Multi-state choices with aggregate feedback on unfamiliar alternatives

Philippe Jehiel and Juni Singh[*]

June 2018

Abstract

This paper studies a multi-state binary choice experiment in which in each state, one alternative has well understood consequences whereas the other alternative has unknown consequences. Subjects repeatedly receive feedback from past choices about the consequences of unfamiliar alternatives but this feedback is aggregated over the states. Varying the payoffs attached to the various alternatives in the various states allows us to test whether unfamiliar alternatives are discounted and whether subjects' use of feedback is better explained by generalized reinforcement learning models (in the spirit of the valuation equilibrium, Jehiel and Samet 2007) or by some variant of Bayesian learning model. Our experimental data suggest that there is no discount attached to the unfamiliar alternatives and that generalized reinforcement learning models have a better explanatory power than their Bayesian counterparts.

Key words: Ambiguity, Bounded Rationality, Experiment, Learning, Coarse feedback

JEL Classification: D81, D83, C12, C91

[*] We would like to thank PSE Research grants, LABEX OSE as well as ERC grant no. 742816 for the funds to conduct the laboratory experiments. A special thanks to Maxim Frolov for the execution of the experiment and logistics. We have benefited from the initial comments of Arian Charpin, Elias Bouacida , Itzhak Gilboa, Jean Francois Lassier, Jean Marc Tallon, Julien Combe, Nicolas Gefflot, Nicolas Jacquemet and Philipp Ketz.

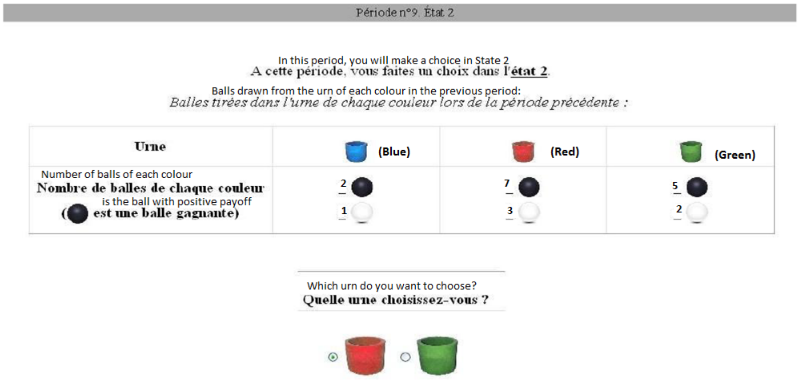
1 Introduction

In many situations, the decision maker faces a choice between two alternatives one of them being more familiar and thus easier to evaluate and another one being less familiar and thus harder to assess. There is generally some information about the less familiar alternative, but this information is typically coarse not being entirely relevant to the specific context of interest.

To give a concrete application, think of the adoption of a new technology by farmers. A farmer has a lot of information about the performance of the current technology but not so much about the new one. The farmer may collect information about the new technology by asking around other farmers who would have previously adopted the new technology. But due to the heterogeneity of the soil and/or the heterogeneity in the ability of the farmers, what works well/poorly for one farmer need not perform in the same way for another. Thus, the feedback received about the new technology is coarse in the sense that it is aggregated over different situations (states in the decision theoretic terminology) as compared to the information held for the old technology. Another example may concern hiring decisions. Consider the hiring for two different jobs, one requiring high skill going together with higher education level and the other requiring lower skills, and assume potential candidates either come from a majority group or a minority group (as determined by nationality, color, caste or religion, say). Presumably, there is a lot of familiarity with the majority group allowing in this group to distinguish the productivity as a function of education as well as past experiences, but in the minority group the information is more likely to be coarse and productivity in that group may not be as clearly related to education or past experiences.

We are interested in understanding how decision makers would make their decisions in multi-state binary decision problems in which decision makers would have precise state-specific information about the performance of one alternative and less precise information about the performance of another alternative where the less precise information takes the form that the decision maker receives aggregate (not state-specific) feedback about the performance of that alternative. Our interest lies in allowing a set of agents to act repeatedly in such environments so as to understand the steady state effects of having agents provided with coarse feedback about some alternatives.

To shed light on this, we consider the following experimental setting. There are two states, $s = 1, 2$. In each state, the decision maker has to choose between two urns identified with a color, Blue and Red in state $s = 1$, Green and Red in state $s = 2$ where the red urn has different payoff implications in states $s = 1$ and 2. Each urn is composed of ten balls, black or white. When an urn is picked, one ball is drawn at random from this urn (and it is immediately replaced afterwards). If a black ball is drawn this translates into a positive payment. If a white ball is drawn there is no payment. One hundred initial draws are made



Période n°9. État 2

In this period, you will make a choice in State 2
A cette période, vous faites un choix dans l'**état 2**.

Balls drawn from the urn of each colour in the previous period:
Balles tirées dans l'urne de chaque couleur lors de la période précédente :

Urne	(Blue)	(Red)	(Green)
Number of balls of each colour **Nombre de balles de chaque couleur** is the ball with positive payoff (● **est une balle gagnante)**	2 ● 1 ○	7 ● 3 ○	5 ● 2 ○

Which urn do you want to choose?
Quelle urne choisissez-vous ?

◉ ○

Figure 1 *Feedback structure for treatment sessions*

for the blue and green urns with no payoff implication for participants, and all subjects are informed of the corresponding compositions of black and white balls drawn from the blue and green urns. Thus, subjects have a precise initial view about the compositions of the Blue and Green urns (these urns correspond to the familiar choices in the motivating examples provided above). In the experiment, the Blue urn has 3 black balls out of ten and the Green urn has 7 black balls out of ten.

Concerning the Red urns, there is no initial information. The Red urns correspond to the unfamiliar choices in the above examples. To guide their choices, subjects are provided with feedback about the compositions of the red urns as reflected by the colors of the balls that were previously drawn when a red urn either in state $s = 1$ or 2 was chosen. More precisely, there are twenty subjects and 70 rounds. In each round, ten subjects make a choice of urn in state 1 and the other ten make a choice of urn in state 2. There are permutations of subjects between rounds so that every subject is in each state $s = 1$ or 2 the same proportion of time. Between rounds, subjects receive feedback about the number of times the Green, Blue and Red urns were picked by the various agents in the previous round, and for each color of urn, they are informed of the number of black balls that were drawn. A typical feedback screen is shown in Figure 1. Note that in the case of the Red urns, this number aggregates the number of black balls drawn from both the red urns picked in state $s = 1$ and the Red urns picked in state $s = 2$ mimicking the kind of coarse information suggested in the motivating examples.

We consider different treatments that differ in the composition of the red urns but maintain the compositions of the Blue and Green urns in all treatments. The initial conditions in these various treatments are thus identical and any difference of behaviors observed in later periods can safely be attributed to the difference in the feedback received by the subjects across the treatments. In treatment 1, the best decision in both states $s = 1$ and 2 would require the choice of the Red urn, but averaging the composition of the Red urns across the two states leads to a less favorable composition than the Green urn. In treatment 2, the best decision would require picking the Red urn in state 1 but not in state 2, but the average composition of the two Red urns dominates that of both the Blue and Green urns. Finally, in treatment 3, the best decision would require picking the Red urn in state 2 but not in state 1. The chosen compositions are depicted in Figure 2.

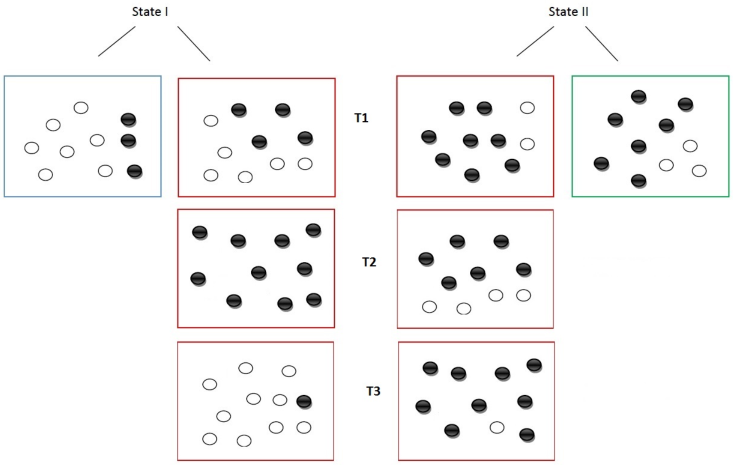


Figure 2 *Set up of the different treatment sessions*

Faced with such an environment, what could be the decision making process followed by subjects? We see the following possible approaches.

