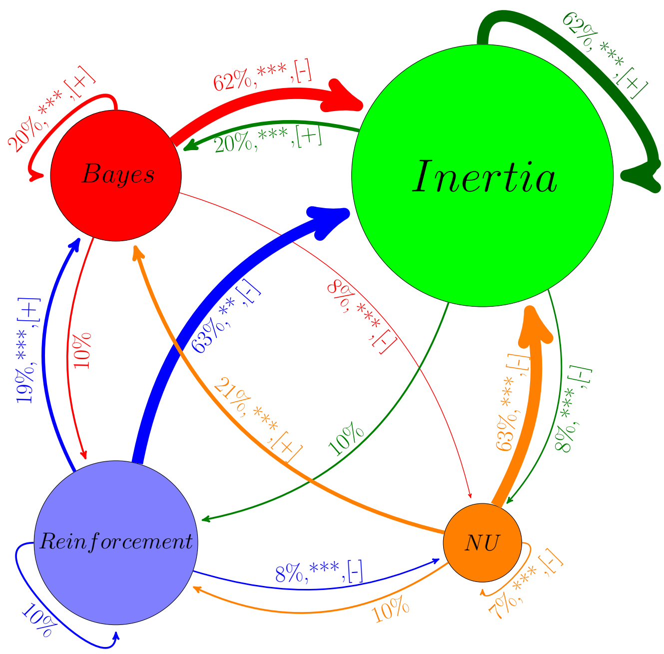
Figure G.1: Average probabilities of switching from one behavioral rule to another conditional on types. Stars indicate MWW between low and high incentives, $*** = p < .01$, $** = p < .05$, $* = p < .10$. [+/−] indicates a significant increase from low to high [+] or vice versa [−]. Arrow thickness indicates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



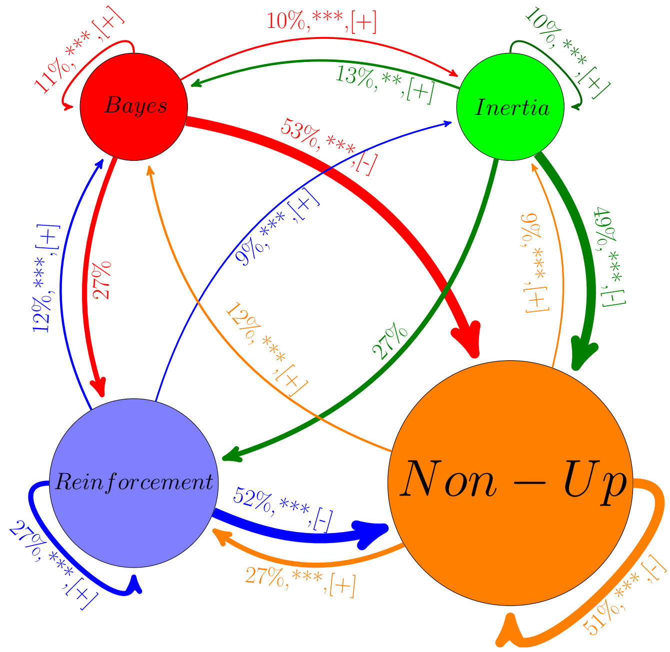
(a) Bayesian subjects.



(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.

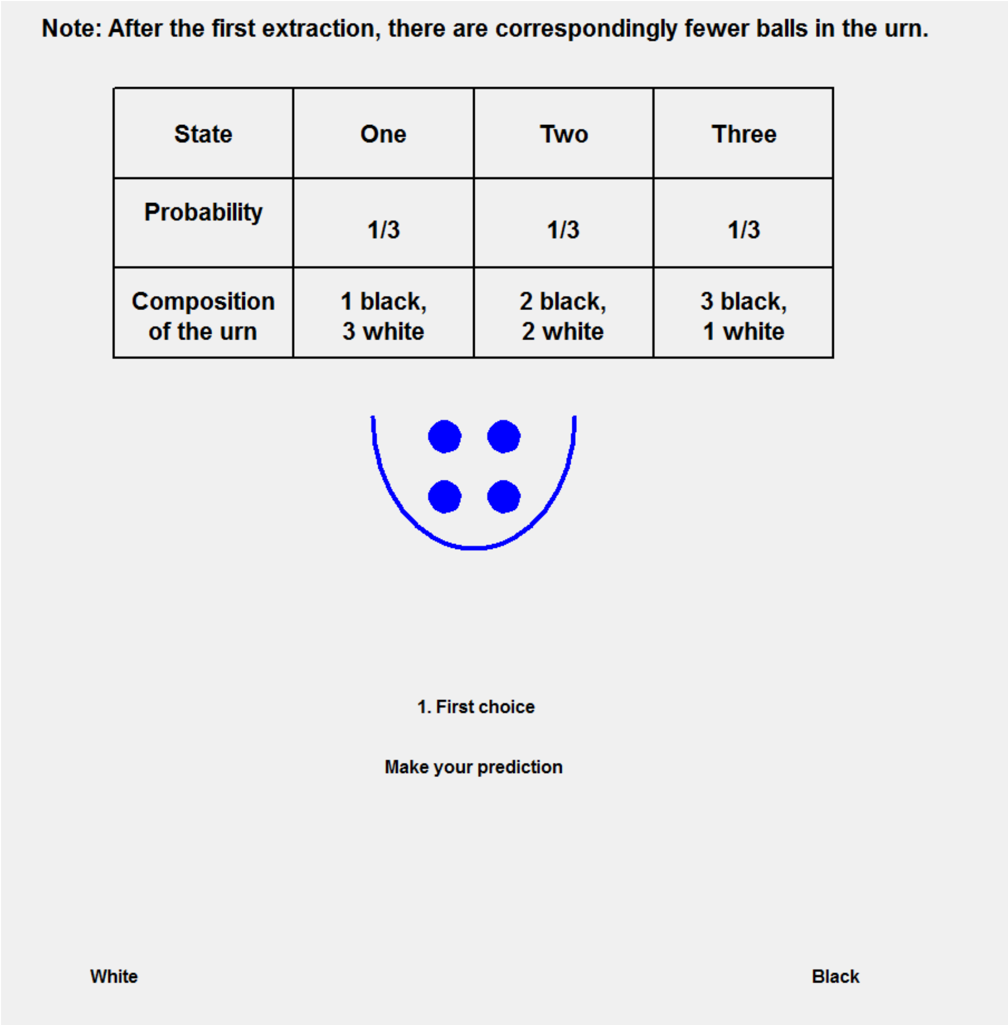
(2) At the bottom (center) you will be informed how many trials you have already completed.

