Updating beliefs with imperfect signals: experimental evidence[*]

François Poinas [†]   Julie Rosaz [‡]   Béatrice Roussillon [§]

July 17, 2012

Abstract

We conduct an experiment on individual choice under risk in which we study belief updating when an agent receives a signal that restricts the number of possible states of the world. Subjects observe a sample drawn from an urn and form initial beliefs about the urn's composition. We then elicit how beliefs are modified after subjects receive a signal that restricts the set of the possible urns from which the observed sample could have been drawn. We find that this type of signal increases the frequency of correct assessments and that prediction accuracy is higher for lower levels of risk. We also show that prediction accuracy is higher after invalidating signals (i.e. signals that contradict the initial belief). This pattern is explained by the lower level of risk associated with invalidating signals. Finally, we find evidence for a lack of persistence of choices under high risk.

JEL Classification: D83, C91

Keywords: Beliefs, Imperfect Information, Experiment

[†]Toulouse School of Economics (University Toulouse 1 Capitole - GREMAQ) - Toulouse F-31000, France - Email: francois.poinas@tse-fr.eu

[‡]GATE Lyon-St Etienne, CNRS, University of Lyon 2, F - 69130 Ecully, France.

[§]Université Pierre Mendès France, UMR 1215 GAEL, F-38000 Grenoble, France. INRA, UMR 1215 GAEL, F-38000 Grenoble, France - Email: beatrice.roussillon@upmf-grenoble.fr

We study an assessment problem in which there are several possible states of the world and agents have to determine which state they are facing. The agent first observes a sample drawn from all possible states of nature and then a signal that restricts the set of the possible states of nature. Before receiving the signal, the agent forms an initial belief based on observing the sample. This initial belief may or may not be included in the subset of possible states of nature that remains after receiving the signal. The objective of this paper is to analyze how this type of signal is used by agents to update their beliefs.

In the theoretical context of incomplete information, belief updating relies on the use of Bayes' theorem. However, experiments in both psychology and economics have shown that this theorem is rarely perfectly implemented in decisions taken under risk. Kahneman, Slovic, and Tversky (1982) describe heuristics that can explain consistently observed behaviors in probability updating. In risky situations, information is key for the economic agent and may take several forms, such as the probability of an event, an indication of the agent's behavior or a restriction of the possible states of nature.

A signal that reduces the number of possible states of nature is found in any situation in which an object or a person undertakes a pass/fail test, such as those preceeding the issuance of a label, a diploma or an audit notation. This type of test provides additional information and is supposed to make the evaluation of an object or a person easier. Many institutions, such as scientific journals, human resource consulting or even state medical licensing, adopt minimum certification standards[1] or pass/fail strategies. These signals do not allow precise quality assessment of an object or a person but indicate whether the object or the person is just above the minimum standard required.

---

[1]A standard is defined as a minimum level to achieve.

For example, we may consider this problem in the context of an employer-employee relationship. The employer, by observing the employee working in his firm, forms an initial belief about the employee's productivity. But the employee's productivity may be hard to assess due to difficulties in observing his behavior or due to multi-tasking. If this is the case, an employee's performance appraisal may be undertaken to acquire more information about his productivity level. For a sales person, for example, the performance appraisal may take the form of a sales objective. What information about the employee's productivity does the employer extract from the observation that the employee succeeds or fails to achieve his sales objective? He does not observe the precise level of the employee's productivity, but instead, a reduction in the range of possible productivity levels.

We examine how this type of signal is used by agents to update their belief by studying a two-stage task in a laboratory setting. In our experiment, subjects initially observe a sample of balls drawn from an urn composed of 20 balls, blue and yellow. We elicit their initial belief about the composition of balls in the urn by asking subjects to predict the total number of yellow balls contained in the urn. They then receive a true but imperfect signal about the urn's composition and are asked to once again predict the number of yellow balls contained in the urn. Varying the size of the observed sample and the precision of the signal allows us to study belief updating in situations characterized by different levels of risk.

On the whole, we find that the signals help subjects in their assessment task and that the prediction accuracy increases with the precision of the signal. Moreover, we show that the subjects' performance is higher after signals that contradict their initial belief than after signals that confirm it. After receiving a signal that confirms their initial belief, subjects' performance is poorer if they change their prediction between the two parts. This change in predictions happens when the initial risk

level associated with the state of nature is high and the signal does not reduce it. Thus, our results suggest beliefs change under risk, whereas the literature (Charness and Levin, 2005; Friedman, 1998) seems to support the assertion that beliefs are persistent under risk, in particular, when they are formed voluntarily. Indeed, we find that the absence of initial belief reinforcement by a validating signal seems to trigger belief updating. With a high initial level of risk, low accuracy validating signals appear to confuse the subjects.

This paper is organized as follows. Section 1 studies the literature on belief updating. Section 2 describes the experimental design. The data are analyzed in Section 3 and in Section 4 we discuss our results and conclude.

1 Related Literature

When faced with risk, economic agents use the available information to form their beliefs. In theory, they should use Bayes' rule to update their beliefs when they receive new information. However, many experiments in psychology and in economics have shown that people are not perfect Bayesian updaters and instead tend to use heuristics in dealing with uncertain events. Kahneman, Slovic, and Tversky (1982) analyze three heuristics used in the updating process: representativeness, availability and anchoring. We will focus our attention on representativeness, which is the most closely related to our work. The intuition behind representativeness is that people tend to associate an observed sample with the population it most resembles. When representativeness and Bayesian reasoning are not in line, the former may drive the updating process to non-maximizing behavior and lead to a number of errors. For instance, Kahneman, Slovic, and Tversky (1982) emphasize insensitivity to the prior probability of outcomes and insensitivity to sample size.

To study the insensitivity to prior probability Kahneman and Tversky (1973) conduct an experiment in which two groups of subjects face different descriptions of people. The first group knows that a population of 100 people is composed of 30 engineers and 70 lawyers. The second group knows that the population is composed of 70 engineers and 30 lawyers. The participants in both groups have to give the probability that a certain description corresponds to an engineer. When the description is uninformative, Baysian individuals in the first group should declare a probability of 0.3 and the participants in the second group should declare a probability of 0.7. However, the authors find participants in both groups declare the same probability of 0.5 independently of their base rate. The prior probability stated for the first (second) group that there are 30% (70%) of engineers in the population is ignored by the participants. Grether (1980) confirms, in an experiment that conforms to economic principles, that individuals do not use the base rate when updating beliefs. He demonstrates that individuals place less weight on the likelihood of information: they update their belief but less than they should do.