First, in the tradition of reinforcement learning (see Barto and Sutton 1998 or Fudenberg and Levine 1998 for textbooks), subjects could assess the strength of the various urns by considering the proportions of black balls drawn in the corresponding urns (aggregating past feedback in some way). One key difficulty in our context is that there is no urn specific feedback for the red urns as the feedback is aggregated over states $s = 1$ and 2, and so the standard reinforcement learning models do not apply. But, following Jehiel and Samet (2007), one could extend the definition of reinforcement learning models to cover situations in which a single valuation would be attached to the red urns whether in state $s = 1$ or 2 and the two red urns would be considered alike in terms of strength by the learning subjects. Jehiel and Samet (2007) have proposed a solution concept called the valuation equilibrium aimed at capturing the limiting outcomes of such generalized reinforcement learning models. In our context, there would be a valuation for each color, Blue, Red and Green; a subject would pick the urn with a color attached to the highest valuation; the valuation attached to the Blue and the Green urns would be 0.3 and 0.7 respectively as reflected by the true compositions of these urns; the valuation of Red would be an average of the proportion of black balls in the Red urns in states $s = 1$ and $s = 2$ where the weight assigned to the Red urns in the various states should respect the proportion of times the Red urn is picked in states $s = 1$ and 2.[1] The valuation equilibrium would predict that in treatment 1, the Red urn is picked in state 1 but not 2; in treatment 2, the Red urns are picked both in states 1 and 2; in treatment 3, the Red urns are picked neither in state 1 nor 2. In our estimation, we will consider a noisy version of such a model in which subjects rely on noisy best-response in the vein of the logit model (as popularized by McKelvey and Palfrey (1995) in experimental economics).

Second, subjects could form beliefs about the compositions of the two Red urns relying on some form of Bayesian updating to adjust the beliefs after they get additional feedback. Note that being informed of the number of times the Blue and the Green urns were picked in the last round is also informative to determine if feedback about the Red urns concerns more state 1 or state 2 (as for example a strong imbalance in favor of the Green urns as opposed to the Blue urns would be indicative that the previous red choices corresponded more to state 1). Of course, such a Bayesian approach heavily depends on the initial prior. When estimating such a model, we will assume that subjects consider a uniform prior over a support that we will estimate and as for the reinforcement learning model we will assume that subjects employ a noisy best response of the logit type.

Another relevant consideration is that the feedback concerning the Red urns is ambiguous to the extent that it does not distinguish between states $s = 1$ and 2. Following the tradition of Ellsberg (1961), one may suspect then that subjects would apply an ambiguity discount to the Red urns (see Gilboa and Schmeidler (1989) for an axiomatization of ambiguity aversion). To cope with this, we propose adding to the previous models (the reinforcement learning and the Bayesian model) an ambiguity discount to the Red urn that we estimate in each case.

Beyond the estimation exercise within each approach, another objective is to analyze which of the generalized reinforcement learning or the generalized Bayesian learning model explains best the observed data.

Our main findings are as follows. First, in our estimation of the generalized reinforcement learning model, we find that that there is no ambiguity discount. That is, despite the inherent ambiguity of the feedback received about the Red urns, the Red urns are not discounted more than the familiar urns. This is simi-

[1] For the sake of illustration, if the Red urns are picked with the same proportion in states $s = 1$ and 2, the valuation should be the unweighted average proportion of black balls in the two Red urns. If the Red urn is only picked in state 1 (resp 2), the valuation of Red should be the proportion of black balls of the Red urn in state 1 (resp 2).

lar to what is being assumed in the valuation equilibrium approach, even if to account for the steady state of the learning model that we propose, there is a need to extend the notion of valuation equilibrium to allow for noisy best-responses in a vein similar to that introduced in the quantal response equilibrium (McKelvey and Palfrey, 1995). Second, we find that the generalized reinforcement learning model explains much better the observed data than the generalized Bayesian learning model. In the last part, we propose some robustness checks and how a generalized reinforcement learning model with an asymmetric ambiguity discount for states $s = 1$ and 2 could slightly outperform our basic reinforcement learning model.

Blue	30 black (B)	70 white (W)
Green	68 black (B)	32 white (W)

Table 1 Information about the relative payoff of actions Blue and Green after 100 random draws is reported at the start of each session.

2 Related Literature

This paper is related to a few other papers beyond those already mentioned. A first line of research related to our study is the framework of case-based decision theory as axiomatized by Gilboa and Schmeidler (1995) in which the various actions in the various states are weighted by some subjectively determined degree of similarity, leading to some linear way of aggregating payoffs obtained in different scenarios. Compared to that literature, our approach provides extra structure on how similarities are formed and the similarity weights given to the various actions in the various states happen to be endogenously shaped by the strategy used by the subjects (which was not considered in the subjective approach pursued in Gilboa and Schmeidler).

Another line of research related to our study includes the possibility that the strategy used by subjects would not distinguish behaviors across different states (Samuelson (2001), Mengel (2012) for theory papers and Grimm and Mengel (2012), Cason et al (2012) or Cownden et al. (2018) for experiments). Our study differs from that line of research in that subjects do adjust their behavior to the state but somehow mix the welfare consequences of some actions (the unfamiliar ones) obtained over different states.

Yet another line related to our study is that of the analogy-based expectation equilibrium (Jehiel (2005) and Jehiel and Koessler (2008)) in which beliefs about other players' behaviors are aggregated over different states (see also Huck et al (2011) for an experiment). Our study differs from that literature in that we are considering decision problems and not games. Yet, viewing nature as a player would allow to see closer connections between the two approaches. To the best of our knowledge, no experiment in this vein has considered environments similar to the type considered here.

Finally, the experimental work of Charness and Levin (2005) considers decision problems in which after seeing a realization of payoff in one urn subjects have to decide whether or not to switch their choices of urns. In an environment in which subjects have a probabilistic knowledge about how payoffs are distributed across choices and states (but have to infer the state from initial information), Charness and Levin observe that when there is a conflict between Bayesian updating and Reinforcement learning, there are significant deviations from optimal choices. While the conclusion that subjects may rely on reinforcement learning more than on Bayesian reasoning is somehow common in their study and our experiment, the absence of ex ante statistical knowledge about the distribution of payoffs across states in our experiment makes it radically different from Charness and Levins experiment.

3 Background and theory

In this Section we define in the context of our experiment a generalization of the valuation equilibrium allowing for noisy best-responses in the vein of the quantal response equilibrium (McKelvey and Palfrey, 1995). We next propose two families of learning models, a generalized reinforcement learning model (allowing for coarse feedback on some alternatives and an ambiguity discount attached to those) as well as a generalized Bayesian model (also allowing for noisy best - responses and a discount on alternatives associated to aggregate feedback). The learning models will then be estimated and compared in terms of fit in light of our experimental data.

3.1 Quantal valuation equilibrium

In the context of our experiment, there are two states $s = 1$ and 2 that are equally likely. In state $s = 1$, the choice is between $Blue$ and Red_1. In state $s = 2$, the choice is between $Green$ and Red_2. The payoffs attached to these four alternatives are denoted by $v_{Blue} = 0.3$, v_{Red_1}, v_{Red_2} and $v_{Green} = 0.7$ where v_{Red_1} and v_{Red_2} are left as free variables to accommodate the payoff specifications of the various treatments.

A strategy for the decision maker can be described as $\sigma = (p_1, p_2)$ where p_i denotes the probability that Red_i is picked in state $s = i$ for $i = 1, 2$. Following the spirit of the valuation equilibrium (Jehiel and Samet, 2007), a single valuation is attached to Red_1, Red_2 so as to reflect that subjects in the experiment only receive aggregate feedback about the payoff obtained when a Red urn was picked either in state $s = 1$ or 2. Accordingly, let $v(Red)$ be the valuation attached to Red. Similarly, we denote by $v(Blue)$ and $v(Green)$ the valuations attached to the Blue and Green urns, respectively.

In equilibrium, we require that the valuations are consistent with the empirical observations as dictated by the equilibrium strategy $\sigma = (p_1, p_2)$. This implies that $v(Blue) = v_{Blue}$, $v(Green) = v_{Green}$ and more interestingly that

$$v(Red) = \frac{p_1 \times v_{Red_1} + p_2 \times v_{Red_2}}{p_1 + p_2} \tag{1}$$

whenever $p_1 + p_2 > 0$. That is, $v(Red)$ is a weighted average of v_{Red_1} and v_{Red_2} where the relative weight given to v_{Red_1} is$p_1/(p_1 + p_2)$ given that the two states $s = 1$ and 2 are equally likely and Red_i is picked with probability p_i.