(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).



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

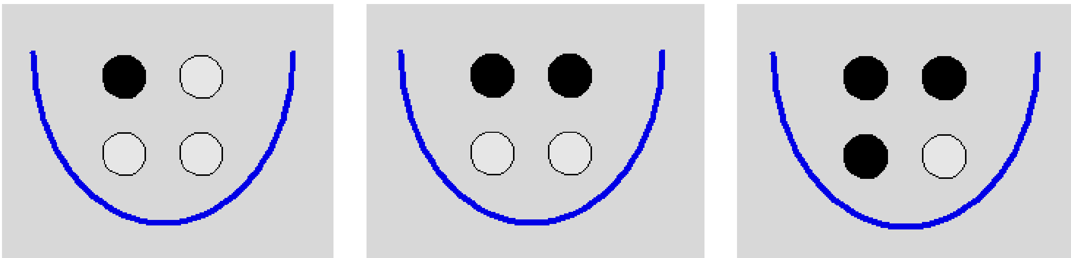
For every correct prediction you receive 18 cents. *[30 cents for the high-incentives treatment]* This means that you will receive this amount if you have predicted a black ball and a black ball is actually drawn, as well as if you have predicted a white ball and a white ball is drawn. If you predicted 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!

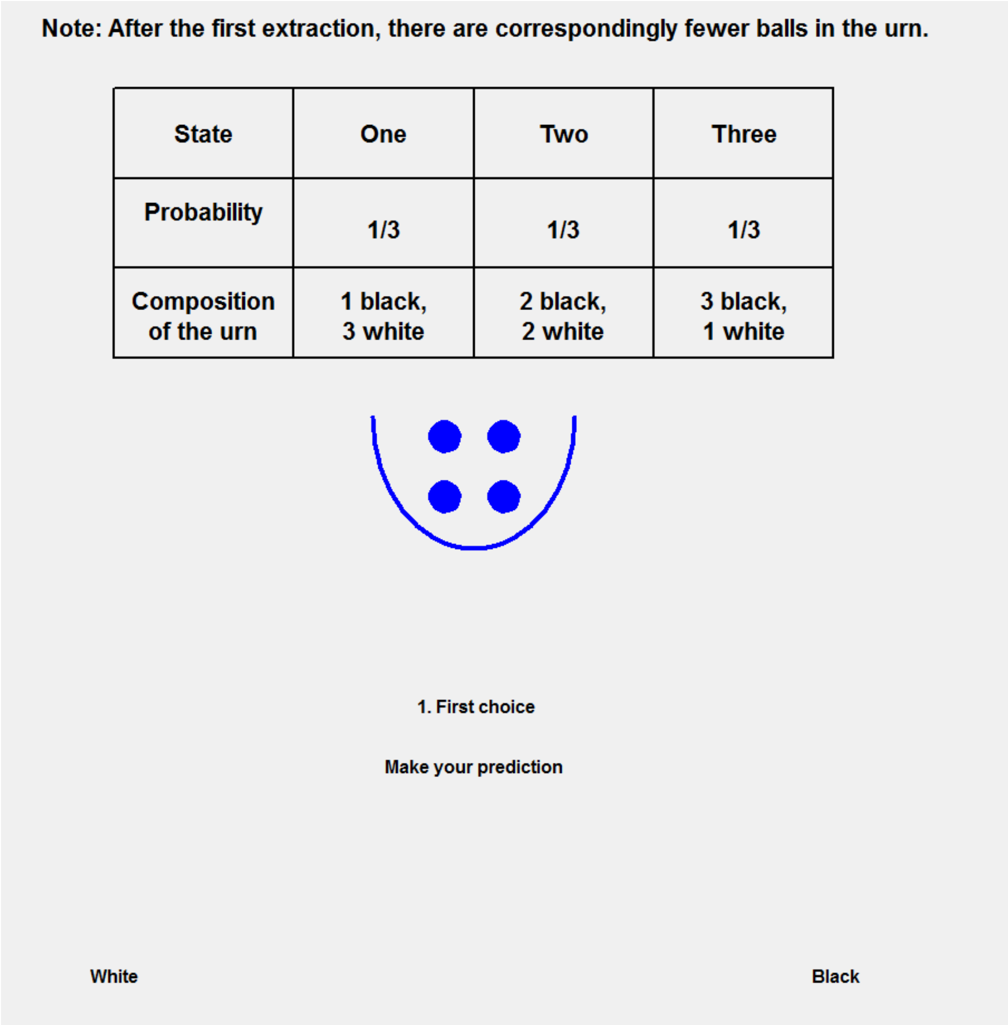
The following table summarizes again the distribution of black and white balls in the different states of the world.

[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.