When using the representativeness heuristic, people tend to be insensitive to the size of the observed sample in the revision process. Tversky and Kahneman (1971) use an urn containing balls of two different colors to study this phenomenon. They show that a high proportion of balls of one color in the sample, even if this sample is small, induces subjects to believe that it comes from the urn that contains a large amount of balls of that color and leaves them more confident about it than if they had observed a bigger sample but with a less extreme proportion.[2] Thus, the composition of the sample seems to matter as much as the size of the sample. In our

---

[2]For instance, consider an urn containing balls of two colors: 2/3 are red and 1/3 are white. A first subject draws 6 balls (5 are red and 1 is white), and a second subject draws 20 balls (12 are red and 8 are white). Asking individuals about who of the two subjects should feel more confident that the urn from which the sample has been drawn contains 2/3 of red balls, they answer the first one.

5

experiment, the composition and the size of the sample vary to identify this possible source of error.

Charness and Levin (2005) and Friedman (1998) show the persistence in subjects' choices, and thus in beliefs, when Bayesian updating would require beliefs to change. Charness and Levin (2005) examine how the difference between clash and reinforcement affects the propensity to use Bayes' rule. Reinforcement supposes that people are more likely to choose actions associated with successful past outcomes. In their experiment, the subjects face two equally likely states of the world, "Up" and "Down", and two urns, "Left" and "Right", containing black or white balls. The Up state of the world is characterized by more black balls in both urns. Moreover, the Left urn has a mixed composition whereas the Right urn contains balls of only one color (only black balls in Up and only white balls in Down), thus indicating the state of the world for sure. The subjects make two draws. For the second draw they choose to continue drawing in the same urn or to switch. Three treatments are conducted. They differ by (i) the composition of the Left urn and (ii) the ability to choose from which urn the first ball is drawn. Bayesian predictions state that the participants should draw a second time from the Left urn (Right urn) after having drawn a white ball (black ball) since this represents a failure (success). Charness and Levin (2005) find that when Bayesian updating and reinforcement are not aligned around 50% of the decisions are inconsistent with Bayesian updating. They also find that when the urn used for the first draw is a choice, it is more likely to be chosen again than when the urn is imposed on the participants. This experiment points out the role of past successes and the degrees of freedom in subject's decisions in changing initial predictions. Our design shares similarities with Charness and Levin (2005) as we also ask for repeated assessments of the state of nature, which is represented by the

composition of an urn. In our experiment, however, we focus on the signal's impact and the precision of initial beliefs.

The consistency of subjects' choices is also studied by Friedman (1998) in his analysis of the three door paradox. The subject faces three face-down cards, one of which is the winning card. First, the subject chooses a card. Second, the experimenter turns over a non-winning card that the subject did not choose. Then the subject chooses one of the two remaining cards. Finally, the two cards are turned over. The participants play 10 trials. Bayes' rule predicts that the subjects should switch between their two choices. However, only 6/104 of the subjects switch more than half of the time. Friedman (1998) explains this recurrent behavior by the illusion of control (Camerer, 1995), i.e. the subjects' belief that their initial choice is the most likely, or by the non rational escalation of commitment (Bazerman, 1990), i.e. the persistence in choices is viewed as a virtue and flip flopping as a vice. Friedman (1998) proposes as an exploration of the three door paradox in which one manipulates the subject's initial choice by increasing the number of cards. In some ways we pursue this endeavour as the number of possible urns in our experiment is higher than the number of cards. Yet the subject's choice is a similar one and lies in choosing the correct urn/card in between all possible ones in two steps, where the first step represents the formation of an initial belief. The main difference between the two experiments is the fact that the signal may invalidate the initial belief of the subject. The experiments in Charness and Levin (2005) and Friedman (1998) are of particular interest to us as they underline the difficulty to predict subjects' choices.

Ashton and Ashton (1988) study the sequential belief revision in five experiments and how subject's beliefs are influenced by the order and the presentation of new evidence. They study the impact of validating/invalidating new information and the degree with which it confirms the initial belief. The participants are auditors.

They have to investigate the payroll records of a hypothetical client. They observe the result of a preliminary investigation that indicates the probability that controls would prevent errors was either 0.20, 0.50, or 0.80. Then, they receive validating or invalidating evidence that can be either strong or weak. The study finds that subjects are more willing to revise their belief after receiving new information, which stands in contrast with Bayesian updating behavior. Moreover, the subjects revise their belief to a greater extent after receiving new evidence that invalidates their initial belief.

Ouwersloot, Nijkamp, and Rietveld (1998) study how the characteristics of messages impact the size of errors in updating beliefs. Subjects are asked to assess a quantity. They are given a prior distribution of the unknown value and then receive information. The experiment confirms that people do not apply Bayes' rule. The authors find that the deviations from Bayesian behavior are impacted by the characteristics of the message, such as its precision, reliability, relevance and timeliness. However, the precision of the message has only a small impact on the error of prediction. The main limit of their experiment is that no proper incentives are given to subjects. Subjects receive a fixed reward that does not depend on the correctness of the answers. In our experiment subjects have to assess a risky state of nature incorporating the information given by a pass/fail test and are incentivized to give the correct answer.

A recent paper by Hoffman, Kagel, and Levin (2011) compares the impact of simultaneous and sequential arrival of information on updating behavior.[3] They also investigate the impact of the order of receiving bad and good news on updating. In a framed experiment, they ask subjects to determine the likelihood of having a genetic disorder. In the simultaneous treatment, subjects receive the results of two

---

[3]We thank an anonymous referee for having suggested this article.

biological tests at the same time. One of the tests declares that the subject has the genetic disorder, whereas the second one claims that he does not have it. In two sequential treatments, subjects receive the two results at a two week interval. After receiving the first result they are asked to determine the likelihood of actually having a genetic disorder. And then after two weeks, they have to determine it once again but this time taking into account the results of both tests. The two sequential treatments differ in the order in which good and bad news arrive. They show that both sequencing and the order of the signals matters. Sequential arrival of information biases the results compared to simultaneous arrival of information. Subjects attach more importance on the latest information they got than to the combined information. The authors note that one source of bias is the incomplete adjustment to the initial test result received.