Based on the valuations $v(Red)$, $v(Blue)$ and $v(Green)$, the decision maker is viewed as picking a noisy best-response where we consider the familiar logit parameterization (with coefficient λ). Formally,

Definition: A strategy $\sigma = (p_1, p_2)$ is a quantal valuation equilibrium if there exists a valuation system $(v(Blue), v(Green), v(Red))$ where $v(Blue) = 0.3$, $v(Green) = 0.7$, $v(Red)$ satisfies (1), and

$$\begin{aligned} p_1 &= \frac{e^{\lambda v(Red)}}{e^{\lambda v(Red)} + e^{\lambda v(Blue)}} \\ p_2 &= \frac{e^{\lambda v(Red)}}{e^{\lambda v(Red)} + e^{\lambda v(Green)}} \end{aligned}$$

We now briefly review how the quantal valuation equilibria look like in the payoff specifications corresponding to the various treatments. In this review, we consider the limiting case in which λ goes to ∞ (thereby corresponding to the valuation equilibria as defined in Jehiel and Samet, 2007).

Treatment 1: $v_{Red_1}= 0.4$ and $v_{Red_2}= 0.8$

In this case, clearly $v(Red) > v(Blue) = 0.3$ (because $v(Red)$ is some convex combination between 0.4 and 0.8). Hence, the optimality of the strategy in state $s = 1$ requires that the Red urn is always picked in state $s = 1$ ($p_1 = 1$). Regarding state $s = 2$, even if Red_2 were picked with probability 1, the resulting $v(Red)$ that would satisfy (1) would only be $\frac{0.4+0.8}{2} = 0.6$, which would lead the decision maker to pick the Green urn in state $s = 2$ given that $v(Green) = 0.7$. It follows that the only valuation equilibrium in this case requires that $p_2 = 0$ so that the Red urn is only picked in state $s = 1$ (despite the Red urns being payoff superior in both states $s = 1$ and 2). In this equilibrium, consistency implies that $v(Blue) < v(Red) = 0.4 < v(Green)$.

Treatment 2: $v_{Red_1}= 1$, $v_{Red_2}= 0.6$

In this case, clearly $v(Red) > v(Blue) = 0.3$ (because any convex combination of 0.6 and 1 is larger than 0.3) and thus $p_1 = 1$. Given that $v_{Red_2} < v_{Red_1}$, this implies that the lowest possible valuation of Red corresponds to $\frac{1+0.6}{2} = 0.8$ (obtained when $p_2 = 1$). Given that this value is strictly larger than $v(Green) = 0.7$, we obtain that it must that $p_2 = 1$, thereby implying that the Red urns are picked in both states. Valuation equilibrium requires that $p_1 = p_2 = 1$ and consistency implies that $v(Blue) < v(Green) < v(Red) = 0.8$.

Treatment 3: $v_{Red_1}= 0.1$, $v_{Red_2}= 0.9$

In this case, we will show that the Red urns are not picked neither in state 1 nor in state 2. To see this, assume by contradiction that the Red urn would (sometimes) be picked in at least one state. This should imply that $v(Red) \geq v(Blue)$ (as otherwise, the Red urns would never picked neither in state $s = 1$ nor 2). If $v(Red) < v(Green)$, one should have that $p_2 = 0$, thereby implying by consistency that $v(Red) = v_{Red_1} = 0.1$ contradicting that $v(Red) \geq v(Blue) = 0.3$. If $v(Red) \geq v(Green)$, then $p_1 = 1$ (given that $v(Red) > v(Blue)$), and thus by consistency $v(Red)$ would be at most equal to $\frac{0.1+0.9}{2} = 0.5$ (obtained when $p_2 = 1$). Given that $v(Green) = 0.7 > 0.5$, we get a contradiction, thereby implying that no Red urn can be picked in a valuation equilibrium.

As explained above the value of $v(Red)$ in the valuation equilibrium varies from being below $v(Blue)$ in treatment 3 to being in between $v(Blue)$ and $v(Green)$ in treatment 1 to being above $v(Green)$ in treatment 2, thereby offering markedly different predictions according to the treatment in terms of long run choices. Allowing for noisy as opposed to exact best-responses would still allow to differentiate the behaviors across the treatments but in a less extreme form (clearly, if $\lambda = 0$ behaviors are random and follow the lottery $50 : 50$ in every state and every treatment, but for any $\lambda > 0$, behaviors would be different across treatments).

3.2 Learning Models

We will consider two families of learning models to explain the choice data observed in the various treatments of the experiment: A generalized version of reinforcement learning model in which choices are made on the basis of the valuations attached to the various colors of urns and valuations are updated based on the observed feedback, and a Bayesian learning model in which subjects update their prior belief about the composition of the Red urns based on the feedback they receive. In each case, we will assume that subjects care only about their immediate payoff and do not integrate the possible information content that explorations outside what maximizes their current payoff could bring. This is -we believe- justified to the extent that in the experiment there are twenty subjects making choices in parallel and that the feedback is anonymous making the informational value of the experimentation by a single subject rather small (it would be exactly 0 if we were to consider infinitely large populations of subjects and we suggest it is negligible when there are twenty subjects).

3.2.1 Generalized reinforcement learning

Usual reinforcement learning models assume that the various strategies are being reinforced as a function of the payoff obtained when these strategies are being used. In the context of our experiment, subjects receive feedback about how the choices made by all subjects in the previous period translated into black (positive payoff) or white (null payoff) draws. More precisely, the feedback concerns the number of Black balls drawn when a $Blue, Green$ or Red urn was picked in the previous period as well as the number of times an urn with that color was then picked. Accordingly, at each time $t = 2, \ldots 70$, one can define for each possible color $C = B, R, G$ (for Blue, Red, Green) of urn(s):

$$UC_t = \frac{\#(\textit{Black balls} \text{ drawn in urns with color } C \text{ at } t-1)}{\#(\text{an urn with color } C \text{ picked at } t-1)}. \qquad (2)$$

UC_t represents the strength of urn(s) with color C as reflected by the feedback received at t about urns with such a color. Note the normalization by #(an urn with color C picked at $t-1$) so that UC_t is comparable to a single payoff attached to choosing an urn with color C.

We will let BC_t denote the value attached to an urn with color C at time t and BC_{init} denote the initial value attached to an urn with that color. For *Green* and *Blue* there is initial information and it is natural to assume that

$$\begin{aligned} BB_{init} &= \frac{30}{100} = 0.3 \\ BG_{init} &= \frac{68}{100} = 0.68 \end{aligned}$$

whereas for *Red*, the initial value BR_{init} is a priori unknown and it will be estimated in light of the observed choice data.

Dynamics of BC_t:

Concerning the evolution of BC_t, we assume that for some (ρ_U, ρ_F), we have:

$$\begin{aligned} BR_t &= \rho_U \times BR_{t-1} + (1-\rho_U) \times UR_t \\ BB_t &= \rho_F \times BB_{t-1} + (1-\rho_F) \times UB_t \\ BG_t &= \rho_F \times BG_{t-1} + (1-\rho_F) \times UG_t \end{aligned}$$

In other words, the value attached to color C at t is a convex combination between the value attached at $t-1$ and the strength of C as observed in the feedback at t. Observe that we allow the weight to be assigned to the feedback to be different for the Red urns on the one hand and the Blue and Green urns on the other to reflect the idea that when a choice is better known as is the case for more familiar alternatives (here identified with urns *Blue* and *Green*) the new feedback may be considered as less important to determine the value of it. Accordingly, we would expect that ρ_F is larger than ρ_U and we will be concerned whether such a comparison shows up when we do the estimations of the model.[2]

Choice Rule:

Given that the feedback concerning the Red urns is aggregated over states $s = 1$ and 2, there is extra ambiguity as to how well BR_t represents the valuation

[2] There are many variants that could be considered. For example, one could have made the weight of the new feedback increases linearly or otherwise with the number of times an urn with that color was observed, or one could have considered that the weight on the feedback is a (decreasing) function of t so as to reflect that as more experience accumulates the new feedback becomes less important. These extensions did not seem to improve much how well we could explain the data and this is why in this presentation we follow the simpler approach described in the main text.

of Red_1 or Red_2 as compared to how well BG_t or BB_t represent the valuations of *Blue* and *Green*.

The valuation equilibrium (or its quantal extension as presented above) assumes that BR_t is used to assess the strength of Red_s whatever the state $s = 1, 2$. In line with the literature on ambiguity aversion as initiated by Ellsberg (1961), it is reasonable to assume that when assessing the urn Red_s, $s = 1, 2$, subjects apply a discount $\delta \geq 0$ to BR_t.[3] Allowing for noisy best-responses in the vein of the logit specification, this would lead to probabilities p_{1t} and p_{2t} of choosing Red_1 and Red_2 as given by

$$p_{1t} = \frac{e^{\lambda(BR_t - \delta)}}{e^{\lambda(BR_t - \delta)} + e^{\lambda BB_t}}$$
$$p_{2t} = \frac{e^{\lambda(BR_t - \delta)}}{e^{\lambda(BR_t - \delta)} + e^{\lambda BG_t}}$$

The learning model just described is parameterized by (ρ_R, ρ_B, δ, λ, BR_{init}). In the next Section, it will be estimated pooling the data across all three treatments using the maximum likelihood method. Particular attention will be devoted to whether $\delta > 0$ is needed to explain better the data, whether $\rho_F > \rho_U$ as common sense suggests, as well as the estimated value of λ and the obtained likelihood for comparison with the Bayesian model to be described next.

3.2.2 Generalized Bayesian Learning Model

As an alternative learning model, subjects could form some initial prior belief regarding the compositions of Red_1 and Red_2, say about the chance that there are k_i black balls out of 10 in Red_i, and update these beliefs after seeing the feedback using Bayes' law.

Let us call $\beta_{init}(k_1, k_2)$ the initial belief of subjects that there are k_i black balls out of 10 in Red_i. In the estimations, we will allow the subjects to consider that the number of black balls in either of the two Red urns can vary between $k_{\inf}$ and $k_{\sup}$ with $0 \leq k_{\inf} \leq k_{\sup} \leq 10$ and we will consider the uniform distribution over the various possibilities. That is, for any $(k_1, k_2) \in [k_{\inf}, k_{\sup}]^2$

$$\beta_{init}(k_1, k_2) = \frac{1}{(k_{\sup} - k_{\inf} + 1)^2},$$

and $\beta_{init}(k_1, k_2) = 0$ otherwise. The values of $k_{\inf}$ and $k_{\sup}$ will be estimated.

[3] One possible rationale following the theoretical construction of Gilboa and Schmeidler (1989) is that the proportion of Black balls in Red_1 and Red_2 is viewed as being in the range [BR - δ, BR + δ] and that subjects adopt a maxmin criterion, leading them to consistently use $BR - \delta$ to assess both Red_1 and Red_2.

Note that considering a convex combination between the most optimistic and pessimistic view along the lines of Hurwicz could not be distinguished from the model we estimate as long as there is no smaller weight on the most pessimistic assessment than on the most optimistic assessment. Given the identification problem, it is reasonable to estimate the version we propose.

Dynamics of the beliefs:

At time $t+1$, the feedback received by a subject can be formulated as (b, g, n) where b, g are the number of blue and green urns respectively that were picked at t, and n is the number of black balls drawn from the *Red* urns. To simplify a bit the presentation, we assume there is no learning on the urns *Blue* and *Green* for which there is substantial initial information. In the robustness checks, we allow for Bayesian updating also on the compositions of the blue and green urns, and obtain that adding learning on those urns do not change our conclusion.

Also to simplify a bit further the presentation, we assume that subjects consider that overall in the feedback they are exposed to there is an equal number of $s = 1$ and $s = 2$ decisions (allowing the subjects to treat these numbers as resulting from a Bernoulli distribution would again not alter our conclusions, see the appendix for details of the corresponding more elaborate Bayesian learning model. In this case, the feedback can be presented in a simpler way, because knowing (b, g, n) now allows the subjects to infer that $m_1 = 10 - b$ choices of *Red* urns come from state $s = 1$ and $m_2 = 10 - g$ choices of *Red* urns come from state $s = 2$. Accordingly, we represent the feedback as (m_1, m_2, n). Clearly, the probability of observing m_1, m_2, n when there are k_1 and k_2 black balls in Red_1 and Red_2 respectively is given by:

$$Pr(m_1, m_2, n|k_1, k_2) = \sum_{\substack{n_1 \leq m_1 \\ n_2 \leq m_2 \\ n_1+n_2=n}} \left(\begin{array}{c} m_1 \\ n_1 \end{array} \right) \left(\begin{array}{c} m_2 \\ n_2 \end{array} \right) (k_1/10)^{n_1}(1-k_1/10)^{m_1-n_1}(k_2/10)^{n_2}(1-k_2/10)^{m_2-n_2}$$

where $\left(\begin{array}{c} m \\ n \end{array} \right) = \frac{m!}{(m-n)!n!}$.

The posterior at $t+1$ about the probability that there are k_1 and k_2 black balls in Red_1 and Red_2 after observing (m_1, m_2, n) at t is then derived from Bayes law by

$$\beta_{t+1}(k_1, k_2) = \frac{\beta_t(k_1, k_2) \cdot \Pr(m_1, m_2, n|k_1, k_2)}{\sum_{r_1, r_2} \beta_t(r_1, r_2) \cdot \Pr(m_1, m_2, n|r_1, r_2)}.$$

with $\beta_1(k_1, k_2) = \beta_{init}(k_1, k_2)$.

Define $v_t^{Bayes}(Red_i) = \sum_{k_i} \frac{k_i}{10} \beta_t(k_i)$ where $\beta_t(k_i) = \sum_{k_{-i}} \beta_t(k_i, k_{-i})$ as the time t expected proportion of black balls in Red_i given the distribution β_t.

Choice Rule:

As for the generalized reinforcement learning model, we allow both for noisy best responses and we introduce an ambiguity discount δ for the evaluation of the Red urns.[4]Accordingly, the probabilities p_{1t} and p_{2t} of choosing Red_1 and

[4] Some might dispute that the ambiguity discount is not so much in the spirit of the Bayesian model in which case one should freeze this parameter to be 0.

Red_2 at time t in the generalized Bayesian learning model are given by:

$$p_{1t} = \frac{e^{\lambda(v_t^{Bayes}(Red_1)-\delta)}}{e^{\lambda(v_t^{Bayes}(Red_1)-\delta)} + e^{\lambda v(Blue)}}$$

$$p_{2t} = \frac{e^{\lambda(v_t^{Bayes}(Red_2)-\delta)}}{e^{\lambda(v_t^{Bayes}(Red_2)-\delta)} + e^{\lambda v(Green)}}$$

where as our simplification implies we assume[5] that $v(Blue) = 0.3$ and $v(Green) = 0.7$. The proposed (generalized) Bayesian learning model is parameterized by $(k_{\inf}, k_{\sup}, \lambda, \delta)$ and these parameters will be estimated by the maximum likelihood method in the next Section.

4 Results

4.1 Further Description of the Experimental Design

The computerized experiments were conducted in the Laboratory at Maison de Sciences Economiques (MSE) between March 2015 and November 2016, with some additional sessions for the Coarse treatment running in March 2017. Upon arrival at the lab, subjects sat down at a computer terminal to start the experiment. Instructions were handed out and read aloud before the start of each session.

The experiment consisted of three main treatments which varied in the payoff of the Red urns as explained above. In addition we had two other treatments referred to as controls in which in the same payoff conditions as those described for treatments 1 and 2, subjects received state specific feedback about the Red urns i.e the feedbacks for Red_1 and Red_2 appeared now in two different columns. The purpose of these control treatments was to check whether convergence to optimal choices was observed in such more standard feedback scenarios.

Each session involved 18-20 subjects[6] and four sessions were run for each treatment and control. In total 235 subjects participated in the experiment, drawn from the participant pool at MSE.

In all treatment and control sessions, subjects were split up equally into two states, State 1 and State 2. Each session consisted of seventy rounds where players were randomly assigned a new state at the start of each round. The players knew the state they were assigned to but did not know the payoff of the actions to be picked within each state[7] In each state, players were asked

[5] We present a model that allows players to update $v(Blue)$ and $v(Green)$ according to Bayes rule in the robustness checks.

[6] **Note that when 18 subjects participated in the session, Bayes updating was modified accordingly where the number of subjects in state** $s = 1$ **was equal to** $s = 2$

[7] With the exception of the initial information for some actions.

to choose between two actions as detailed in Figure 1. The feedback structure for the main treatments has been explained above (see the appendix for the feedback in control).