The updating process is still to be investigated and understood. People do not seem to be exclusively Bayesian thinkers and there is a need for understanding the different processes used in their decision making. Our experiment contributes to this goal by analyzing specifically one kind of information: a signal that restricts the set of possible states of the world, or more precisely, the use of a pass/fail test to assess the characteristics of an object.

2   Experimental Design

2.1   The task

Subjects have to determine the number of yellow balls contained in an urn. The urn is composed of 20 balls that can be either yellow or blue. The number of yellow balls (between 0 and 20) is randomly drawn. Therefore, each of the 21 possible urn compositions (representing a particular state of nature) has the same probability of

being drawn. This is common information.

The subjects observe only a random sample drawn from this urn, whose size is either 6, 10 or 14 balls. The size of the sample is also randomly drawn.[4]

Each period consists of two parts. In the first part, subjects observe the sample drawn from the urn. They are then asked to predict the number of yellow balls in the urn based on observing the sample. We call this first prediction the *initial belief*.

In the second part, the same sample is kept visible and the subjects receive a signal about the urn's composition. The signal indicates that the urn is composed of either weakly more or strictly less than X yellow balls. The value on which the test is based is randomly chosen from amongst the three possible values of 6, 10 or 14.[5] This is common information. Therefore one of the following six signals is sent to the subjects: There are LESS THAN (AT LEAST) X yellow balls in the urn, with X being either 6, or 10, or 14. Subjects are then once again asked to predict the number of yellow balls in the urn.

Each subject plays 50 periods of this experiment. In each period, the urn's composition, the observed sample and the signal change.

2.2 The payoffs

Subjects were paid in cash in a separate room. They received a show-up fee of €10. Subjects were also paid according to their predictions in three randomly drawn periods. When they entered the payment room, they tossed a coin three times to determine which part of each selected period was paid. If one of the predictions over the three selected periods was correct, subjects received €10. If two were correct, they received €15 and if three were correct, €20.

---

[4] As an illustration, see the instructions in Appendix 1.

[5] Given that the urn and the sample composition are random, the numbers 6, 10 and 14 are chosen so as to ensure enough observations in each case.

This payoff rule induces subjects to provide, as a prediction, the number of yellow balls which they believe to have the highest probability to be correct.

2.3 Optimal predictions

Given the fact that the urn's composition, the observed sample and the signal are drawn randomly from known distributions, it is possible to compute the number of yellow balls that has the highest probability of being correct from a theoretical perspective. This number, which we call the optimal prediction, is obtained using Bayes' rule.[6] From the available information, Bayes' rule allows us to compute the probability associated with each urn from which the observed sample could have been drawn. And the optimal prediction corresponds to the number of yellow balls contained in the urn with the highest theoretical probability. The way this probability distribution is computed is provided in Appendix 2.

In the first part, the most probable urn is the one having the same proportion of yellow balls as the observed sample, since the number of yellow balls contained in the urn is drawn from a uniform distribution.[7] Therefore, in our setup, Bayesian revision and representativeness are in line.

The computation of the optimal prediction in the second part follows the same logic. Bayes' rule is applied to take into account the information provided by the signal, which restricts the set of potential urns from which the observed sample could have been drawn. Given that the signal is randomly selected, the urns that remain possible candidates after the signal is received are still ranked in the same order

---

[6]The objective of the paper is not to test if the subjects apply Bayes' rule in forming their beliefs. Instead, our goal is to study how individuals update their beliefs after observing a pass/fail test, regardless of the law or heuristic they use in their revision process. Nevertheless, we think that predictions computed with Bayes' rule represent an interesting point of comparison.

[7]For instance, if the observed sample contains 6 yellow balls out of 10, the most likely number of yellow balls contained in the urn is 12.

with respect to their probability of occurrence. Therefore, the optimal prediction in the second part is the same as the one in the first part when the corresponding urn is still present in the set of possible urns. When it is not, the optimal prediction corresponds to the urn in which the number of yellow balls is the closest to the first part's optimal prediction (see Appendix 3 for an example).

2.4 Procedures

The experiment was conducted at the GATE laboratory, Lyon, France, using the Regate software (Zeiliger, 2000). Using the ORSEE software (Greiner, 2004), we recruited 81 under-graduate students from local business and engineering schools.

At the beginning of each session each subject was randomly assigned to a computer. The subjects participated in the 50 periods of the experiment. At the end of the session, they filled out a demographic questionnaire and were asked to provide a brief description of how they made their decisions.

On average, a session lasted 60 minutes and subjects earned €15.

3   Results

First, we compare the predictions made by the subjects with the optimal ones, i.e. the ones that maximize the expected gain. On average, the optimal prediction is given in 36% of cases.[8] In 60% of the cases, the prediction error is only one ball with respect to the optimal prediction. We also observe that 20% of the subjects give the optimal prediction in more than 50% of the cases. The data do not reveal any other recurrent strategy. On average, subjects earned €15.16. If they had given the optimal prediction, the average gain would have been €17.70 (see calculations in

---

[8]In the post experimental questionnaire 28% of the subjects declare having used the representativeness heuristics to make their decisions.

Appendix 4).

When the subjects receive a signal that invalidates their initial belief, they modify their prediction in 98% of the cases and they choose the number of yellow balls that is closest to their first prediction in the new set of possible urns in 30% of the cases. When the subjects receive a signal that validates their initial belief, they change their prediction in 50% of the cases.

Overall, we can conclude that in a majority of cases subjects do not use Bayesian updating to make their predictions.

In what follows, we analyze how the signal affects the probability that the subject predicts the exact number of yellow balls in the urn. We then investigate whether signals that invalidate the initial belief affect the prediction accuracy differently than signals that validate it. Finally, we explain why subjects change their predictions when the signal validates their initial belief.

3.1 Impact of the signal on the prediction accuracy

For each subject we compute the proportion of correct predictions as the number of predictions equal to the number of yellow balls contained in the urn divided by the total number of decisions taken. Figure 1 plots the density of the proportion of correct predictions in the whole sample. Figure 1.A (resp. Figure 1.B) shows this density for predictions realized in the first (respectively second) part. A Wilcoxon matched-pairs signed-ranks test[9] shows that the two distributions of the proportion of correct predictions in part 1 and in part 2 are significantly different (p-value = 0.000) and that the median of the distribution of correct predictions is higher in part 2 than in part 1 (p-value = 0.000). Therefore, on average, signals sent to the agents help them to correctly assess the number of yellow balls contained in the urn.