Subjects were paid a turn-up fee of 5 € and in addition to this were given the opportunity to earn 10 € depending on their choice in the experiment. Specifically, for each subject, two rounds were drawn at random and a subject earned an extra 5 € for each black ball that was drawn from their chosen urn in these two rounds. The average payment was around 11 € per subject, including the turn-up fee. All of the sessions lasted between 1 hour and 1.5 hour, and subjects took longer to consider their choices at the start of the experiment.

4.2 Results

We first present descriptive statistics and next present the statistical analysis.

4.2.1 Preliminary findings

In Figs 3 and 4, we report how the choices of urns vary with time and across treatments[8]. Across all these sessions, initially, subjects are more likely to choose the Red urn than the Blue urn in state 1 and they are more likely to choose the Green urn than the Red urn in state 2. This is presumably because the Green urn being more rewarding than the Blue urn and the Red urns looking alike in states 1 and 2, the difference of payoffs attached to the more familiar urns tilt the choice toward the Red urn more in state 1 than in state 2.

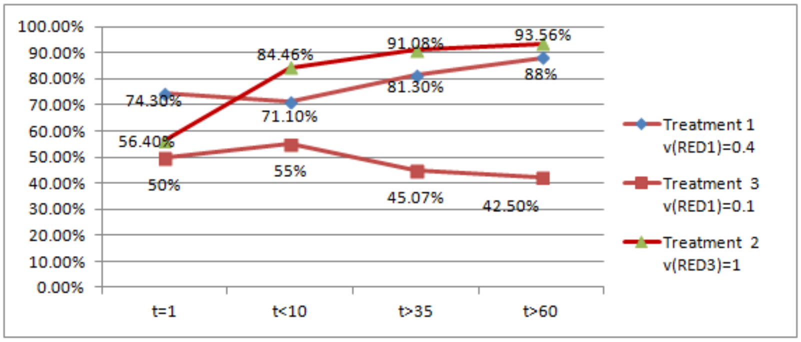


Figure 3 *Red in State 1: Coarse Information Case*

[8] In Appendix B, we report the choices in the Control sessions in which subjects receive state-specific feedback for the Red urns.

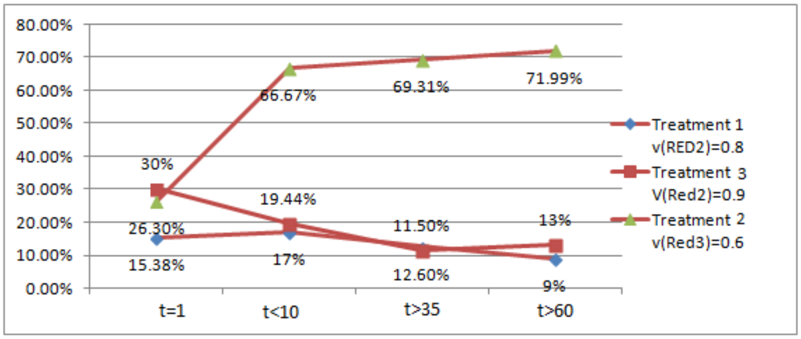


Figure 4 *Red in State 2: Coarse Information Case*

The more interesting question concerns the evolution of choices. Roughly, in state 1, we see toward the final rounds, a largely dominant choice of the Red urns in treatments 1 and 2 whereas Red in state 1 is chosen less than half the time in treatment 3.

Concerning state 2, we see that in the final rounds, the Red urns are rarely chosen in treatments 1 and 3 and chosen with high probability in treatment 2.

The qualitative differences of the choices in the final rounds among the three treatments and the two states are in line with the prediction of the valuation equilibrium even if some noise in the best-response is presumably needed especially for treatment 3 in state 1 to explain why about 40% of choices correspond to Red.

4.2.2 Statistical estimations

Generalized reinforcement learning

The estimations of the parameters of the generalized reinforcement learning model together with the log likelihood[9] are given in the following Table.

Table 2 Parameters for generalized reinforcement learning model

ρ_U	ρ_F	δ	BR_{ini}	λ	Likelihood
0.45 (0.008)	0.599 (0.009)	0.00 (0.009)	0.44 (0.008)	5.23 (0.0093)	7626.6 -

By way of comparison, a complete random choice model where in every state, subjects would randomize 50:50 between the two choices would result in

[9] Likelihood throughout the paper refers to the negative of the log likelihood. Thus, the lower the Likelihood, the better the model. Standard errors are reported in brackets.

a negative log likelihood of L=11402, which is much higher than 7626.6.[10]

We now discuss the most salient aspects of the estimations.

The finding that $\rho_F > \rho_U$ seems natural to the extent that for the familiar urns, the feedback should affect less how the valuations are updated.

The finding that BR_{init} is slightly below 0.5 may be interpreted along the following lines. In the absence of any information, an initial value of 0.5 would be the one dictated by the principle of insufficient reason, but the uncertainty attached to the unfamiliar urns may lead to some extra discount in agreement with ambiguity aversion as reported in Ellsberg.[11]

The most interesting observation concerns δ which is estimated to be 0. Even though the feedback for the Red urns is ambiguous (because it is aggregated over the two states), the valuations for Red are not discounted as if subjects were ambiguous neutral from that perspective.

Thus, what our estimation suggests is that while there may be some initial ambiguity aversion relative to the unfamiliar choices (as reflected by BR_{init} being smaller than 0.5), this ambiguity aversion disappears as the subjects have made more decisions despite the fact that some incompressible level of ambiguity remains attached to the Red urns due to the coarseness of the feedback about the Red urns.

Generalized Bayesian learning model:

The estimated parameters for the generalized Bayesian learning model are given in the following Table.

Table 3 Parameter for Bayesian model with bounds

λ	k_{inf}	k_{sup}	δ	Likelihood
7.488	3	7	0.003	8816.2
(0.021)	-	-	(0.0012)	-

The value of δ =0.003 implies that with the Bayesian model, the subjects show some mild form of ambiguity aversion. However we cannot statistically reject the hypothesis that $\delta = 0$, which implies that with the bayesian model too, there is no significant ambiguity discount similarly as what we found in the generalized reinforcement learning estimations. For the support of initial prior, we found that k_{inf}=3 and k_{sup}=7.[12] our finding in the generalized reinforcement learning estimations.We also note that the value of λ is slightly higher than that for the reinforcement model.randomness with the Bayesian model and that the bounds for the number of black balls in either of the two Red urns varies between

[10] Optimizing on the probability of Red_1 vs Red_2 would lead to assume that Red_1 is chosen with probability p_1=0.7 and Red_2 is chosen with probability p_2= 0.3 with a negative log likelihood of L=9899.6. As seen from Table 2, reinforcement learning explains the data better than models that would assume behaviors are not responsive to feedback.

[11] The difference $0.5 - 0.44 = 0.06$ can be interpreted as measuring the ambiguity aversion of choosing an unfamiliar urn.

[12] The value of the bounds correspond to v_{Blue}=0.3 and v_{Green}=0.7 respectively and so maybe the compositions of the familiar urns serve as anchoring the support of the priors.

Comparing the two models:
Maybe the most important question is which of the Bayesian learning model or the reinforcement learning model explains best the experimental data. We consider three methods of comparisons, all establish that the reinforcement learning model outperforms the Bayesian learning model. First, looking at the likelihood for the two models, we see that the Bayesian Model explains less well the data than the generalized reinforcement learning model. Second, we perform a Vuong test to compare the performance of the two model statistically. Under the null hypothesis that both models perform equally well, we conclude that the null can be rejected in favor of the reinforcement model. Specifically,

$$H_0 = E(L(\theta_R; x_d)) = E(L(\theta_B; x_d))$$
$$H_a = E(L(\theta_R; x_d)) \neq E(L(\theta_B; x_d))$$

where x_d is the individual data point, θ_R is the set of parameters estimated via reinforcement learning, θ_B is the set of parameters estimated via Bayesian learning, $L(\theta_R; x_d)$ is the log likelihood under reinforcement model and $L(\theta_B; x_d)$ is the log likelihood under bayesian for each data point d. The vuong statistics is then defined by,

$$V_{stat} = \sqrt{N}\frac{\bar{m}}{S_m}$$

where $\bar{m}$= log($Pr(x_i|\theta_R)$-log($Pr(x_i|\theta_B)$, N is the total number of observations and S_m is the sample standard deviation.

V_{stat} tests the null hypothesis that the two models are equally close to the true data generating process, against the alternative that one of the model is closer [13]. The obtained $V_{stat} = 25.01$ implies that the reinforcement model is a better fit to our experiment data than the Bayesian model. Finally, we use the Bayesian Information Criterion (BIC) [14] to determine the goodness of fit in the two models. The corresponding finding is in line with the findings of Vuong test and reinforcement model is shown to outperform the Bayesian one according to the BIC approach.

Table 4 BIC values for the two competing models

Valuation model	Bayesian model
1.52 x 10_4	1.763 x 10_4

The model with the lower BIC corresponds to the model that is closer to the data generating process. As seen from table 4, we can conclude that the reinforcement model performs better than the Bayesian one in explaining the data.

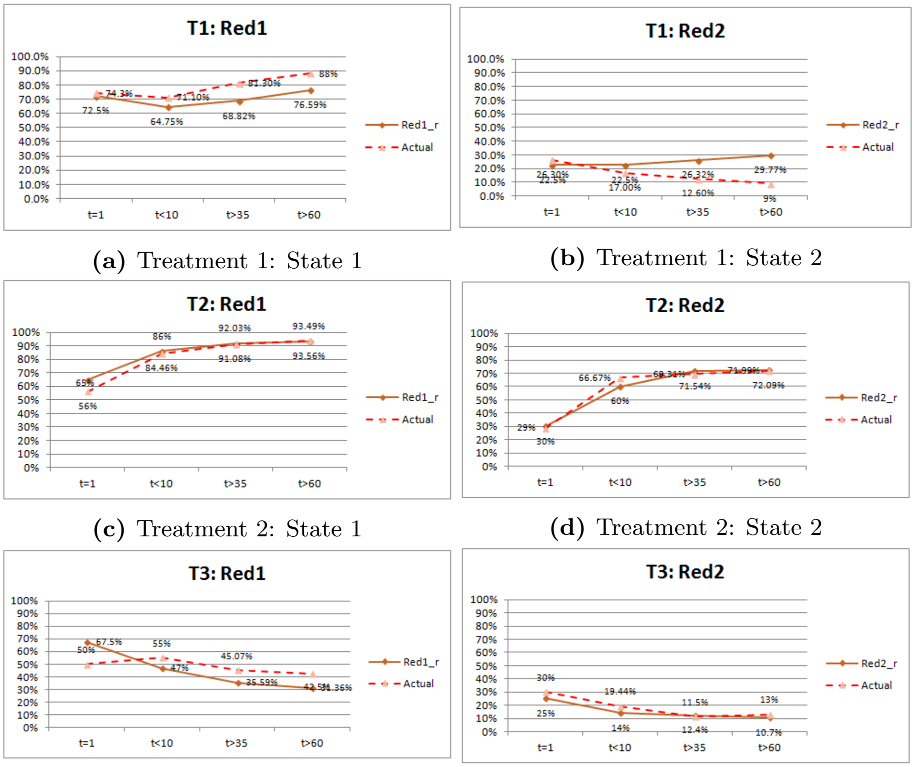
[13] Vuong test compares the predicted probabilities of two non nested models. It computes the difference in likelihood for each observation i in the data. A high positive V_{stat} implies Model 1 is better than Model 2 where $\bar{m}$= log($Pr(x_i|Model1)$-log($Pr(x_i|Model2)$

[14] In statistics, the Bayesian information criterion (BIC) or Schwarz criterion (also SBC, SBIC) is a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is based, in part, on the likelihood function.

In the next section, we run several robustness checks on both the reinforcement and the bayesian models, but the conclusion on the comparison between them remains unchanged.

4.3 Comparing the Reinforcement learning model to the data

While we have established that the Generalized reinforcement learning model explains better the data than its Bayesian counterpart, it is of interest of see how the obtained frequencies of choices as generated by such a model with estimations as reported in Table 2 compare to the observed frequencies that come from our experimental data. In Figure 5, we report the simulated frequencies of urn choices using the reinforcement model across all time periods and treatments. Across all these sessions, our simulated frequencies remain close to the actual frequencies in Treatment 1. Allowing for a different λ in Treatment 1, we obtain a much better fit and improve significantly the fit in this treatment as shown in Appendix D.



(a) Treatment 1: State 1

(b) Treatment 1: State 2

(c) Treatment 2: State 1

(d) Treatment 2: State 2

(e) Treatment 3: State 1

(f) Treatment 3: State 2

Figure 5 Simulated choices using the reinforcement model

4.3.1 Robustness Checks

As many variants of reinforcement learning models and Bayesian models could be considered, we review here a few of these and suggest that our basic conclusions remain the same in these variants.

Generalized reinforcement learning model

Regarding reinforcement models, we consider the following variants. First,we allow the speed of adjustment of the valuation of the Red urns to differ across the two states as a large imbalance in the number of green urns as opposed to blue urns in the feedback is indicative that the feedback concerned more Red urns in state 1 than in state 2. Specifically, we now introduce a new parameter μ and specify the weight on the previous valuation to satisfy

$$\rho_{U1} = \rho_U \times 1 - \mu(\frac{NB}{NG + NB} - 0.5)]$$
$$\rho_{U2} = \rho_U \times [1 - \mu(\frac{NG}{NG + NB} - 0.5)]$$

where NB and NG are the respective numbers of Blue and Green urns appearing in the feedback. Intuitively, one would expect that as $NB < NG$, more weight on the feedback would be assigned to the new valuation of Red in state 1 so that we expect $\mu > 0$. The estimations of this extended model are reported in the following table.

Table 5 Parameters for generalized reinforcement learning model

ρ_U	ρ_F	δ	BR_{ini}	λ	μ	Likelihood
0.45 (0.236)	0.599 (0.04)	0.00 (0.26)	0.49 (0.188)	5.21 ((0.009)	-0.218 (0.19)	7625.4 (-)

Surprisingly, we find μ with a sign different from what we expected. Yet, the gain in likelihood is small and one cannot reject the hypothesis that $\mu = 0$. Thus, this extended model does not explain the data better than our previously proposed version.

A different idea somewhat related to the one just discussed is that subjects would apply a different discount to the Red urn in state 1 and 2 maybe because after some while they would consider the feedback for the Red urns to be more indicative of Red in state 1 than in state 2 (again because of the imbalance of the number of Blue and Green urns in the feedback). This leads us to consider an extended version with two different discounts δ_1 and δ_2 for Red in state 1 and 2 while keeping the other aspects of the dynamics unchanged as compared to the main reinforcement learning model. That is, the only change in this variant is in the choice rule

Choice Rule:

$$Prob(y = R) = \frac{exp^{\lambda(BR_1 - \delta_1)}}{exp^{\lambda(BR_1 - \delta_1)} + exp^{\lambda BB}}$$

$$And \quad in \quad State2$$

$$Prob(y = R) = \frac{exp^{\lambda(BR_2 - \delta_2)}}{exp^{\lambda(BR_2 - \delta_2)} + exp^{\lambda BG}}$$

and one would expect $\delta_1 < \delta_2$ according to the logic just discussed. the estimated parameters for this variant are reported in the following table.

Table 6 Parameters for robustness check. Standard errors are reported in brackets

ρ_U	ρ_F	δ_1	δ_2	BR_{ini}	λ	Likelihood
0.42 (0.007)	0.57 (0.009)	0.00 (0.028)	0.03 (0.005)	0.46 (0.005)	5.093 (0.031)	7610.6 (-)

In this variant, we see a slight discount for Red2 but not for Red1. The likelihood for this model is better than for the original model and the hypothesis $\delta_1 = \delta_2 = 0$ is rejected under significance level 0.01. While this extension does improve the explanation of the data, we find only a modest level of ambiguity aversion applied to Red in state 2 when allowed to differ from the ambiguity aversion to Red in state 1.[15]

Generalized Bayesian learning model

For the Bayesian model, one could argue that instead of fixing v_{Blue}=0.3 and v_{Green}=0.7, the values could be updated via

$$Pr(x, n_x|k_x) = \sum \left(\begin{array}{c} b \\ n_x \end{array} \right) (k_x/10)^{n_x}(1 - k_x/10)^{x-n_x}$$

where x is the number of times Blue (Green) was chosen in $s_1(s_2)$, n_x is the number of times a black ball was picked out of the x draws and $k_x \in [k_{\inf}, k_{\sup}]$.