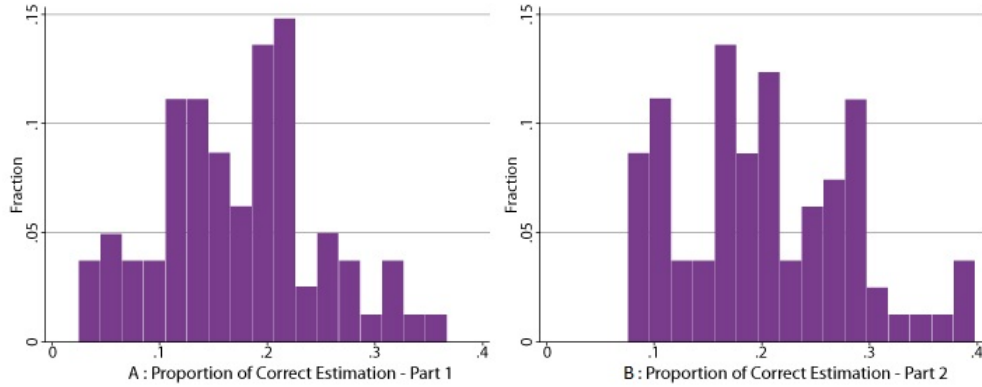[9]Test of equality of distributions on matched data.

Figure 1: Density of the proportion of correct predictions.

Result 1: The signal improves the accuracy of the subjects' predictions.

In order to understand how the signal helps the subjects in assessing the number of yellow balls contained in the urn we analyze the determinants of the probability that the prediction made in part 2 (i.e. after the signal is received) is correct. The potential determinants are the composition of the observed sample, the initial risk level (before the signal is received), and the final risk level (after the signal is received). Each subject makes predictions during a subsequent number of periods. The post-experimental questionnaire shows an important heterogeneity in the strategies reported by the subjects. This heterogeneity also appears in the data, as illustrated in Figure 1. Therefore, there might be some underlying characteristics (observable or not) that affect the decisions of the subjects. We account for the individual characteristics persistent over time by (i) introducing individual observed characteristics (reported by the subjects in the questionnaire) and (ii) allowing for the presence of unobserved permanent characteristics by exploiting the panel dimension of our data.

We estimate a random effect logit model, that takes the following form:

$$\Pr(Y_{it} = 1) = \beta_{OS} OS_{it} + \beta_{IR} IR_{it} + \beta_{FR} FR_{it} + \beta_X X_i + \alpha_i + \epsilon_{it}, \qquad (1)$$

where Y_{it} is an indicator variable equal to 1 if the prediction made in part 2 by

14

subject i in period t is correct. OS_{it} represents the composition of the observed sample, measured by the observed proportion of yellow balls and discretized into three categories.[10] IR_{it} represents the initial risk level. In the context of our experimental design, the initial risk level may either be high (when the observed sample is composed of 6 visible balls), medium (10 visible balls) or low (14 visible balls). Therefore, a high risk level means that there are 15 possible urns from which the observed sample may be drawn. Medium and low risk levels correspond to 11 and 7 possible urns respectively. FR_{it} represents the final risk level, measured as the number of possible urns remaining after the signal is received. X_i is a set of time-invariant observed characteristics captured in the questionnaire (it includes indicators for gender, the field of study and the first participation in an experiment). α_i is a time-invariant unobserved term assumed to follow a normal distribution and ϵ_{it} is a period-individual specific error term following a standard logistic distribution.

The results are reported in Table 1. In the first specification (column 1), we include only the potential determinants that are observed by the subjects before the signal is received: the composition of the observed sample and the initial risk level. We observe that the higher the initial risk level, the lower the prediction accuracy. The composition of the observed sample also plays a key role: subjects perform better when the observed sample is composed of almost only yellow balls (high proportion) or almost only blue balls (low proportion). We find evidence similar to that of Tversky and Kahneman (1971), who show that the composition of the sample is influencing the updating behavior. In the second specification, we include the final risk level as a third potential explanatory factor. This element represents the information provided by the signal: the more precise the signal, the

---

[10]The three categories are: (i) the proportion of yellow balls is lower than 0.1, (ii) the proportion of yellow balls is between 0.1 and 0.9, and (iii) the proportion of yellow balls is higher than 0.9.

lower the number of possible urns in part 2. We find that a decrease in the final risk level has a positive impact on prediction accuracy. More interestingly, the coefficients associated to the initial risk level become insignificant. Hence, only the risk level remaining after the signal is received seems to influence prediction accuracy in part 2.

Result 2: Prediction accuracy increases with the precision of the signal.

3.2 Impact of invalidating signals on the prediction accuracy

Let us now turn to the question of how the prediction accuracy in part 2 is affected by signals that confirm the initial belief (validating signals) on the one hand and signals that do not (invalidating signals) on the other. There are 85% of validating signals and 15% of invalidating signals. The first column of Table 2 reports the results of the estimation of equation (1) in which an indicator of an invalidating signal is introduced as an additional explanatory variable. The first column provides results when the final risk level is not controlled for. We find that the probability that the prediction is correct in part 2 is higher after an invalidating signal. This means that a signal that does not confirm the subject's initial belief improves a subject's prediction by more than a signal that confirms it.

In order to find an explanation for this pattern one has to remember that invalidating signals happen for two reasons: the subjects make a bad prediction in part 1 (far from the optimal prediction) and/or the observed sample is a bad representation of the urn's composition. Thus, the positive impact of invalidating signals on prediction accuracy may be caused (totally or partially) by the fact that invalidating signals happen in situations in which the subjects are left with a low number of possible urns.[11] We examine this potential explanation by running the same regression,

---

[11]Indeed, the coefficient of correlation between the number of possible urns after a signal and invalidating signals is negative and high (-0.301).