The rest of the Bayesian updating follows from Section 3.2.2. This model performs better than the generalized Bayesian one in terms of likelihood. However, this extended model is still statistically [16] worse than the generalized valuation model.

A more elaborate approach would be to take into account the probability of having r_i state i in the 20 observation of the feedback. We represent the feedback as (b, g, n) where b, g are the number of draws from blue and green

[15] We also considered the possibility that subjects would use a different slope to appreciate payoffs above 0.5 and payoffs below 0.5 in the spirit of prospect theory (with a reference payoff fixed at 0.5), but such a variant did not result in an improvement of the likelihood, hence we do not report it here.

[16] The Vuong test was conducted with the null hypothesis that both models explain the data equally well. The null was rejected in favor of the generalized reinforcement learning model with V_{stat}= 24.72.

Table 7 Parameter for Bayesian model with Blue and Green updating

λ	k_{inf}	k_{sup}	δ	Likelihood
8.69 (0.25)	3 (-)	7 (-)	0.008 (0.003)	8583.8 (-)

urns and n is the number of black balls in *Red*. We modify the generalized bayesian model by taking into account the probability of having x states out of $s = 1$ 20.

$$Pr(b, g, n|k_1, k_2) = \sum_x \left(\begin{array}{c} 20 \\ x \end{array} \right) \left(\ \frac{1}{2^{20}} \ \right) Pr(m_1 = x - b, m_2 = 20 - x - g, n|k_1, k_2)$$

where x is the number of times state $s = 1$ was observed in one round, $Pr(m_1 = x - b, m_2 = 20 - x - g, n|k_1, k_2)$ is defined as in section 3.2.2 where the total number of players in each session is 20.

The dynamics of beliefs is now given by

$$\beta_{t+1}(k_1, k_2) = \frac{\beta_t(k_1, k_2) \cdot \Pr(b, g, n|k_1, k_2)}{\sum_{r_1, r_2} \beta_t(r_1, r_2) \cdot \Pr(b, g, n|r_1, r_2)}.$$

with $\beta_1(k_1, k_2) = \beta_{init}(k_1, k_2)$. The other aspects of the Bayesian learning model are identical to those considered in section 3.2.2.

After running the estimation of this model (see Table 8), we note that the corresponding likelihood further improved compared to the other two Bayesian models. However, even with the improved likelihood, the model still under performs when compared to the reinforcement model. The Vuong test[17] statistically favors the generalized reinforcement learning model.

Table 8 Parameter for elaborate Bayesian model

λ	k_{inf}	k_{sup}	δ	Likelihood
6.56 (0.01)	3 (-)	7 (0.02)	0 (0.18)	8584.4 (-)

5 Conclusion

In this paper, we have considered the choices to be made between familiar alternatives and unfamiliar alternatives for which there is feedback that is aggregated over different states of the economy. The literature on ambiguity aversion would

[17] The Vuong test was conducted with the null hypothesis that both models explain the data equally well. The null was rejected in favor of the generalized reinforcement learning model with V_stat= 20.18.

suggest that the unfamiliar alternatives would be discounted as compared with the familiar ones, but that literature has largely ignored how behaviors would change in the face of continuously coming new feedback that would remain aggregated over different states.

Several competing learning models could be considered to tackle the choices in the face of new feedback: either extensions of reinforcement models in the spirit of the valuation equilibrium (Jehiel and Samet, 2007) or Bayesian models in which subjects would start with some diffuse priors and update as well as they can based on the coarse feedback they receive. Clearly, ideas of ambiguity aversion can be combined with such learning models as well as the idea that subjects would make noisy best-responses to their representations of the alternatives as is ususally done in the discrete choice literature (McFadden) or in the literature on quantal response equilibrium (McKelvey and Plafrey, 1995).

Our results indicate that generalized reinforcement learning models outperform their Bayesian counterparts and that little discount seems to be applied to unfamiliar choices even when the feedback relative to them is aggregated over different states. As in other experimental findings, our results also indicate that subjects' choices are heterogeneous, which we have tackled by assuming that subjects employ noisy best-responses.

We believe such a work could be viewed as a starting point for an ambitious research agenda that aims at understanding how subjects make choices in the face of a mix of coarse and state-specific feedback. It seems well suited to cope with a number of choice problems in which one alternative is familiar and another one is not. Questions of whether subjects seek to generate state-specific feedback and when should also be part of this broader agenda.

References

[1] Camerer, C. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, pages 325–370.

[2] Camerer, C. (1999). Experience-Weighted Attraction Learning in Normal form games. *Econometrica*, 67.

[3] Camerer, C. (2003). *Behavorial Game Theory.* Princeton University Press.

[4] Cason, T. N., Sheremeta, R. M., and Zhang, J. (2012). Communication and efficiency in competitive coordination games. *Games and Economic Behavior*, 76:26–43.

[5] Charness, G. and Levin, D. (2005a). When optimal choices feel wrong: A laboratory study of bayesian updating, complexity,and affect. *The American Economic Review*, 95(4):1300–1309.

[6] Charness, G. and Levin, D. (2005b). When optimal choices feel wrong:a laboratory study of bayesian updating, complexity, and affect. *The American Economic Review*, 95(4):1300–1309.

[7] Cownden, D., Eriksson, K., and Strimling, P. (2018). The implications of learning across perceptually and strategically distinct situations. *Synthese*, 195:511–528.

[8] Elsberg, D. (1961). Risk, ambiguity and the savage axioms. *The Quarterly Journal of Economics*, pages 643–661.

[9] Epstein, L. G. and Schneider, M. (2007). Learning under ambiguity. *The Review of Economic Studies*, 74:1275–1303.

[10] Erev, I. and Roth, A. E. (1998). Predicting how people play games : Reinforcement learning in experimental games with unique mixed strategy. *The American Economic Review*, 88(4):848–881.

[11] Fudenberg, Drew.and Levine, D. K. (1998). *The theory of learning in games*. MIT Press.

[12] Gilboa, I. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153.

[13] Gilboa, I. and Schmeidler, D. (1993). Updating ambiguous beliefs. *The Journal of Economic Theory*, 59:33–49.

[14] Gilboa, I. and Schmeidler, D. (2001). *A theory of Case based decision*. Cambridge University Press.

[15] Gluck, M. A., Shohamy, D., and Myers, C. (2002). How do people solve the weather prediction task: Individual strategies for probabilistic category learning. *Learning and Memory*, 9(6):408–418.

[16] Grimm, V. and Mengel, F. (2012). An experiment on learning in a multiple games environment. *Journal of Economic Theory*, 147(6):2220–2259.

[17] Huck, S., Jehiel, P., and Ruttera, T. (2011). Feedback spillover and analogy-based expectations: A multi-game experiment. *Games and Economic Behavior*, 71(2):351–365.

[18] Hurwicz, L. (1951). The generalized bayes minmax principle: A criterion for decision making under uncertainty. *Cowles Comission Discussion Papers: Statistics no. 355.*

[19] Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic Theory*, 123(2):81–104.

[20] Jehiel, P. and Koessler, F. (2008). Revisiting games of incomplete information with analogy based equilibrium. *Games and Economic Behavior*, 62(2):533–557.

[21] Jehiel, P. and Samet, D. (2005). Learning to play extensive games by valuation. *Journal of Economics Theory*, 121(557):129–148.

[22] Jehiel, P. and Samet, D. (2007). Valuation equilibrium. *Theoretical Economics*, pages 163–185.

[23] Kenneth, T. (2003). *Discrete Choice Methods and Simulations.* Cambridge University Press.

[24] Machina, M. J. and Siniscalchi, M. (2013). *The Handbook of Economics of Risk and Uncertainty*, chapter 13. Elsevier North-Holland.

[25] McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10:6–38.

[26] Merkle, E. C., You, D., and Preacher, K. J. (2016). Testing non nested structural equation models. *Psychological Methods*, 21.

[27] Mullainathan, S., Schwartzstein, J., and Shleifer, A. (2008). Coarse thinking and persuasion. *The Quarterly Journal of Economics*, 123(2):577–619.

[28] Roland, F. and Jackson, M. (2008). A categorical model of cognition and biased decision making. *The B.E. Journal of Theoretical Economics*, 8(1):1–44.

[29] Samuelson, L. (2001). Analogies, adaptations and anomalies. *Journal of Economic Theory*, 97:320–367.

[30] Sobel, J. (2000). Economist model of Learning. *Journal of Economic Theory*, pages 241–261.

[31] .Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction.* Cambridge Press.

[32] Teck, H., Camerer, C. F., and Chong, J.-K. (2007). Self tuning experience weighted attraction learning in games. *Journal of Economic Theory*, 133.

[33] Tversky, A. (1992). Thinking through uncertainty: Non consequential reasoning and choice. *Cognitive Psychology.*

[34] Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, New Series*, 185(4157):1124–1131.

[35] Tversky, A. and Kahneman, D. (1979). Prospect theory, an analysis of decision under risk. *Econometrica*, 47(2):263–291.

[36] Weinstein, A. N. (2009). The ambiguity aversion literature a critical assessment. *Economics and Philosophy*, pages 249–284.

[37] Zeelenberg, E. (2003). The discounting of ambiguous information in economic decision making. *Journal of Decision making.*

6 Appendix

Appendix A

Instruction sheet for the players (In the lab the instructions in French):

Control Group:

Welcome to the experiment and I thank you for your participation. Please listen to these instructions carefully. If you have any questions kindly raise your hand and it shall be addressed. You receive 5 euros for participating and then your payoff depends on your performance in the experiment.

The Experiment:

The experiment consists of 70 rounds. It is a simple decision task. There are two situations you may face referred to as states 1 and 2. In each state, you have to choose one of two urns. Each urn is composed of ten balls either black or white in color. When you choose an urn, one of the balls in the urn is drawn at random (by the computer) and it is immediately replaced after the computer has noted the color of the ball. If the ball drawn is Black, you can receive extra payment (see below for details) whereas if the ball drawn is White you receive no payment.

The two urns available in state 1 are Blue and Red, respectively. The two urns available in state 2 are Green and Red, respectively. While the compositions of the various urns remain the same throughout the experiment, note that the compositions of the Red urn in state 1 need not be the same as the composition of the Red urn in state 2. These are two different urns.

As the experiment goes, on your computer screen, you will be informed whether you have to make a choice of urns in state 1 (Blue or Red) or in state 2 (Green or Red). The sequence of choices from states 1 or 2 is decided randomly by the computer. Your task is to choose one urn out of the two in each state.

Note: We drew 100 times a ball (replacing the ball in the urn after each draw) out of the Blue and Green urn. We obtained the following composition

Blue	30 Black	70 White
Green	68 Black	32 White

At the beginning of the experiment:

- Your terminal is randomly assigned a State of the world. If in State 1, you choose between a Red and Blue urn. If in State 2, you choose between a Red and Green urn.
- After you choose the color of the urn that you want to pick, you click on the screen. A ball (the color of which could be either Black or White) will be drawn from that urn by the computer. You will not know the color of the ball drawn. This implies you will not have the information for your choice.
- Once all participants have made their choices, we provide you with some feedback. The total number of black and white balls drawn in previous

rounds by all subjects according to the color of the urn (Blue, Red1, Red2, Green).

- Following the feedback, your terminal is randomly assigned a state of the world again. The state may vary from the previous round or remain the same.
- We then repeat the same experiment again until the completion of the 70 rounds.

For determining your payoff, two of the rounds will be randomly chosen at the end of the experiment. If one of your balls in these two rounds is Black, you will get an extra 5 euros. If both of your balls in these two rounds are Black, you will have an extra 10 euros. Otherwise (if both balls are White), you will have no extra return. So if you have no questions let us begin!

Treatment Group:

Welcome to the experiment and I thank you for your participation. Please listen to these instructions carefully. If you have any questions kindly raise your hand and it shall be addressed. You receive 5 euros for participating and then your payoff depends on your performance in the experiment.

The Experiment:

The experiment consists of 70 rounds. It is a simple decision task. There are two states of the world each containing two urns. One of the urns is Red and the other is either Blue or Green depending on the state (Note: The composition of the Red urn across states is not the same). On your computer screen, you will play different states of the world randomly. Your task is to choose one urn out of the two in each state. Each urn consists of 10 balls either black or white in color. The Black gives you a payoff of 5 euro whereas the White gives no payoff .

Note: We drew 100 balls randomly out of the Blue and Green urn which gave us the composition

Blue	30 Black	70 White
Green	68 Black	32 White

At the beginning of the experiment:

- Your terminal is randomly assigned a State of the world. If in State 1, you choose between a Red and Blue urn. If in State 2, you choose between a Red and Green urn.
- After you choose the color of the urn that you want to pick, you click on the screen. A ball (color of which could be either Black or White) will be picked up from that urn. You will not know the color of the ball drawn. This implies you will not have the information of your choice
- Once every participant has made their choice, we provide you with the feedback. The no. of black and white balls drawn from each colored urn(Blue, Red, Green) across states based on only the previous round draw is reported.

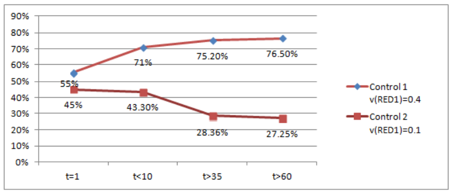
- Following the feedback, your terminal is randomly assigned a state of the world again. The state may vary from the previous round or remain the same. Note that the composition of the urn is however fixed throughout the experiment.

- We then repeat the same experiment again till we complete 70 rounds.

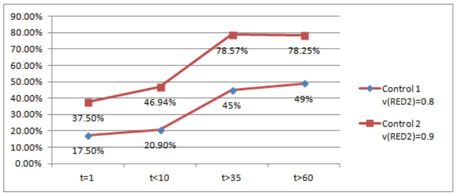
For determining your payoff, two of the rounds will be randomly chosen at the end of the experiment. If you have picked up B in that particular round, you end up with 5 euros more for each B otherwise no returns. So if you have no questions let us begin!

Appendix B

We report the choices made by individuals in the Control group below. The individuals were provided state specific feedback on the performance of the Red urn. We see a clear trend toward the optimal choices even if some noise would be needed to explain why only 49% of choices correspond to Red in state 2 in Control 1.



(a) Treatment 1: State 1



(b) Treatment 1: State 2

Appendix C

The diagrams below show that the parameters are normally distributed via Monte Carlo simulations

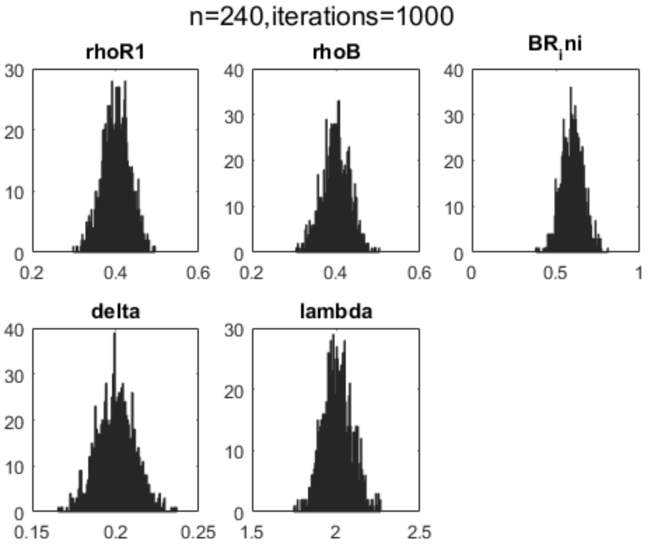
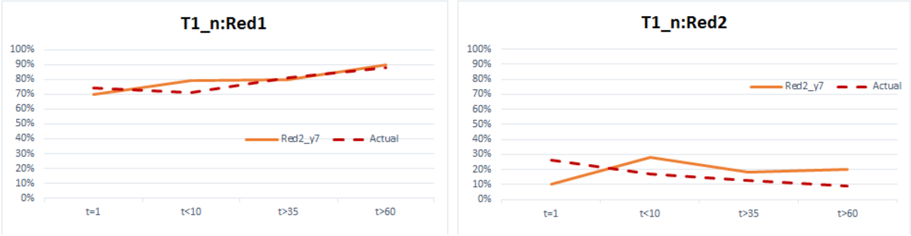


Figure 7 *Results for Montecarlo simulations for Model 1.There were 100 iterations with n=240 which is in line with number of players in our actual experiment.It can seen that the parameters are normally distributed and well identified.*

Appendix D

The figure shows simulated proportions of choices for the reinforcement model over 70 rounds with the estimated parameters for Treatment 1. Instead of using the noise parameter, λ= 5.23, we use λ= 7 to introduce less noise. This improves the fit of the simulated data with the actual one.



(a) Treatment 1: State 1

(b) Treatment 1: State 2