Table 1: Impact of the Signal on the Prediction Accuracy in Part 2

Dependent variable : Indicator of correct prediction in part 2		
	(1)	(2)
Composition of the observed sample:		
Proportion of yellow balls $\in [0, 0.1]$	0.271**	0.325***
	(0.118)	(0.120)
Proportion of yellow balls $\in (0.1, 0.9)$	ref.	ref.
Proportion of yellow balls $\in [0.9, 1]$	0.242**	0.421***
	(0.121)	(0.124)
Initial risk level:		
Low	0.248***	-0.106
	(0.097)	(0.104)
Medium	ref.	ref.
High	-0.264**	-0.070
	(0.105)	(0.108)
Final risk level:		
Number of possible urns in part 2	-	-0.163***
		(0.019)
Constant	-1.548***	-0.216
	(0.126)	(0.194)
Control for individual observed characteristics	yes	yes
Observations	3,850	3,850
Log-likelihood	-1905.429	-1866.237
ρ	0.027**	0.026**

Logit model with random effects. Standard errors in parentheses. Significance levels: *** 1%; ** 5%; * 10%. ρ denotes the share of the individual (within-group) variance in the total variance. High, medium and low initial risk levels correspond to situations in which the number of possible urns in part 1 is respectively 15, 11 and 7. Individual observed characteristics are indicators for gender, the field of study and the first participation in an experiment.

Table 2: Impact of Invalidating Signals on Prediction Accuracy in Part 2

Dependent variable : Indicator of correct prediction in part 2		
	(1)	(2)
Invalidating signal indicator	0.535***	0.080
	(0.115)	(0.130)
Number of possible urns in part 2	-	-0.157***
		(0.035)
Constant	-1.633***	-0.275
	(0.130)	(0.216)
Control for initial risk level	yes	yes
Control for observed sample composition	yes	yes
Control for individual observed characteristics	yes	yes
Observations	3,850	3,850
Log-likelihood	-1895.162	-1866.049
ρ	0.030***	0.027***

Logit model with random effects. Standard errors in parentheses. Significance levels: *** 1%; ** 5%; * 10%. ρ denotes the share of the individual (within-group) variance in the total variance. Individual observed characteristics are indicators for gender, the field of study and the first participation in an experiment.

this time controlling for the final risk level, measured as the number of possible urns left after the signal is received. The results are shown in the second column of Table 2. We find that the higher the final risk level, the lower the prediction accuracy (the impact of the number of possible urns in part 2 is negative). Moreover, the impact of invalidating signals looses significance. This means that the higher probability of prediction accuracy after an invalidating signal can be attributed to the fact that invalidating signals reduce the risk level associated with the urn's composition by more than validating signals and subjects are thus left with fewer options.

Result 3: Prediction accuracy is higher after invalidating signals than validating signals. This is explained by the lower risk level left after invalidating signals.

Let us now examine the prediction accuracy after validating signals more closely, since it appears that the prediction accuracy is lower in such situations. When

subjects receive a signal that confirms their initial belief, they change their prediction from the one made in part 1 in half of the cases. We examine if the decision of changing the prediction affects the prediction accuracy. In order to do so, we estimate equation (1) in which we introduce an indicator variable for a prediction change between the two parts as an explanatory variable. This estimation is conducted only on observations for which the signal validates the initial belief. Results are presented in Table 3. The first column provides results when the final risk level is not controlled for. It shows that the probability of a correct prediction is lower when the subjects change their prediction between the two parts. Again, one may wonder if this result can be explained by the risk level remaining after the signal is received. We thus run the same regression, this time introducing the number of possible urns in part 2 as an explanatory variable. Consistent with the results reported in the second column of Table 2, we find that a higher risk level has a negative impact on the prediction accuracy. Nevertheless, the impact of a prediction change remains negative and highly significant after we control for the final risk level. We can conclude that validating signals have an adverse effect. In half of the cases in which subjects receive a signal that confirms their initial belief they change their prediction. And this updating behavior has a negative impact on the prediction accuracy.

Result 4: Signals that validate the initial belief make subjects change their predictions in half of the cases. The prediction change has a negative impact on subjects' performances.

Table 3: Impact of Prediction Change on Prediction Accuracy in Part 2 when the Signal Validates the Initial Belief

Dependent variable : Indicator of correct prediction in part 2		
	(1)	(2)
Indicator of prediction change	-0.420***	-0.403***
	(0.100)	(0.100)
Number of possible urns in part 2	-	-0.138***
		(0.022)
Constant	-1.426***	-0.244
	(0.141)	(0.233)
Control for initial risk level	yes	yes
Control for observed sample composition	yes	yes
Control for individual observed characteristics	yes	yes
Observations	3,316	3,316
Log-likelihood	-1591.493	-1572.065
ρ	0.023***	0.021***

Logit model with random effects. Standard errors in parentheses. Significance levels: *** 1%; ** 5%; * 10%. ρ denotes the share of the individual (within-group) variance in the total variance. Individual observed characteristics are indicators for gender, the field of study and the first participation in an experiment.

3.3 Explaining prediction change when the signal validates the initial belief

In this section, we analyze situations in which subjects change their initial belief when the signal validates it. Such situations are interesting in light of the results in section 3.2, which showed that this behavior has a negative impact on the probability that a subject's prediction is correct. Moreover, the existing literature exhibits belief persistence instead of belief change when dealing with risky situations (Charness and Levin, 2005; Friedman, 1998). Thus, we investigate whether there are some situations in which information revealed to the subjects may make them more likely to change their prediction.

We estimate a random effect logit model in which the dependent variable is the indicator of a prediction change between parts 1 and 2. This estimation is performed on observations for which the signal confirms the initial belief. We explore the role

of the two factors that may explain why people change their beliefs: the evolution of the risk level associated with the state of nature between the first and the second part and the (objective) probability that the prediction made in part 1 is correct.

The evolution of the risk level takes into account both the risk level in the first part and the reduction of the risk level between parts 1 and 2 induced by the signal. The initial risk level may be either high (15 possible urns from which the observed sample is drawn), medium (11 urns) or low (7 urns). The signal reduces the number of possible urns from which the sample might be drawn. In order to obtain the final risk level, the number of possible urns left after receiving the signal is discretized into three categories, high, medium and low, if the number of possible urns is 12 or more, between 8 and 11, or 7 or less respectively. The evolution of the risk level between the two parts therefore falls into one of six categories: high-to-high, high-to-medium, high-to-low (taken as the reference category because it corresponds to the highest decrease in risk between the two parts), medium-to-medium, medium-to-low, and low-to-low.

The probability that the prediction in part 1 is correct corresponds to the probability that the number of yellow balls contained in the urn is equal to the prediction made by the subject in part 1, given the signal he received. It is calculated using Bayes' rule. This variable takes into account the composition of the observed sample and the prediction that the subjects make in part 1. Therefore, it allows us to control for both the difficulty of the assessment problem (some urns are more difficult to assess than others) and the subject's performance in part 1.

Table 4 shows the results of this estimation. We find that only being in a situation of high initial risk level that is not reduced after the signal is received has an increasing impact on the propensity to update the belief when the signal received validates the initial belief. This result suggests that the decision to change one's

Table 4: Determinants of Belief Updating when the Signal Validates the Initial Belief

Dependent variable: Indicator for prediction change		
Evolution of the risk level between the two parts:		
high → high	0.516**	(0.260)
high → medium	0.381	(0.256)
high → low	ref.	
medium → medium	0.355	(0.237)
medium → low	0.055	(0.261)
low → low	0.327	(0.231)
Probability of correct prediction at part 1	-2.596***	(0.349)
Constant	0.595	(0.569)
Control for individual observed characteristics	yes	
Observations	3,316	
Log-likelihood	-1491.910	
ρ	0.596***	

Logit model with random effects. Standard errors in parentheses. Significance levels: *** 1%; ** 5%; * 10%. ρ denotes the share of the individual (within-group) variance in the total variance. High, medium and low initial risk levels correspond to situations in which the number of possible urns in part 1 is respectively 15, 11 and 7. High, medium and low final risk levels correspond to situations in which the number of possible urns in part 2 is respectively 12 or more, between 8 and 11, and 7 or less. Individual observed characteristics are indicators for gender, the field of study and the first participation in an experiment.

prediction is neither affected only by the initial risk level nor only by the fact that the signal reduces the number of possible urns. Rather, it is the combination of these two factors that plays a key role: the prediction change is more likely when the subjects observe a sample of 6 balls and afterward receive a signal of low accuracy. Concerning the impact of the sample composition and the subject's first prediction, we find that the probability that the prediction made in part 1 is correct decreases the propensity to change the prediction when a validating signal is received. This result suggests that when the subjects are more confident about their initial belief, they are less likely to change it after receiving a validating signal.[12]

Result 5: When the signal confirms the initial belief, the subjects are more likely to change their prediction when the initial risk level is high

---

[12]A closer look at the data makes it apparent that the subjects are less likely to change their prediction when the observed sample is composed of around 100% or 0% of yellow balls and subjects play the optimal strategy (their predictions are close to 20 or 0 yellow balls respectively) and regardless of the sample size.

and remains high after the signal is received.

4 Conclusion

In this paper, we analyze how individuals update their beliefs when they receive an imperfect signal. Our original contribution to the literature resides in the type of signal sent to agents, which is a restriction of the set of possible states of nature. More precisely, we analyze how subjects update their beliefs after receiving this kind of signal for different levels of risk associated with the state of nature and for different signal accuracies. We examine these questions in a laboratory setting with a two-stage task. The subjects observe a partially revealed urn composed of 20 balls (yellow or blue) and have to assess the total number of yellow balls in the urn twice, once before a signal of the urn's composition is observed, and once after such a signal is observed.

We find that, on average, signals significantly help subjects predict the number of yellow balls contained in the urn. Prediction accuracy decreases with the level of risk (less precise signals). A signal that disproves a mistaken initial belief, i.e. an invalidating signal, increases the probability of finding the exact composition of the urn. Validating signals, however, may have an adverse effect. Subjects change their prediction in half of the cases following a signal that validates their initial belief and this change in prediction is associated with lower levels of prediction accuracy. Furthermore, we show that in cases in which the signal confirms their initial beliefs, subjects are more likely to change their prediction when the initial risk level associated with the state of nature is high, i.e. the observed sample is small, and the signal is not precise.

Our results suggest that signals corresponding to pass/fail tests are useful in improving the assessment of the state of nature, especially when the first prediction

is wrong. However, we find a limitation on the ability of this type of signal to better inform agents about the state of nature. Moreover, individuals are more likely to revise their beliefs in situations where the state of nature remains risky after the signal is received.

The literature on updating beliefs has highlighted the persistence of subjects' beliefs, whereas Bayesian updating requires changes (Charness and Levin, 2005; Friedman, 1998). Our results suggest that subjects change their beliefs, even when Bayesian updating requires persistence. We show that the evolution of risk associated with the state of the world significantly affects subjects' behavior. Agents are more willing to change their belief facing a riskier state of the world.

Our framework could readily be extended to introduce a measure of the degree of confidence subjects have in their predictions. In a new experimental design, we could ask subjects to provide a confidence interval around their predictions. We expect that the higher a subject's confidence level in their initial belief (first prediction), the lower the likelihood that they change their initial belief. In this new setting, we could also investigate if there exist some specific signals or situations that make agents want to change their initial beliefs even if they are confident about them.

References

Ashton, A.H., and R.H. Ashton. 1988. "Sequential Belief Revision in Auditing." *The Accounting Review* 63:pp. 623–641.

Bazerman, M. 1990. *Judgement in managerial decision making*. New York: Wiley.

Camerer, C. 1995. *Handbook of Experimental Economics*, Princeton University Press, chap. Individual Decision Making. pp. 587–703.

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

Friedman, D. 1998. "Monty Hall's Three Doors: Construction and Deconstruction of a Choice Anomaly." *American Economic Review* 88 (4):933–946.

Greiner, B. 2004. "An Online Recruitment System for Economic Experiments." *Kurt Kremer, Voler Macho (eds). Forschung und wissenschaftliches Rechnen, GWDG Bericht 63, Gttingen: Ges. fr Wiss. Datenverarbeitung*, pp. 79–93.

Grether, D. 1980. "Bayes Rule as a Descriptive Model: The Representativness Heuristic." *Quaterly Journal of Economics* 95:537–557.

Hoffman, R.M., J.H. Kagel, and D. Levin. 2011. "Simultaneous versus sequential information processing." *Economics Letters* 112:16–18.

Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kahneman, D., and A. Tversky. 1973. "On the psychology of prediction." *Psychological Review* 80:237–251.

Ouwersloot, H., P. Nijkamp, and P. Rietveld. 1998. "Errors in probability updating behaviour : Measurement and impact analysis." *Journal of Economic Psychology* 19:535–563.

Tversky, A., and D. Kahneman. 1971. "Belief in the law of small numbers." *Psychological Bulletin* 76:105–110.

Zeiliger, R. 2000. "A Presentation of Regate, Internet Based Software for Experimental Economics."

Appendix 1: Instructions (original in French)

You are about to participate in an experiment during which you can earn money. Your earnings will depend on your decisions taken during the experiment. Your earnings will be paid to you in cash in private and in a separate room.

There are 50 periods in the experiment. Each period has two parts. In each period, you see a partially revealed urn containing 20 balls. The balls may be yellow or blue. In each part, you have to predict the total number of yellow balls in the urn. Your earnings depend on your prediction of the total number of yellow balls. You increase your earnings if you find the exact total number of yellow balls in the urn.

Period description   An urn is composed of **20 balls**: yellow or blue. The number of yellow balls is randomly determined between 0 and 20. There are equal chances that the urn is composed of 0 yellow balls, or 1 yellow ball,..., or 20 yellow balls. The urn stays the same during the period. You can observe only 6, 10 or 14 balls in the urn. The visible balls are randomly drawn from the urn.

Each period has 2 parts. The set of visible balls is identical for the 2 parts.

   PART 1: You have to predict the total number of yellow balls in the urn. Enter your answer with the scroll bar and validate it.
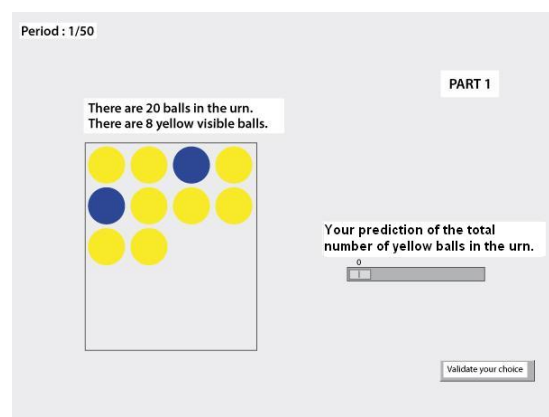
   PART 2: The urn stays the same as in the 1st part. You receive complementary

information on the urn's composition. This information indicates if the urn is composed with at most, or at least 6, 10 or 14 yellow balls. Then, you have to predict a second time the total number of yellow balls in the urn. Enter your answer with the scroll bars, even if your answer does not differ from your prediction in the 1st part, and validate it.
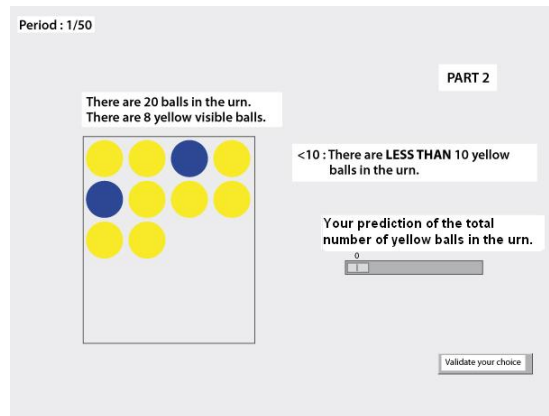
Description of the added information on the urn's composition   You can receive one of the following pieces of information:

- ≥ 6, there are AT LEAST 6 yellow balls in the urn.

- < 6, there are LESS THAN 6 yellow balls in the urn.

- ≥ 10, there are AT LEAST 10 yellow balls in the urn.

- < 10, there are LESS THAN 10 yellow balls in the urn.

- ≥ 14, there are AT LEAST 14 yellow balls in the urn.

- < 14, there are LESS THAN 14 yellow balls in the urn.

All six possibilities are not simultaneously possible for a given composition of the urn. For example, when an urn has a total number of 8 yellow balls, only three things out of six are possible:

- ≥ 6, there are AT LEAST 6 yellow balls in the urn.

- < 10, there are LESS THAN 10 yellow balls in the urn.

- < 14, there are LESS THAN 14 yellow balls in the urn.

The computer draws randomly and shows one of the possible pieces of information.

At the end of this 2nd part, you observe the complete urn and the predictions you made in part 1 and in part 2.

A new period starts automatically when you validate the last screen. At each new period, your task remains the same, but the following elements can be different:

- the urn's composition

- the number of visible balls

- the information on the maximum or minimum total number of yellow balls in the urn.

Payment   Three periods over the 50 played are selected for your payoff. The three periods are randomly drawn by the computer at the end of the experiment. They can be different among the participants. Thus, you do not know in advance which periods will be selected. Each period has the same chance to be selected.

For each of the three selected periods, only one part is chosen for your payoff. When you enter the payment room, you determine the part selected for the payoff by tossing a coin:

- if you toss tails, the 1st part is selected for your payoff.

- if you toss heads, the 2nd part is selected for your payoff.

To summarize, three periods are selected randomly by the computer and, for each of the selected periods, you determine which part will be paid by tossing a coin. Thus, three parts are selected for your payoff. Your payoff depends on the number of exact predictions you have made for these three parts:

- If none of your three predictions is exact: your payoff is €0.

- If one of your three predictions is exact: your payoff is €10.

- If two of your three predictions are exact: your payoff is €15.

- If all your three predictions are exact: your payoff is €20.

In all the cases, you will receive €10 more for your participation in this experiment.

It is forbidden to communicate with other subjects during the experiment. If you have any question regarding these instructions, please raise your hand.

Appendix 2: Predicted Probabilities

Predicted Probabilities in Part 1

An urn is composed of 20 balls and the number of yellow balls contained in the urn is chosen randomly (from 0 to 20). Therefore, there are 21 possible urns, with each one having the same probability (1/21) of being drawn. Then, the probability to observe sample S, given the urn from which the sample is drawn is U_i is:

$$P(S|U_i) = \frac{C_{y_u}^{y_s} * C_{n_u-y_u}^{n_s-y_s}}{C_{n_u}^{n_s}}$$

where n_u (resp. n_s) is the number of balls contained in urn U_i, $n_u = 20$, (resp. in sample S) and y_u (resp. y_s) represents the number of yellow balls contained in urn U_i (resp. in sample S). Then, according to Bayes' rule and given an observed sample S, the probability that this sample is drawn from urn U_i is equal to:

$$P(U_i|S) = \frac{P(S|U_i) \cdot P(U_i)}{\sum_{x=1}^{21} P(S|U_x) \cdot P(U_x)}$$

Given an observed sample S, the distribution of probabilities over the 21 possible urns is obtained using this formula. The optimal choice for the subject is thus to give as a prediction the number of yellow balls contained in the urn having the highest probability.

Predicted Probabilities in Part 2

Now, we have to compute the probability that a particular observed sample is drawn from each of the 21 possible urns, after a signal is received. According to Bayes' rule, the probability that sample S is drawn from urn U_i given signal I (I stands for information) is:

$$P(U_i|S,I) = \frac{P(S,I|U_i) \cdot P(U_i)}{\sum_{x=1}^{21} P(S,I|U_x) \cdot P(U_x)}$$

The signal sent to the agent is only based on the composition of the urn. In particular, it is independent from the composition of the observed sample drawn from the urn in part 1. Therefore, $P(S, I|U_i) = P(S|U_i) \cdot P(I|U_i)$. $P(I|U_i)$ is computed according to the following logic. For a particular urn, among the six existing signals, only three can be sent to the subject. Indeed, the number of yellow balls contained in an urn cannot be both inferior and superior to the same number. Since the signal is randomly chosen, the probability that a signal is sent is equal to one third if the signal is a possible one, and 0 if the signal is inconsistent with the composition of the urn. This leads to the following:

$$P(U_i|S, I) = \frac{P(S|U_i) \cdot P(I|U_i) \cdot P(U_i)}{\sum_{x=1}^{21} P(S|U_x) \cdot P(I|U_x) \cdot P(U_x)}$$

with

$$P(I|U_i) = \begin{cases} 1/3 & \text{if signal } I \text{ is consistent with the composition of urn } U_i \\ 0 & \text{if not} \end{cases}$$

As is the case in the first part, the optimal choice for the subject is to give as a prediction the number of yellow balls contained in the urn having the highest probability, given the sample and the signal.

Appendix 3: Example of optimal prediction

Figures 2 and 3 display the probability distribution that different predictions are correct for an urn where the observed sample is composed of 6 yellow balls and 4 blue balls.
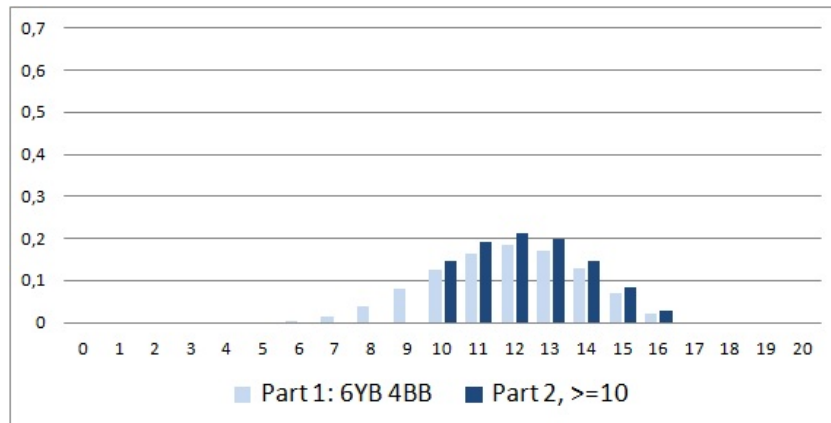

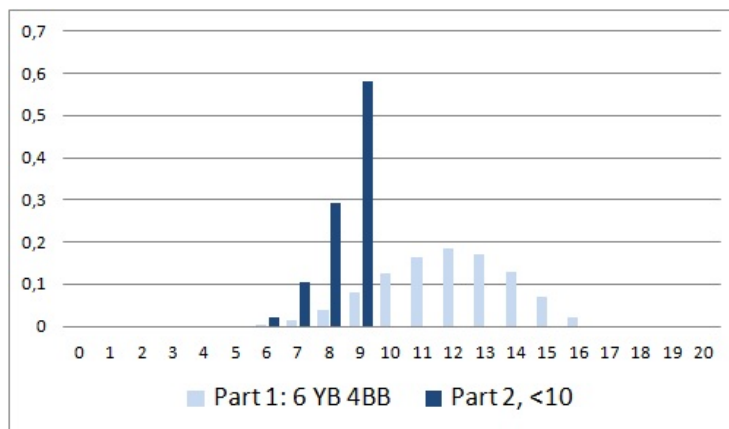Figure 2: Optimal predictions for a validating signal.

Figure 3: Optimal predictions for an invalidating signal.

In both figures, the optimal prediction in the first part (represented by light bars) is 12 yellow balls as it is the number of yellow balls associated with the urn having the highest probability to be correct. Figure 2 represents the case where the subject receives a signal that validates his initial belief (indeed the signal states that "there are at least 10 yellow balls"). Each one of the urns remaining in the new

restricted set has a higher probability of being correct. Thus, the optimal prediction in part 2 (represented by dark bars) remains the same (12) and its probability of being correct increases. Figure 3 represents the case where the signal invalidates the first prediction. The signal states that "there are less than 10 yellow balls" in the urn and proves the optimal prediction to be wrong. In this case, the optimal prediction in the second part corresponds to 9, which is the closest number to 12 included in the new restricted set.

Appendix 4: Expected revenue

The expected gain for one subject depends both on the number of correct predictions he gives and on the likelihood that the period(s) at which he makes correct predictions is(are) drawn at the end of the experiment. Three periods are drawn. The probability that, among those three periods, a subject makes i correct predictions (and therefore makes $3 - i$ wrong predictions) is given by:

$$P(i) = \frac{C_n^i * C_{n-c}^{3-i}}{C_n^3}, \text{ for } i = 0, \ldots, 3$$

where n represents the total number of periods played and c the total number of correct predictions.

The expected gain is the average of the gains associated with 0, 1, 2 and 3 correct answers drawn, weighted by the probability that the corresponding number of correct answers is drawn:

$$ER(n,c) = \frac{C_c^0 * C_{n-c}^3}{C_n^3} * 10 + \frac{C_c^1 * C_{n-c}^2}{C_n^3} * 20 + \frac{C_c^2 * C_{n-c}^1}{C_n^3} * 25 + \frac{C_c^3 * C_{n-c}^0}{C_n^3} * 30$$

Given that, for the whole experience, the subjects made 7900 predictions, among which 1492 were correct ones, the total expected revenue is $ER(7900, 1492) =$€15.16. If every subject had played with the optimal strategy (consisting in giving as a prediction the urn associated with the highest probability), they would have made 2340 correct predictions. In this case, the expected revenue would have been $ER(7900, 2340) =$€17.70.